Formation of Structure in the Univer

Nicolas Laporte

Lent Term 2023

UNIVERSITY OF CAMBRIDGE

Formation of Structure in the Universe

Nicolas Laporte

Lent Term 2023

UNIVERSITY OF CAMBRIDGE

Contents

1 Introduction 1

1.1 The first 3 minutes of the Universe ... 3
1.2 Formation of first stars and galaxies ... 4

2 Physical processes in baryonic gas 11

2.1 Radiation processes in Astrophysics ... 12
2.1.1 Description of a radiation field ... 13
2.1.2 Radiative Transfer ... 15
2.1.3 Thermal Radiation ... 19
2.2 Line emission ... 21
2.2.1 The Einstein coefficients ... 21
2.2.2 Absorption and Emission coefficients in terms of Einstein coefficients ... 23
2.2.3 Excitation of lines by collisions ... 25
2.2.4 Astrophysics terminology ... 26
2.2.5 An example : the Hydrogen atom ... 27
2.3 Heating and cooling ... 29
2.3.1 Definition ... 29
2.3.2 The cooling curve ... 30
2.3.3 Cooling by line emission ... 30
2.3.4 Cooling by free-free emission in ionised gas ... 32
2.3.5 Cooling of molecular gas ... 35

2.3.6 Cooling by dust 36
2.3.7 Radiative heating and cooling by recombination 37
2.3.8 Mechanical heating 39
2.4 The multi-phase ISM 39

3 Gravitational stability and instability 42

3.1 Equations of hydrodynamics and hydrostatic equilibrium 43
3.2 The isothermal sphere 44
3.2.1 Singular isothermal sphere 45
3.2.2 General solution 46
3.3 The polytropic sphere 48
3.4 Virial equilibrium for the self-gravitating sphere 48
3.5 Stability of an isothermal cloud 49
3.6 Jeans instability 51
3.7 Magnetic fields 53
3.8 Application to molecular clouds 54

4 Gravitational collapse 55

4.1 Free-fall time 56
4.2 Inside-out collapse 57
4.2.1 Similarity Analysis 58
4.2.2 Physics Analysis 59

5 From gas cloud to collapsed object 62

5.1 Basic physics of object formation 63
5.2 Evolution of the first core 64
5.3 Structure around the protostar 64
5.3.1 Accretion luminosity 65
5.3.2 Accretion shock and dust envelope 65
5.4 Evolution of the protostar 66
5.4.1 Protostellar structure during accretion 66
5.4.2 Onset of deuterium burning and convection 70
5.4.3 Deuterium shell burning 71
5.4.4 Contraction and hydrogen burning 71

6 Galaxies and star-formation on galactic scales 75

6.1 Properties of Galaxies 76
6.1.1 The Galaxy Zoo 76
6.1.2 The galaxy Luminosity Function 79
6.1.3 Stellar population 81
6.2 Cloud fragmentation 88
6.3 Galactic-wide star formation 93
6.4 Simple models of gas and star formation evolution in galaxies 95
6.4.1 Closed-box model of star formation in a galaxy 96
6.4.2 The effect of inflows and the gas regulator (or "bathtub") model 96
6.4.3 The effect of outflows 99
6.5 Metallicity evolution of galaxies 100

6.6 Stellar orbits and spiral structure 103
6.6.1 Rotation curves in galaxy disks 103
6.6.2 Stellar Orbits 105
6.6.3 Resonant orbits 107
6.6.4 Stability of a rotating disc - spiral density waves 108
6.6.5 Lindblad Resonances 111
6.6.6 The threshold of the Schmidt-Kennicutt relation 111

7 Feedback processes in star formation 113

7.1 Stellar Winds 114
7.2 Feedback from O-B stars 115
7.3 Supernovae 115
7.4 Starburst and ultraluminous galaxies 117
7.5 Global stellar feedback 118
7.6 Feedback from Black Holes 118
7.6.1 The Eddington Limit 118
7.6.2 Mass to energy efficiency 119
7.6.3 AGN Feedback 120
7.6.4 The relative role of different feedback processes 121

8 Galaxies interaction and triggering star formation 122

8.1 Collisions between stellar system 123
8.1.1 Stellar collisional cross section and relaxation time 123
8.1.2 Collisions in stellar systems 125
8.1.3 Dynamical friction 126

8.2 Interacting galaxies 127
8.2.1 ULIRGs as merging systems 133
8.2.2 Nuclear fuelling through stellar bars 133

9 Gravitational instabilities in the cosmological context 137

9.1 Starting point 138
9.2 Jeans Instability 139
9.2.1 Fluid equation in co-moving coordinates 139
9.2.2 Governing equations for the overdensity 142
9.2.3 The growth of instabilities 142
9.3 The need for dark matter 144
9.4 Perturbations with dark matter 144
9.5 Evolution of perturbations 145

10 Evolution of fluctuations, non-linear collapse and hierarchical structure formation 147

10.1 The power spectrum of the fluctuations 148
10.1.1 The two-point correlation function 148
10.1.2 The power-spectrum of the fluctuations 149
10.1.3 The initial Power-Spectrum 150
10.1.4 The Harisson-Zel'dovich Power Spectrum 151
10.1.5 Evolution of the Power Spectrum and transfer functions 151
10.2 Non-linear collapse of a spherical overdensity 155
10.2.1 Background Universe 156

10.2.2 Evolution of the overdensity 156
10.3 Application to the Milky Way 159
10.4 Hierarchical structure formation 159
10.4.1 The Press-Schechter Mass Function 160

11 Galaxy formation and evolution, star formation history of the Universe 163

11.1 The first objects 164
11.2 Baryonic gas in dark matter halos 166
11.2.1 Hydrostatic equilibrium 167
11.2.2 Halo growth 168
11.2.3 Gas cooling 168
11.2.4 Star Formation 170
11.3 The galaxy Luminosity Function and the galaxy populations 170
11.4 Feedback 172
11.5 Comments and Conclusion 174

Appendices 177

A Examples 1 178
B Examples 2 182

1. Introduction

> Earlier theories ... were based on the hypothesis that all the matter in the universe was created in one big bang at a particular time in the remote past.
>
> — Sir Fred Hoyle

This course aims to address *the formation of structure in the universe*, one of the most active topics in modern astrophysics. Over the past two decades, huge advances have been done in understanding how structures develop on both large and small scales with the arrival of large ground-based and space telescopes (e.g. *Hubble* Space Telescope, *Spitzer* Space Telescope, the Very Large Telescope, Keck Observatory, the Gran Telescopio Canarias, etc.). The arrival within the next 10 years of even larger facilities, such as the Extremely Large Telescope or SKA, will have strong impact on our understanding of these processes.
The current model on structure formation describes how initial density perturbations develop into proto-galaxies, and how these small galaxies merge to form larger galaxies over cosmic time. At a smaller scale, we know how star formation takes place in a group of stars and how the gaseous medium in galaxies (hereafter the Inter-Stellar Medium - ISM) is structured and evolved. However the link between the large and small scales is not well understood yet.

This course will build on the knowledge obtained from two previous courses :

- The Astrophysics and Cosmology major option from last term
- The Part II Astrophysical fluid dynamics

Several concepts of dynamics, electrodynamics, Special Relativity, and thermodynamics will be used throughout this course, and we will assume that you are familiar with the material from last term. We will also make use of some areas of physics we do not have time to cover in detail, but we will need to use results from them nevertheless. Key references will be given, but we do not expect you to look these up.

1.1 The first 3 minutes of the Universe

According to the current standard paradigm, the Big-Bang is the starting point of the history of the Universe (Figure 1.1). Shortly after the Universe undergoes a period of exponential inflation (from 10^{-36} s and 10^{-32} s after the Big-Bang) which predicts :

- a near spatially flat universe as a result of the rapid inflation with a total density close to the critical density (defined as the watershed point between an expanding and a contracting Universe).
- the exponential growth of the scale factor during inflation caused quantum fluctuations of the inflation field to be stretched to macroscopic scales and these fluctuation had the same amplitude on all physical scales.

After the inflation phase, the Universe's volume has increased by a factor of at least 10^{78}, it is mainly composed of quarks. The Universe continued to decrease in density and fall in temperature, hence the typical energy of each particle was decreasing. At about 1μs, quarks and gluons combined to form baryons such as protons and neutrons. At that time, protons and electrons are not bounded together, while nuclei of deuterium, helium and lithium are formed. Essentially all of the elements that are heavier than lithium form much later, by stellar nucleosynthesis in evolving and exploding stars.

Few minutes after the Big-Bang, the Universe is composed of :

- radiation (photons and neutrinos)
- baryonic matter (protons, neutrons, electrons, etc.)
- dark matter: non baryonic matter which interacts only weakly with ordinary matter (typically a human body will interact with only 1 dark matter particle over its lifetime), and is 'cold' (i.e. 'non relativistic').
- dark energy dominating the mass/energy budget

1.2 The formation of the first stars and galaxies

The Universe cools down as it is expanding. Over-dense regions grow from the initial perturbation and are dominated by dark matter. Eventually, when the universe is sufficiently cold (≈370 000 years after the Big-Bang according to the latest *Planck* results), matter and radiation are decoupled (transparency increases), and the electrons and protons start to be bound to form neutral hydrogen following :

$$e + p \rightarrow H + \gamma \tag{1.1}$$

Once photons decoupled from matter (also known as the recombination phase), they travelled freely through the universe without interacting with matter and constitute what is observed today as Cosmic Microwave Background radiation (hereafter CMB - Figure 1.2). The imprint of the initial structure of the Universe is left on the CMB. After the recombination, the Universe continues to expand and cool. The baryonic matter (mainly composed of hydrogen atoms) is neutral, and there is no source of light (e.g, stars) in the Universe : this epoch is called 'The Dark Ages'.
The over-dense regions continue to grow in the dark matter distribution, forming well defined dark matter potential wells. Their density becomes sufficiently large that their gravitational field is dominated by their own mass and they decouple from the Hubble flow[1]. We say that they are self-gravitating objects. Gravitational interactions lead to mergers of the dark-matter potential wells forming larger structures. Some baryonic mass falls into the dark matter potential wells, and at the centre of potential wells in high density region the gas cools leading to the formation of the first luminous objects.
Focusing now on the self-gravitating systems : baryonic matter forms a complex multi-phase system subject to thermal and gravitational instabilities. Within this complex gaseous phase, star formation proceeds on small scales against the back drop of cosmological evolution.
The heating and cooling of the gas in a gravitational field leads to a complex multi-phase medium. Cool clouds exist in dynamical equilibrium with hotter, low-density phases. The densest region can cool sufficiently fast that they are unstable to gravitational collapse, and collapse of these unstable regions within baryonic gas leads to fragmentation of the cold dense clouds. Then, proto-stars form as the collapsing gas heats up. Since the parent cloud has angular momentum, disks form in the protostellar regions. The final step of this evolution, is the formation of planetary systems from the cooling discs around stellar systems.

This course start with the understanding of the processes responsible for star

[1] The 'Hubble flow' describes the motion of galaxies due solely to the expansion of the Universe.

formation, and in particular the formation of stars within molecular clouds, which fragment and collapse as a consequence of gravitational instabilities and radiative cooling.

Important equations we will use in this course

- Euler's equation : in fluid mechanics, they are a set of quasilinear hyperbolic equations governing adiabatic and inviscid flow (i.e. $\nabla\rho = 0$), given by

$$\boxed{\rho\frac{\partial\vec{v}}{\partial t} + \rho(\vec{v}\cdot\nabla)\vec{v} = -\nabla P + \nabla\Phi_g} \tag{1.2}$$

  where $\vec{v}$ is the flow velocity, P the pressure and Φ_g the gravitational potential

- The Poisson's equation is an elliptic partial differential equation of broad utility in theoretical physics given by :

$$\boxed{\Delta\phi = \nabla^2\phi = f} \tag{1.3}$$

  In three-dimensional Cartesian coordinates, it takes the form :

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\Phi(x,y,z) = f(x,y,z) \tag{1.4}$$

  In the case of a gravitational potential, we can write :

$$\vec{g} = -\nabla\Phi_g \tag{1.5}$$

  and the Gauss law gives :

$$\nabla\cdot\vec{g} = -4\pi G\rho \tag{1.6}$$

  therefore :

$$\nabla(-\nabla\Phi_g) = -4\pi G\rho \tag{1.7}$$

  which gives the Poisson's equation for the gravitational potential :

$$\boxed{\nabla^2\Phi_g(\vec{r}) = 4\pi G\rho(\vec{r})} \tag{1.8}$$

- The equation of continuity is an equation that describes the transport of some quantity. It is particularly simple and powerful when applied to a conserved quantity, but it can be generalised to apply to any extensive quantity. Since mass, energy, momentum, electric charge and other natural quantities are conserved under their respective appropriate conditions, a variety of physical phenomena may be described using continuity equations. It is given by :

$$\boxed{\frac{\partial \rho}{\partial t} + \nabla \cdot \vec{j} = 0} \tag{1.9}$$

where ρ is the density in mass per unit volume and $\vec{j} = \rho \vec{v}$ is the flux.

- the equation of state for an ideal gas given by :

$$\boxed{P = \frac{\rho k_B T}{\mu}} \tag{1.10}$$

where ρ is the density, k_B is the Boltzmann constant, T is the temperature, and μ the mass of the particle.

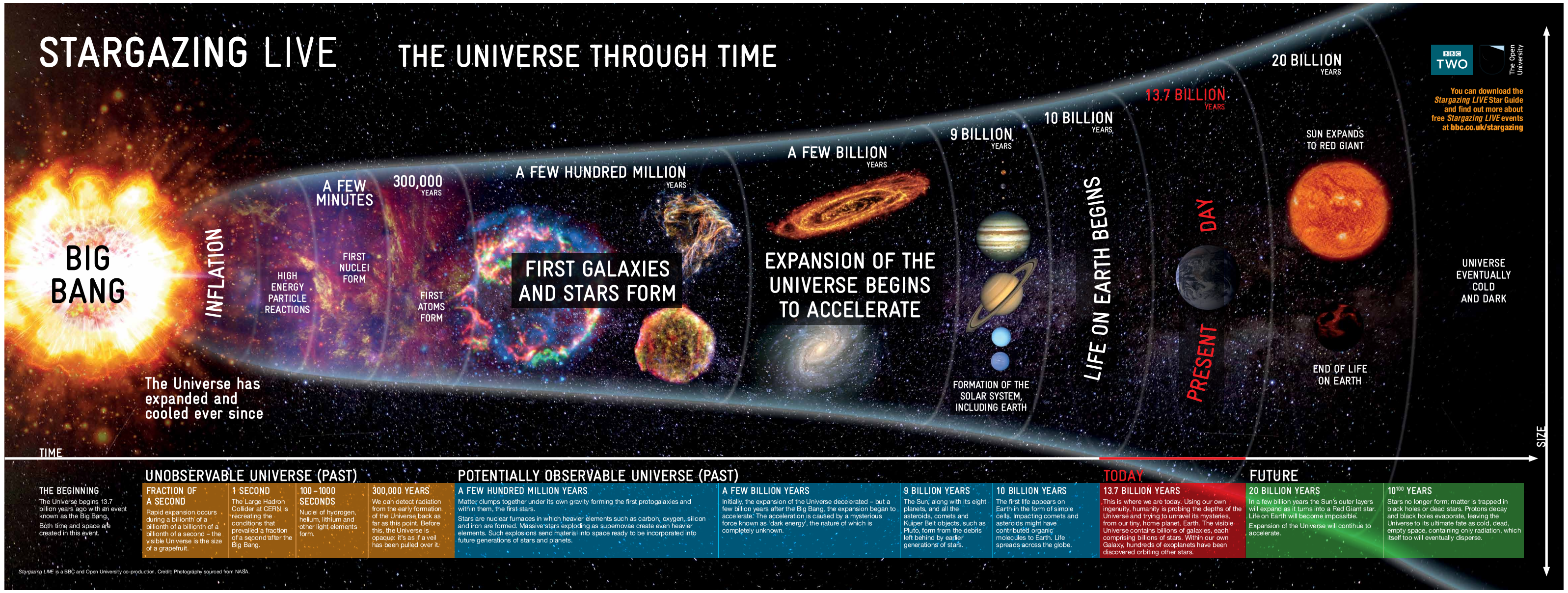

Figure 1.1 : The history of the Universe: (bottom) timeline of the history of the Universe with specific events, such as the inflation and the formation of neutral hydrogen. Source: BBC

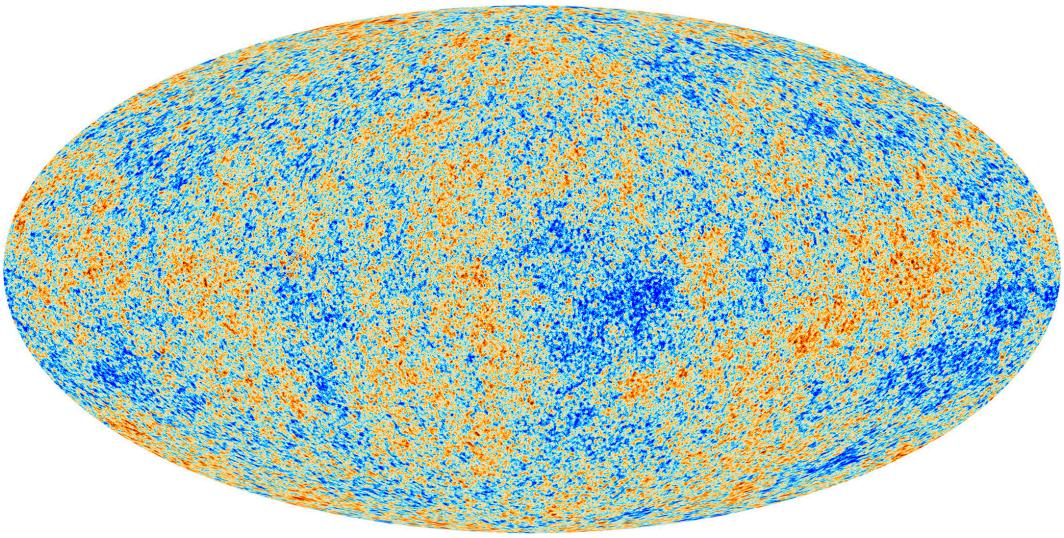

Figure 1.2 : The Cosmic Microwave Background observed by the Planck satellite. The colours trace the temperature of the radiation. It demonstrates the non-uniformity of the universe 370 000 years after the Big-Bang. Source : Planck collaboration

The Epoch of Reionisation

When the first stars formed in the Universe (the so-called 'Cosmic Dawn' epoch), their photons started to ionise the neutral hydrogen formed at the recombination phase. The neutral hydrogen atoms absorb CMB photons: observing the CMB at different epochs gives the evolution of the amount of neutral hydrogen as a function of time. According to *Planck* observations of the CMB, the ionisation of the neutral hydrogen is completed 1 billion years after the Big-Bang (Figure 1.4).

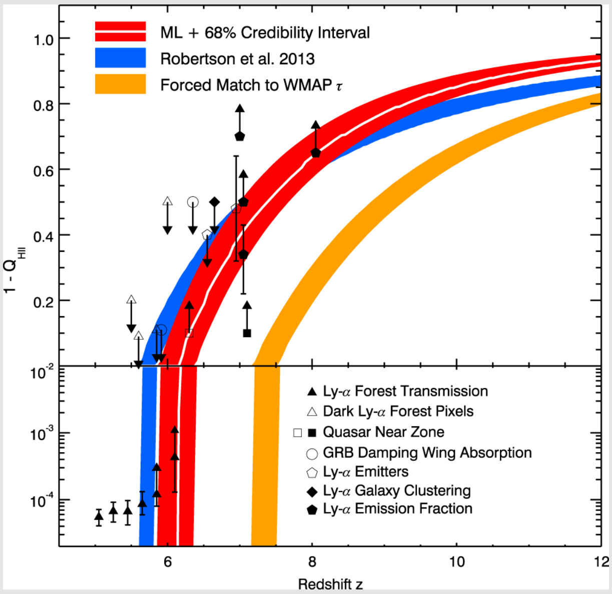

Figure 1.3 : Evolution of the neutral fraction of hydrogen (1-Q_{HII}) as a function of redshift. The data points show the constrain from galaxies/quasars observations. The coloured regions are from CMB observations with different assumptions (blue and red used *Planck* data and the yellow curve is based on WMAP data. Source : Robertson et al. 2015, ApJL, 802, 19

Studying the process of reionisation of the neutral hydrogen during the first billion years of the Universe is crucial to shape future evolution of structure. Three conclusions arise from these studies :

- first structure grows hierarchically by merger of small dark-matter halos with further infall of baryonic matter into the potential wells formed by the dark matter.
- These early galaxies evolve very quickly as large amount of dense gas (the fuel for star formation) is made available by this hierarchical build-up of self-gravitating objects
- an evolving distribution of galaxies of different masses is formed.

The Search for the Most distant galaxies

One of the most active topics of modern extragalactic astronomy is the search for the most distant galaxies, formed few million years after the Big-Bang. This quest for the first object illuminating the Universe for the first time started in the 1950s with the opening of 2m-class telescopes. At this time, the most distant galaxies was at a redshift of z =0.20 (11.7 billion years after the Big-Bang - Humason et al., 1956, AJ, 61, 97). In January 2023, the confirmed most distant galaxy (confirmed by spectrosocpy) is at z =13.27 (323 million years after the Big-Bang - Harikane et al. 2022, ApJ, 929, 1).

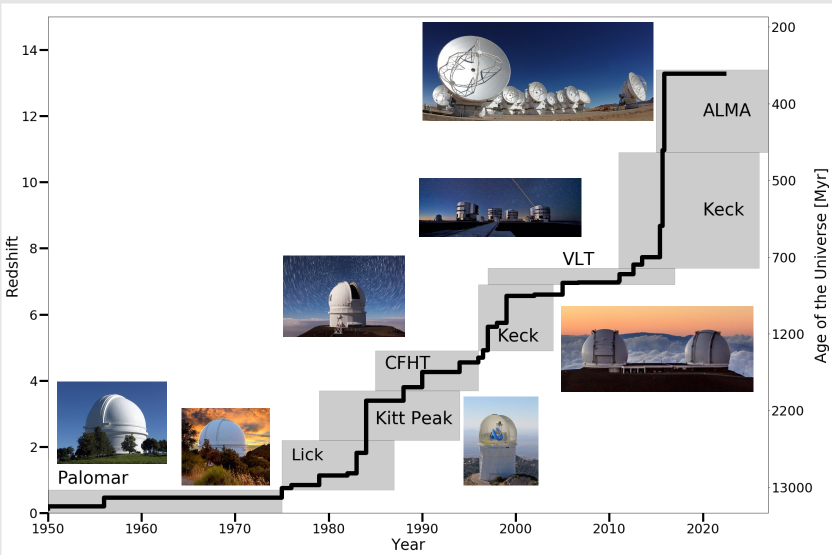

Figure 1.4 : Evolution of the redshift of the most distant galaxy known as a function of time. The name and picture of each telescope involved in the discovery is indicated. This evolution is exponential and follows the opening of new observatory.

In December 2021, from French Guyana, an Ariane V rocket launched in Space the *James Webb Space Telescope* (JWST), a space telescope with a 6.5m diameter mirror. After one month of travel to reach L2, the commissioning of the 4 instruments started. On July 13th 2022 the first images obtained with this telescope have been released (including one by the US president). In less than a week after this release, a dozen of galaxy candidates at z ≥13 has been announced, confirming the very high activity of this field of research. Spectroscopic follow-up of these candidates are now on the way, and new redshift record could happen shortly, even during this term. Stay tuned !

2. Physical processes in baryonic gas

> An experiment is a question which science poses to Nature and a measurement is the recording of Nature's answer.
>
> — Max Planck

The main goal of the following chapter is to discuss the physical mechanisms by which the baryonic gas is heated and cooled in the interstellar medium. We will therefore start by discussing the radiation processes in astrophysics before exploring in details the heating and cooling mechanisms of astrophysical gas and thermal stability.

2.1 Radiation processes in Astrophysics

When the scale of a system greatly exceeds the wavelength of radiation (e.g., light shining through a keyhole), we can consider radiation to travel in straight lines. One of the most primitive concepts is that of *energy flux* : consider an element of area A exposed to radiation for a time $\mathrm{d}t$. The amount of energy passing through the element should be proportional to $\mathrm{d}A \times \mathrm{d}t$, and we write it as :

$$\mathrm{d}E = \mathrm{d}F \times \mathrm{d}A \times \mathrm{d}t. \tag{2.1}$$

The energy flux F is usually measured in erg s^{-1} cm^{-2}.
A source of radiation is called *isotropic* if it emits energy equally in all directions. An example would be a spherically symmetric, isolated star. If we put imaginary spherical surfaces S_1 and S at radii r_1 and r, respectively, we know by conservation of energy that the total energy passing through S_1 must be the same as that passing through S. Thus :

$$F(r_1)4\pi r_1^2 = F(r)4\pi r^2 \tag{2.2}$$

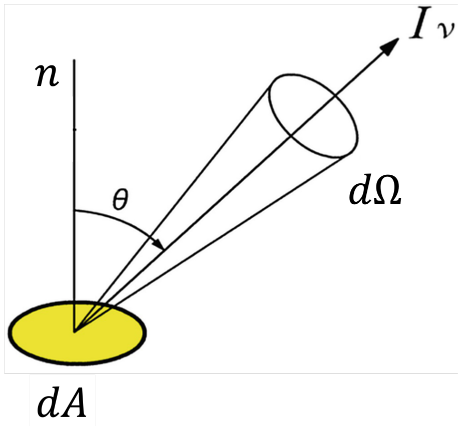

Figure 2.1 : Geometry for obliquely incident rays

or

$$F(r) = \frac{F(r_1)r_1^2}{r^2} \tag{2.3}$$

and then, if we regard the sphere S_1 as fixed, then :

$$F(r) = \frac{\text{constant}}{r^2} \tag{2.4}$$

2.1.1 Description of a radiation field

The flux is a measure of the energy carried by all rays passing through a given area. If we consider an area dA normal to the direction of a given ray, and if we consider also all rays passing through dA whose direction is within a solid angle dΩ of the given ray, then the energy crossing dA in a time dt and in a frequency range dν is then defined by the relation :

$$\mathrm{d}E = I_\nu \mathrm{d}A\,\mathrm{d}t\,\mathrm{d}\Omega\mathrm{d}\nu \tag{2.5}$$

where I_ν is the **specific intensity** or brightness. It has a dimension of :

$$I_\nu(\nu,\Omega) = \text{energy(time)}^{-1}\text{(area)}^{-1}\text{(solid angle)}^{-1}\text{(frequency)}^{-1} \tag{2.6}$$

$$= \text{erg s}^{-1}\,\text{cm}^{-2}\,\text{sr}^{-1}\,\text{Hz}^{-1} \tag{2.7}$$

Suppose now that we have a radiation field (rays in all direction) and construct a small element of area dA at some arbitrary orientation $\vec{n}$ (Figure 2.1). Then the differential amount of flux from the solid angle dΩ is :

$$\mathrm{d}F_\nu = I_\nu \cos\theta\,\mathrm{d}\Omega \tag{2.8}$$

The net energy flux is obtained by integrating dF over all solid angles :

$$F_\nu = \int I_\nu \cos\theta \, \mathrm{d}\Omega \tag{2.9}$$

Note that if I_ν is an isotropic radiation field (not a function of angle), then the net energy flux is 0 since $\int \cos\theta \, \mathrm{d}\Omega = 0$

The specific energy density u_ν is defined as the energy per unit volume per unit frequency range. To determine this it is convenient to consider first the energy density per unit of solid angle $u_\nu(\Omega)$ by :

$$\mathrm{d}E = u_\nu(\Omega) \, \mathrm{d}V \, \mathrm{d}\Omega \, \mathrm{d}\nu \tag{2.10}$$

where dV is a volume element.
Consider a cylinder about a ray of length (i.e. ct), since the volume of the cylinder is $\mathrm{d}A \times c \times \mathrm{d}t$,

$$\mathrm{d}E = u_\nu(\Omega) \, \mathrm{d}A \, c\mathrm{d}t \, \mathrm{d}\Omega \, \mathrm{d}\nu \tag{2.11}$$

but since radiation travels at velocity c, within dt all the radiation in the cylinder will pass out of it :

$$dE = I_\nu \, \mathrm{d}A \, \mathrm{d}\Omega \, \mathrm{d}t \, \mathrm{d}\nu \tag{2.12}$$

similar to eq. 2.5.
Equating eq. 2.5 and eq. 2.11 yields :

$$u_\nu(\Omega) = \frac{I_\nu}{c} \tag{2.13}$$

Integrating over solid angles gives :

$$u_\nu = \int u_\nu(\Omega) \, \mathrm{d}\Omega = \frac{1}{c} \int I_\nu \, \mathrm{d}\Omega \tag{2.14}$$

We also need to define the *mean density* as :

$$J_\nu = \frac{1}{4\pi} \int I_\nu \, \mathrm{d}\Omega \tag{2.15}$$

Therefore, the energy density can be simplified as :

$$u_\nu = \frac{4\pi}{c} J_\nu \tag{2.16}$$

Finally the total radiation density (in erg cm^{-3}) is simply obtained by integrating u_ν over all frequencies :

$$u = \int u_\nu \, \mathrm{d}\nu = \frac{4\pi}{c} \int J_\nu \, \mathrm{d}\nu \tag{2.17}$$

2.1.2 Radiative Transfer

If a ray passes through a matter, energy may be added or subtracted from it by emission or absorption, and the specific intensity will not in general remain constant.

Emission

The spontaneous *emission coefficient* j is defined as the energy emitted per unit time per unit solid angle and per unit volume :

$$dE = j\,\mathrm{d}V\,\mathrm{d}\Omega\,\mathrm{d}t \tag{2.18}$$

A monochromatic emission coefficient can be similarly defined so that :

$$\mathrm{d}E = j_\nu\,\mathrm{d}V\,\mathrm{d}\Omega\,\mathrm{d}t\,\mathrm{d}\nu \tag{2.19}$$

where j_ν has units of erg cm^{-3} s^{-1} sr^{-1} Hz^{-1}.
In general, the emission coefficient depends on the direction into which emission takes place. For an *isotropic* emitter, or for a distribution of randomly oriented emitters, we can write :

$$j_\nu = \frac{1}{4\pi} P_\nu \tag{2.20}$$

where P_ν is the radiated power per unit volume per unit frequency.
Sometimes the spontaneous emission is defined by the *emissivity* ϵ_ν, defined as the energy emitted spontaneously per unit frequency per unit time per unit mass, with units of erg g^{-1} s^{-1} Hz^{-1}, if the emission is isotropic then :

$$\mathrm{d}E = \epsilon_\nu \rho\,\mathrm{d}V\,\mathrm{d}t\,\mathrm{d}\nu \frac{\mathrm{d}\Omega}{4\pi} \tag{2.21}$$

where ρ is the mass density of the emitting medium.
Comparing eq.2.19 and eq.2.21, we have the relation between ϵ_ν and j_ν :

$$j_\nu = \frac{\epsilon_\nu \rho}{4\pi} \tag{2.22}$$

holding for isotropic emission.
In going a distance ds, a beam of cross section $\mathrm{d}A$ travels through a volume $\mathrm{d}V = \mathrm{d}A \times \mathrm{d}s$. Thus the intensity added to the beam by spontaneous emission is :

$$\mathrm{d}I_\nu = j_\nu \mathrm{d}s \tag{2.23}$$

Absorption

By definition, the *absorption coefficient*, $\alpha_\nu(\mathrm{cm}^{-1})$, represents the loss of intensity in a beam as it travels a distance $\mathrm{d}s$[1] :

$$\mathrm{d}I_\nu = -\alpha_\nu I_\nu \,\mathrm{d}s \tag{2.24}$$

This phenomenological law can be understood in terms of microscopic model in which particles with density n (number per unit volume) each present an effective absorbing area, or *cross section*, of magnitude σ_ν (cm^2). These absorbers are assumed to be distributed at random. Let us consider the effect of these absorbers on radiation through $\mathrm{d}A$ within solid angle $\mathrm{d}\Omega$. The number of absorbers in the element equals $n \times \mathrm{d}A \times \mathrm{d}s$. The total absorbing area presented by absorbers equals $n \times \sigma_\nu \times dA \times ds$. The energy absorbed out of the beam is :

$$-\,\mathrm{d}I_\nu \,\mathrm{d}A \,\mathrm{d}\Omega \,\mathrm{d}t \,\mathrm{d}\nu = I_\nu(n\sigma_\nu \,\mathrm{d}A \,\mathrm{d}s) \,\mathrm{d}\Omega \,\mathrm{d}t \,\mathrm{d}\nu \tag{2.25}$$

thus

$$\mathrm{d}I_\nu = -n\sigma_\nu I_\nu \,\mathrm{d}s \tag{2.26}$$

which is precisely the above phenomenological law where :

$$\alpha_\nu = n\sigma_\nu \tag{2.27}$$

Often α_ν is written as :

$$\alpha_\nu = \rho\kappa_\nu \tag{2.28}$$

where ρ is the mass density and $\kappa_\nu(\mathrm{cm}^2\ \mathrm{g}^{-1})$ is called the *mass absorption coefficient*, κ_ν is also sometimes called the *opacity* coefficient.

The radiative transfer equation

We can now incorporate the effects of emission and absorption into a single equation giving the variation of specific intensity along a ray. From the above expressions for emission and absorption, we have the combined expression :

$$\boxed{\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\alpha_\nu I_\nu + j_\nu} \tag{2.29}$$

The transfer equation provides a useful formalism within which to solve for the intensity in an emitting and absorbing medium. Once α_ν and j_ν are known it is relatively easy to solve the transfer equation for the specific intensity. When scattering is present, solution of the radiative transfer equation is more difficult, because emission into $\mathrm{d}\Omega$ depends on I_ν in solid angles $\mathrm{d}\Omega'$ integrated over the latter (scattering from $\mathrm{d}\Omega$ into $\mathrm{d}\Omega$').
Here we can give solutions to two simple limiting cases :

[1] α_ν is positive for energy taken out of a beam

1. **Emission Only :** α_ν=0. In this case we have :

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = j_\nu, \tag{2.30}$$

which has the solution :

$$I_\nu(s) = I_\nu(s_0) + \int_{s_0}^{s} j_\nu(s')\,\mathrm{d}s' \tag{2.31}$$

The increase brightness is equal to the emission coefficient integrated along the line of sight

2. **Absorption Only :** j_ν=0. In this case we have :

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\alpha_\nu I_\nu, \tag{2.32}$$

which has the solution :

$$I_\nu(s) = I_\nu(s_0)\exp\left[-\int_{s_0}^{s} \alpha_\nu(s')\,\mathrm{d}s'\right] \tag{2.33}$$

The brightness decreases along the ray by the exponential of the absorption coefficient integrated along the line of sight.

Optical depth and Source Function

Here it is useful to define the *optical depth* τ_ν which is a measure of the transparency of a medium. It is defined as :

$$\mathrm{d}\tau_\nu = \alpha_\nu\,\mathrm{d}s \tag{2.34}$$

or :

$$\tau_\nu(s) = \int_{s_0}^{s} \alpha_\nu(s')\,\mathrm{d}s' \tag{2.35}$$

A medium is said to be *optically thick* or *opaque* when τ_ν, integrated along a typical path through the medium, satisfies $\tau_\nu \geq$1. When $\tau_\nu \leq$1, the medium is said to be *optically thin* or *transparent.* In other words, *an optically thin* medium is one in which the typical photons of frequency ν can traverse the medium without being absorbed, whereas an optically thick medium is one in which the average photon of frequency ν cannot traverse the entire medium without being absorbed.
The transfer equation can now be rewritten as :

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu} = -I_\nu + S_\nu \tag{2.36}$$

where S_ν is the source function, defined as the ratio of the emission coefficient to the absorption coefficient :

$$S_\nu = \frac{j_\nu}{\alpha_\nu} \tag{2.37}$$

We can now solve the equation of radiative transfer by regarding all quantities as functions of the optical depth τ_ν instead of s, multiplying the equation by the integrating factor e^{τ_ν} :

$$\frac{\mathrm{d}}{\mathrm{d}\tau_\nu}(I_\nu e^{\tau_\nu}) = S_\nu e^{\tau_\nu} \tag{2.38}$$

then the formal solution of the previous equation is :

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} + \int_0^{\tau_\nu} e^{-(\tau_\nu-\tau'_\nu)} S_\nu(\tau'_\nu)\,\mathrm{d}\tau'_\nu \tag{2.39}$$

Since τ_ν is just the dimensionless $e-$folding factor for absorption, the above equation is easily interpreted as the sum of two terms : the initial intensity diminished by absorption plus the integrated source diminished by absorption.
As an example consider a *constant* source function S_ν, then eq.2.39 gives the solution :

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} + S_\nu(1-e^{-\tau_\nu}) \tag{2.40}$$

$$= S_\nu + e^{-\tau_\nu}(I_\nu(0) - S_\nu) \tag{2.41}$$

Mean Free Path

A useful concept, which describes absorption in an equivalent way, is that of the *mean free path* of radiation (or photons). This is defined as the average distance a photon can travel through an absorbing material without being absorbed. It may be easily related to the absorption coefficient of a homogeneous material. Consider the photon mean free path as it tries to escape from an emitting region, when $\tau_\nu = 1$ then $s\alpha_\nu = 1$, or :

$$s = \frac{1}{\alpha_\nu} \tag{2.42}$$

$$= \frac{1}{n\sigma_\nu} \tag{2.43}$$

This distance is the definition of *the mean free path* l_ν.
A photon escaping from a region with an optical depth τ_ν will undergo a random walk with N scatterings. For a region of size L :

$$L = \sqrt{N l_\nu} \Rightarrow N \sim \frac{L^2}{l_\nu^2} \sim (\alpha_\nu L)^2 = \tau_\nu^2 \tag{2.44}$$

2.1.3 Thermal Radiation

By definition, thermal radiation is radiation emitted by matter in thermal equilibrium. The best example of thermal radiation is the *blackbody radiation*. To obtain such radiation we keep an enclosure at temperature T and do not let radiation in or out until equilibrium has been achieved. Using some general thermodynamic arguments plus the fact that photons are massless, we can derive several important properties of blackbody radiation.

An important property of I_ν is that it is independent of the properties of the enclosure and depends only on the temperature :

$$I_\nu = B_\nu(T) \tag{2.45}$$

where $B_\nu(T)$ is the *Planck function*. Its form is discussed below.

Now consider an element of some thermally emitting material at temperature T, so that its emission depends solely on its temperature and internal properties. Put this into the opening of a blackbody enclosure at the same temperature T. Let the source function be S_ν. If $S_\nu > B_\nu$, then $I_\nu > B_\nu$, and if $S_\nu < B_\nu$, then $I_\nu < B_\nu$ (see eq.2.43). But the presence of the material cannot alter the radiation, since the new configuration is also a blackbody enclosure at temperature T. Thus we have the two following relations :

$$S_\nu = B_\nu(T) \tag{2.46}$$

$$j_\nu = \alpha_\nu B_\nu(T) \tag{2.47}$$

The transfer radiation for thermal radiation can be rewritten as :

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\alpha_\nu I_\nu + \alpha_\nu B_\nu(T) \tag{2.48}$$

or

$$\boxed{\frac{dI_\nu}{d\tau_\nu} = -I_\nu + B_\nu(T)} \tag{2.49}$$

The *Planck function* is usually defined as :

$$B_\nu(T) = \frac{2h\nu^3/c^2}{\exp\left(h\nu/kT\right) - 1} \tag{2.50}$$

At this point, it is well to draw the distinction between *blackbody radiation*, where $I_\nu = B_\nu$, and *thermal radiation* where $S_\nu = B_\nu$. Thermal radiation becomes a blackbody radiation only for optically thick media.

Blackbody radiation, like any system in the thermodynamic equilibrium can be treated by thermodynamic methods. Let us make a blackbody enclosure with a

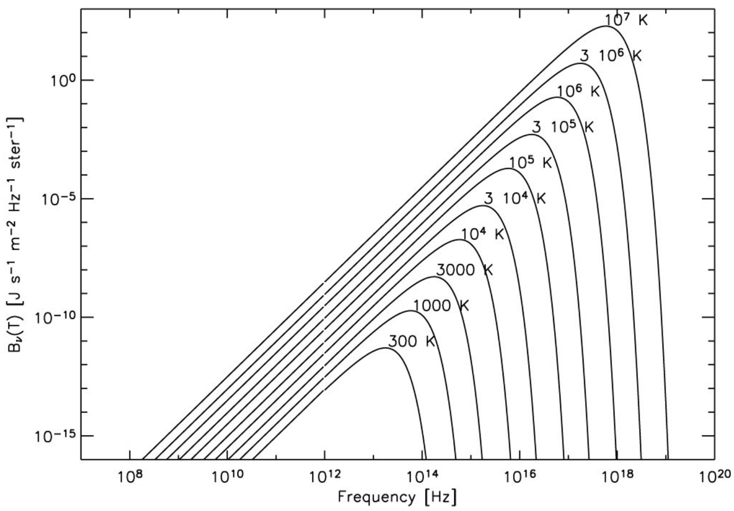

Figure 2.2 : Spectrum of blackbody radiation at various temperature (from Kraus J.D. 1966)

piston, so that work may be done on or extracted from the radiation. Now, by the first law of thermodynamics, we have :

$$\mathrm{d}U = \mathrm{d}Q - P\mathrm{d}V \tag{2.51}$$

where Q is the heat and U is the total energy.
By the second law of thermodynamics, we have :

$$\mathrm{d}S = \frac{\mathrm{d}Q}{T} \tag{2.52}$$

where S is the entropy.
But $U = uV$, and $P = \frac{u}{3}$, and u depends only on the temperature T since $u = (4\pi/c)\int J_\nu \mathrm{d}\nu$ and $J_\nu = B_\nu(T)$, thus we have :

$$\mathrm{d}S = \frac{V}{T}\left(\frac{\mathrm{d}u}{\mathrm{d}T}\right)\mathrm{d}T + \frac{u}{T}\mathrm{d}V + \frac{1}{3}\frac{u}{T}\mathrm{d}V \tag{2.53}$$

$$= \frac{V}{T}\left(\frac{\mathrm{d}u}{\mathrm{d}T}\right)\mathrm{d}T + \frac{4u}{3T}\mathrm{d}V \tag{2.54}$$

Since dS is a perfect differential,

$$\left(\frac{\partial S}{\partial T}\right)_V = \frac{V}{T}\frac{\mathrm{d}u}{\mathrm{d}T} \qquad \left(\frac{\partial S}{\partial V}\right)_T = \frac{4u}{3T} \tag{2.55}$$

Thus we obtain :

$$\frac{\partial^2 S}{\partial T \partial V} = \frac{1}{T}\frac{du}{dT} = -\frac{4u}{3T^2} + \frac{4}{3T}\frac{du}{dT} \tag{2.56}$$

so that

$$\frac{du}{dT} = \frac{4u}{T}, \qquad \frac{du}{u} = 4\frac{dT}{T} \tag{2.57}$$

$$\log u = 4\log T + \log a \tag{2.58}$$

where $\log a$ is a constant of integration. Thus we obtain the *Stefan-Boltzmann* law:

$$\boxed{u(T) = aT^4} \tag{2.59}$$

This may be related to the Planck function, since $I_\nu = J_\nu$ for isotropic radiation (see eq.2.16) :

$$u = \frac{4\pi}{c}\int B_\nu(T)\,\mathrm{d}\nu = \frac{4\pi}{c}B(T) \tag{2.60}$$

where the integrated Planck function is defined by :

$$B(T) = \int B_\nu(T)\,\mathrm{d}\nu = \frac{ac}{4\pi}T^4 \tag{2.61}$$

The emergent flux from an isotropically emitting surface (such as a blackbody) is $\pi\times$ brightness, so that :

$$F = \int F_\nu \mathrm{d}\nu = \pi\int B_\nu \mathrm{d}\nu = \pi B(T) \tag{2.62}$$

This leads to another form of the *Stefan-Boltzmann* law :

$$F = \sigma T^4 \tag{2.63}$$

where :

$$\sigma = \frac{ac}{4} = 5.67\times 10^{-5}\text{erg cm}^{-2}\text{deg}^{-4}\text{s}^{-}1, \tag{2.64}$$

$$\alpha = \frac{4\sigma}{c} = 7.56\times 10^{-15}\text{erg cm}^{-3}\text{deg}^{-4} \tag{2.65}$$

2.2 Line emission

2.2.1 The Einstein coefficients

The Kirchoff's law, $j_\nu = \alpha_\nu B_\nu$, relating emission to absorption for a thermal emitter, clearly must imply some relationship between emission and absorption at a microscopic level. This relationship was first discovered by Einstein in a beautifully simple analysis of the interaction of radiation with an atomic system.

He considered the simple case of two discrete energy levels : the first of energy E with statistical weight g_1, the second of energy $E + h\nu_0$ with statistical weight g_2. The system makes a transition from 1 to 2 by absorption of a photon of energy $h\nu_0$. Similarly, a transition from 2 to 1 occurs when a photon is emitted. Einstein identified three processes :

1. **Spontaneous Emission:** This occurs when the system is in level 2 and drops to level 1 by emitting a photon, and it occurs even in the absence of a radiation field. We define the *Einstein A-coefficient* by :

$$A_{21} = \text{transition probability per unit time for spontaneous emission } (s^{-1})$$

2. **Absorption :** This occurs in the presence of photons of energy $h\nu_0$. This system makes a transition from level 1 to level 2 by absorbing a photon. Since there is no self-interaction of the radiation field, we expect that the probability per unit time for this process will be proportional to the density of photons (or to the mean intensity) at frequency ν_0. To be precise, we must recognise that the energy difference between the two levels is not infinitely sharp but is described by a *line profile function* $\phi(\nu)$, which is sharply peaked at $\nu = \nu_0$ and which is conveniently taken to be normalized :

$$\int_0^\infty \phi(\nu)\, \mathrm{d}\nu = 1 \tag{2.66}$$

This line profile function describes the relative effectiveness of frequencies in the neighbourhood of ν_0 for causing transitions. These arguments leads us to write :

$$B_{12}\overline{J} = \text{transition probability per unit time for absorption, where}$$

$$\overline{J} = \int_0^\infty J_\nu \phi(\nu)\, \mathrm{d}\nu \tag{2.67}$$

The proportionality constant B_{12} is the *Einstein B-coefficient*

3. **Stimulated Emission :** Einstein found that to derive Planck's law another process was required that was proportional to $\overline{J}$ and caused *emission* of photon. As before we define :

$$B_{21}\overline{J} = \text{transition probability per unit time for stimulated emission}$$

B_{21} is another *Einstein B-coefficient*

In thermodynamics equilibrium, we have that the number of transitions per unit time per unit volume out of state l equal the number of transitions per unit time per unit volume into state l. If we let n_1 and n_2 be the number densities of atoms of level 1 and 2 respectively, this reduces to :

$$n_1 B_{12} \overline{J} = n_2 A_{21} + n_2 B_{21} \overline{J} \tag{2.68}$$

Now solving for $\overline{J}$:

$$\overline{J} = \frac{A_{21}/B_{21}}{(n_1/n_2)(B_{12}/B_{21}) - 1} \tag{2.69}$$

In thermodynamics equilibrium, the ratio of n_1 to n_2 is :

$$\frac{n_1}{n_2} = \frac{g_1 \exp(-E/kT)}{g_2 \exp[-(E + h\nu_0)/k_B T]} = \frac{g_1}{g_2} \exp(h\nu_0/k_B T) \tag{2.70}$$

so that :

$$\overline{J} = \frac{A_{21}/B_{21}}{(g_1 B_{12}/g_2 B_{21}) \exp(h\nu_0/kT) - 1} \tag{2.71}$$

But in the thermodynamics equilibrium we also know $J_\nu = B_\nu$ and the fact that B_ν varies slowly on the scale of $\Delta\nu$ implies that $\overline{J} = B_\nu$. For the expression in eq.2.71 to equal the Planck function for all temperatures, we must have the following *Einstein relation* :

$$g_1 B_{12} = g_2 B_{21} \tag{2.72}$$

$$A_{21} = \frac{2h\nu^3}{c^2} B_{21} \tag{2.73}$$

2.2.2 Absorption and Emission coefficients in terms of Einstein coefficients

To obtain the emission coefficient j_ν we must make some assumption about the frequency distribution of the emitted radiation during a spontaneous transition from level 2 to 1. The simplest assumption is that this emission is distributed in accordance with the same line profile function $\phi(\nu)$ that describes absorption. The amount of energy emitted in volume $\mathrm{d}V$, solid angle $\mathrm{d}\Omega$, frequency range $\mathrm{d}\nu$, and time $\mathrm{d}t$ is, by definition, $j_\nu \, \mathrm{d}V \, \mathrm{d}\Omega \, \mathrm{d}\nu \, \mathrm{d}t$. Since each atom contributes an energy $h\nu_0$ distributed over a 4π solid angle for each transition, this may also be expressed as $(h\nu_0/4\pi)\phi(\nu) n_2 A_{21} \, \mathrm{d}V \, \mathrm{d}\Omega \, \mathrm{d}\nu \, \mathrm{d}t$, so that the emission coefficient is :

$$j_\nu = \frac{h\nu_0}{4\pi} n_2 A_{21} \phi(\nu) \tag{2.74}$$

To obtain the absorption coefficient, we first note that the total energy absorbed in time dt and volume dV is :

$$\mathrm{d}V\,\mathrm{d}t\,\frac{h\nu_0}{4\pi}n_1 B_{12}\int \mathrm{d}\Omega \int \mathrm{d}\nu\phi(\nu)I_\nu \tag{2.75}$$

Therefore the energy absorbed out of a beam in frequency range dν solid angle dΩ time dt and volume dV is :

$$\mathrm{d}V\,\mathrm{d}t\,\mathrm{d}\Omega\,\mathrm{d}\nu\frac{h\nu_0}{4\pi}n_1 B_{12}\phi(\nu)I_\nu \tag{2.76}$$

Assuming the volume element is dV = d$A \times$ ds, the absorption coefficient is given by :

$$\alpha_\nu = \frac{h\nu}{4\pi}n_1 B_{12}\phi(\nu) \tag{2.77}$$

What about the stimulated emission ? At first sight one might be tempted to add this as a contribution to the emission coefficient ; but notice that it is proportional to the intensity and only affects the photons along the given beam, in close analogy to the process of absorption. Thus it is much more convenient to treat stimulated emission as *negative absorption* and include its effects through the absorption coefficient. In operational terms these two processes always occur together and cannot be disentangled by experiments. By reasoning entirely analogous to that leading to the previous equation, we can find the contribution of stimulated emission to the absorption coefficient. The result for the absorption coefficient, corrected for stimulated emission is :

$$\alpha_\nu = \frac{h\nu_0}{4\pi}\phi(\nu)(n_1 B_{12} - n_2 B_{21}) \tag{2.78}$$

As atoms collide, their electrons can be knocked up to the next higher energy level if the colliding atoms have enough energy (collisional excitation). The electron can even be knocked entirely away from the atom (ionization). Looking at the problem from a statistical standpoint, the probability of the atom's being in one energy state, s_a, is

$$P(s_a) \sim \exp(-E_a/kT) \tag{2.79}$$

and the probability for state s_b is :

$$P(s_b) \sim \exp(-E_b/kT) \tag{2.80}$$

where E_a and E_b are the energies of the two states. The ratio of these probabilities is then :

$$\frac{P(s_b)}{P(s_a)} = \frac{\exp(-E_b/kT)}{\exp(-E_a/kT)} \tag{2.81}$$

However, more than one state in an atom can have the same energy (i.e. they can be degenerate), e.g. in He I (neutral helium) two electrons in the ground state have $n = 1$, but the two electrons spin in opposite directions (m_s = +1/2, -1/2). We define the degeneracy (the number of states with the same energy E_a) as g_a, which is called the statistical weight of state a. The above expression then becomes :

$$\frac{P(s_b)}{P(s_a)} = \frac{g_b \exp(-E_b/k_B T)}{g_a \exp(-E_a/k_B T)} \tag{2.82}$$

For a large number of atoms, the ratio of probabilities must be the same as the ratio of numbers of atoms in the two energy levels, e.g. :

$$\boxed{\frac{N_b}{N_a} = \frac{g_b}{g_a} \exp[-(E_b - E_a)/k_B T]} \tag{2.83}$$

This is the Boltzmann Equation.
Therefore another important transition rates to take into account is the collisional processes leading excitation or de-excitation $n_0 C_{21}$ and $n_0 C_{12}$ respectively, where n_0 is the density of colliding particles. When the gas is ionised, it is often the case that the collisions are dominated by electron-ion collisions in which case $n_0 \approx n_e$
If the gas is in thermodynamic equilibrium at a temperature T then detailed balance requires :

$$C_{12} = \frac{g_2}{g_1} C_{21} \exp(-[E_1 - E_2]/kT) \tag{2.84}$$

2.2.3 Excitation of lines by collisions

This is a crucial cooling processes in astrophysical gas whereby thermal energy can be dissipated via radiation. Assuming the previous two-level system : the number of transitions from $1 \rightarrow 2$ must equal the number of transition from $2 \rightarrow 1$, including the collisional effects, we can now rewrite eq.2.68 :

$$n_1(n_0 C_{12} + B_{12}\overline{J}) = n_2(A_{21} + B_{21}\overline{J} + n_0 C_{21}) \tag{2.85}$$

N.B.: the $n_1 n_0$ terms before the collisional coefficient takes into account the number of electron n_0 and the number density of atoms in state 1 and 2, n_1 and n_2 respectively.
We can easily make the assumption that induced processes are much less important than spontaneous transitions and collisions, therefore :

$$n_1 n_0 C_{12} = n_2(A_{21} + n_0 C_{21}) \tag{2.86}$$

and

$$\frac{n_2}{n_1} = \frac{n_0 C_{12}}{A_{21}} \frac{1}{1 + \dfrac{n_0 C_{21}}{A_{21}}} \tag{2.87}$$

The line emissivity corresponds to the amount of energy emitted by the total number of atoms :

$$\epsilon = n_2 A_{21} h\nu_{21} = n_0 n_1 C_{12} h\nu_{21} \frac{1}{1 + \frac{n_0 C_{21}}{A_{21}}} \tag{2.88}$$

We can now estimate easily the emissivity of the line in two simple regime :

1. **Low density limit :** in that case $n_0 C_{21} \ll A_{21}$:

$$\epsilon \approx n_0 n_1 C_{12} h\nu_{21} \tag{2.89}$$

   which means that every upwards transition due to a collision rise to a downward radiative transition.

2. **High density limit :** where $n_0 C_{21} \gg A_{21}$

$$\epsilon \approx n1 \frac{g_2}{g_1} e^{-h\nu_{21}/k_B T} A_{21} h\nu_{21} \tag{2.90}$$

   In that case, the emissivity of the line is determined by the conditions of thermal equilibrium of the excited state ; not all the downward transitions are now associated with photon emission, instead many downward transitions are caused by collisions and we say that *the line is collisionally de-excited.*

There exists a *critical density* above which the line is predominantly collisionally de-excited (collisional de-excitation rate higher than spontaneous emission rate) and for the two level example this is : $n_{0c} \approx A_{21}/C_{21}$.
Note that generally : $n_0 \sim n_e \propto n_{\rm atoms} \propto n$, where n is the total gas density. Also, $n_1, n_2 \propto n_{\rm atoms} \propto n$. Therefore the emissivity expression becomes :

$$\text{at } n \ll n_{0c} : \epsilon_\nu \propto n^2 \tag{2.91}$$

$$\text{at } n \gg n_{0c} : \epsilon_\nu \propto n \tag{2.92}$$

$$\tag{2.93}$$

N.B.: For systems in which there are more than two energy states which need to be considered then the analysis is more complicated, but the same ideas apply and an expression determined fro the critical density at which a given quantum state becomes collisionally de-excited.

2.2.4 Astrophysics terminology

In Astrophysics, the timescales are so long that we may detect emission lines which are not seen in laboratory (also called *forbidden lines*) because gases can not be

	Timescale	Nature of line	Example
Dipole	short	permitted	Haλ6563
Quadrupole	long	forbidden	[OIII]λ 5007
Intercombination	Intermediate	semi-forbidden	CIII]λ1909

Table 2.1: Notation of several types of emission lines in astrophysics. In the last column: the letter gives the element, the roman numerical indicate the ionisation state (I : neutral ; II : singly ionised, etc...), the wavelength is in Angstroms, the square brackets indicate the nature of transition.

rarefied enough. The term forbidden is misleading; a more accurate description would be "highly improbable." The emissions result from electrons in long-lived orbits within the radiating atoms—i.e., the transition from an upper energy level to a lower energy level that produces the emissions requires a long time to take place. As a result, emission lines corresponding to such atomic transitions are extremely weak compared with other lines. In the laboratory, moreover, an excited atom tends to strike another particle or the walls of the gas container before it emits a photon, thereby further reducing the possibility of observation. There is a well-defined astrophysical notation used to label these which emphasises the lifetimes of the excited states (see Table2.1).

2.2.5 An example : the Hydrogen atom

Hydrogen is the most important element in the Universe : it is the first atom formed after the Big-Bang (see previous chapter) and it is responsible for a vast majority of light in the Universe. The electronic energy states are determined by :

$$\Delta E_{mn} = R\left(\frac{1}{n^2} - \frac{1}{m^2}\right) \tag{2.94}$$

where R is the Rydberg constant (R =1.09×10^7 m^{-1}). Figure 2.3 shows several transitions with the wavelength of the emitted photon.

Young stars are usually UV emitters and emit radiation at wavelength below the Lyman edge (λ=91.1753 nm). Therefore, the hydrogen in the surrounding of young stars is usually ionised. Extended regions of photoionised hydrogen are produced in the vicinity of young stars, also called HII regions (it is important to note that these regions are not only composed of Hydrogen, they are also hosting metals in various ionisation stages). Hot accretion disks around a supermassive black hole can also create ionise region.

In the ionised gas, electrons recombine in H^+ mostly in the upper levels and then decay to the fundamental level via multiple transitions (generally all permitted

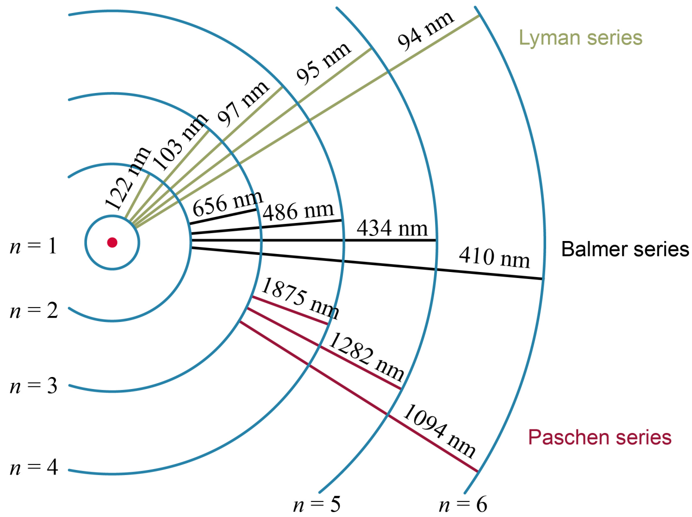

Figure 2.3 : Level transition in an Hydrogen atom : transition to the fundamental state ($n = 1$) are the Lyman transition, to level $n = 2$ the Balmer transition, etc...

transitions). The recombination rate is defined as :

$$\beta = \alpha(T) n_p n_e \tag{2.95}$$

where $\alpha(T)$ is the *recombination coefficient*. Therefore, the rate of recombination passing through the i level to the $j - level$ is :

$$\beta_{i\to j} = \alpha^{eff}_{i\to j} n_p n_e \tag{2.96}$$

where $\alpha^{eff}_{i\to j}$ is the *effective recombination coefficient* which gives the probability that recombination passes through the $i \to j$ transition.
The *emissivity of a recombination line* is given by :

$$\epsilon_{i\to j} = h\nu_{i\to j} \alpha^{eff}_{i\to j} n_p n_e \tag{2.97}$$

It is important to note that the intensity (and therefore the cooling capability) of collisionally excited line drops at $n > n_c$, relative to recombination lines (or drops relative to other collisionally excited lines with higher critical density, such as the "permitted" lines). As an example, the *Broad Line Region* (hereafter BLR) is a very compact region (R <1kpc) surrounding accreting supermassive black holes. In these regions, clouds are photo-ionised by the strong UV radiation field emitted by the accretion disk and have a density of 10^{11}cm^{-3}. As a consequence of $\epsilon \propto n^2$, the permitted and recombination lines (e.g. CIVλ1549, Hα, Hβ) emitted by the BLR reach luminosities that are much higher, not only of the forbidden

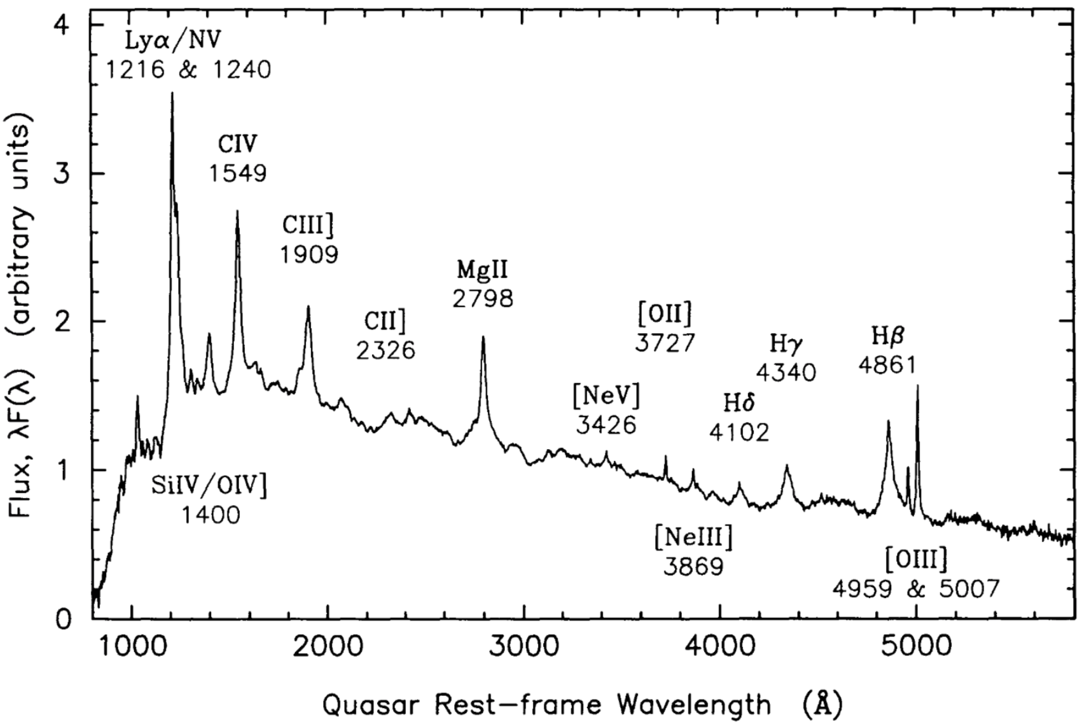

Figure 2.4 : Composite spectrum obtained by stacking 718 quasars spectra plotted as $\lambda F(\lambda)$ vs. rest-frame wavelength with the principal emission features identified. The flux scale is in arbitrary units. From Francis et al. (1991).

lines from the bLR itself (e.g. [OIII] λ5007, which is undetected), but also much more luminous than any line coming from the host galaxy, despite the size of the latter being much larger (Fig. 2.4).

2.3 Heating and cooling

2.3.1 Definition

Heating and cooling of gas is of central importance to our understanding of the formation of structure in the Universe. The net heating rate (Q) is defined as the difference between the total heating rate ($\Gamma(n,T)$) and the total cooling rate ($\Lambda(n,T)$) for the gas :

$$Q(n,T) = \Gamma(n,T) - \Lambda(n,T) \tag{2.98}$$

The equilibrium temperature of the gas is defined as the temperature when the cooling rate equal the heating rate. The stability of this equilibrium state can be found by considering deviations from equilibrium. For example, considering a gas at constant pressure and defining $\Delta T = T - T_E$, the enthalpy can be given by :

$$\frac{\mathrm{d}\Delta H}{\mathrm{d}t} = Q(T) \approx Q(T_E) + \Delta T \left(\frac{\partial Q}{\partial T}\right)_P (T_E) \tag{2.99}$$

but $Q(T_E)$ =0 then the gas will be thermally stable if :

$$\left(\frac{\partial Q}{\partial T}\right)_P \bigg|_{T_E} < 0 \tag{2.100}$$

Finally, we can define a useful timescale over which gas will cool : the *cooling time* as :

$$\boxed{\tau_c = \frac{U}{\Lambda}} \tag{2.101}$$

where U is the thermal energy of the gas.

2.3.2 The cooling curve

The cooling function Λ as a function of temperature and other physical conditions is called the *cooling curve* (Figure 2.5). It provides an overall description of the way in which gas will cool taking into account the different physical processes, which are effective over a wide range of temperatures and physical conditions. Detailed calculations of cooling curves are available which incorporates different processes operation at different temperatures. In general :

$$\Lambda(T) = \sum_i \Lambda_i \tag{2.102}$$

where the sum is over all processes contributing to the cooling at a given temperature. In the following, we will consider some of the processes which contributes to the details of this cooling curve and also heating processes.

2.3.3 Cooling by line emission

We have described in the previous section, that at high densities the emissivity of a line is given by :

$$\epsilon = \frac{g_1}{g_2} n_1 A_{21} h\nu_{12} e^{-\Delta E_{12}/k_B T} \tag{2.103}$$

we also demonstrated that at low densities below the critical density for collisional de-excitation, the emissivity is given by :

$$\epsilon = n_0 n_1 C_{12} h\nu_{12} \tag{2.104}$$

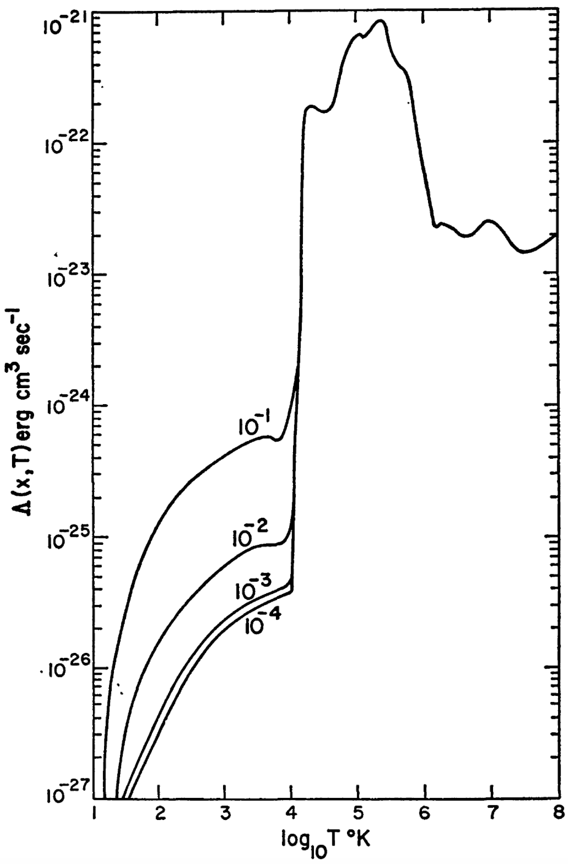

Figure 2.5 : Net cooling rate as a function of the temperature, also referred to the *cooling curve*. The curve is plotted for several density ranging from 10^{-4} to 0.1 cm^{-3}. From Dalgarno & McCray 1972.

We write the upwards collision rate in terms of the collision rate for downwards transitions as :

$$C_{12} = \frac{g_1}{g_2} C_{21} e^{-\Delta E_{12}/k_B T} \tag{2.105}$$

In either limit (high or low density) for line emission to be an effective cooling process the energy difference between the states must therefore be $\Delta E \sim k_B T$. Although hydrogen is very abundant in the Universe, the typical energy spacing between the levels $\Delta E \sim$10 eV for transitions out of the ground state. Therefore, hydrogen only becomes important to cooling for temperature of order 10^4 K. That's why we can see a huge break in the cooling curve at $T \sim 10^4$ K [2]

There are some ions with energy spacings which corresponds to lower temperature. For example, an important case is C^+ for which the energy of the $^2P_{1/2} \rightarrow ^2 P_{3/2}$ is $\Delta E/k_B$ =92 K. Collisional excitation can occur via collisions either with electrons or neutral hydrogen atoms. For $c^+ \leftrightarrow e^-$ collisions, the cooling rate is :

$$\Lambda_{C^+} = n_e n_{c^+} 8 \times 10^{-33} T^{-1/2} \exp(-92/T) \mathrm{J\ m^{-3} s^{-1}} \tag{2.106}$$

We can also note other ions with appropriate transitions :

- Si^+ $(^2P_{1/2} \rightarrow {}^2P_{3/2})$ $\Delta E/k_B$ =413 K
- O $(^3P_2 \rightarrow {}^3P_1)$ $\Delta E/k_B$ =228 K
- O $(^3P_2 \rightarrow {}^3P_0)$ $\Delta E/k_B$ =326 K

Note that in the low density regime (most cases) the cooling function and, therefore the cooling time is proportional to n^2. This has important implications for the expected cooling timescale of gaseous systems, both on small and large scales, as we shall see in the next lectures.

2.3.4 Cooling by free-free emission in ionised gas

For hot fully ionised gas (typically $T \gg 10^5$ K), radiation is produced via *Bremsstrahlung*. It is electromagnetic radiation produced by the deceleration of a charged particle when deflected by another charged particle, typically an electron by an atomic nucleus. The moving particle loses kinetic energy, which is converted into radiation (i.e., photons), thus satisfying the law of conservation of energy. The term is also used to refer to the process of producing the radiation. As demonstrated in PartII EM course, the Bremsstrahlung emissivity is given by :

$$\epsilon_\nu^{ff} = \frac{\mu_0 Z^2 e^6}{3\pi^2 c \epsilon_0^2 m^2} \left(\frac{\pi m}{6k}\right)^{1/2} g_{ff} n_e n_i T^{-1/2} e^{-h\nu/kT} \quad = a_1 g_{ff} n_e n_i Z^2 T^{-1/2} e^{-h\nu/kT} \tag{2.107}$$

[2] k_B =1.38×10^{-23} J K^{-1} and 1 eV=1.60×10^{-19} J

The factor g_{ff} is a *Gaunt factor* which is tabulated and is included to accommodate the results of detailed calculations in various limits.
Knowing that :

$$j_\nu = \frac{\epsilon_\nu}{4\pi} \tag{2.108}$$

$$j_\nu = \alpha_\nu B_\nu(T) \tag{2.109}$$

$$\text{and } B_\nu(T) = \frac{2h\nu^3/c^2}{\exp(h\nu/k_BT) - 1} \tag{2.110}$$

then the absorption coefficient is given by :

$$\alpha_\nu^{ff} = \frac{\mu_0 Z^2 e^6 c}{24\pi^3 \epsilon_0^2 m^2 h}\left(\frac{\pi m}{3k}\right)^{1/2} g_{ff} n_e n_i \nu^{-3} T^{-1/2}\left(1 - e^{-h\nu/k_BT}\right) \tag{2.111}$$

The Rayleigh-Jeans limit is a limit for low energies (or short frequencies), such as $h\nu \ll k_BT$, therefore :

$$1 - e^{-h\nu/k_BT} \approx h\nu/k_BT \tag{2.112}$$

then the absorption coefficient becomes :

$$\alpha_\nu^{ff} = \frac{\mu_0 Z^2 e^6 c}{24\pi^3 \epsilon_0^2 m^2 h}\left(\frac{\pi m h^2}{3k^3}\right)^{1/2} g_{ff} n_e n_i \nu^{-2} T^{-3/2} \tag{2.113}$$

$$= a_2 g_{ff} n_e n_i Z^2 \nu^{-2} T^{-3/2} \tag{2.114}$$

Example of the hydrogen atom: We can use the above results to calculate the thermal emission from fully ionised hydrogen at a temperature T. We will assume we are in a Rayleigh-Jeans limit (i.e. $h\nu \ll kT$) and the absorption coefficient is given by $(Z = 1)$:

$$\alpha_\nu^{ff} = a_2 g_{ff} n_e n_i \nu^{-2} T^{-3/2} \tag{2.115}$$

if the distance through the region is L then the optical depth is given by :

$$\tau_\nu = a_2 n_e^2 T_e^{-3/2} \nu^{-2} g_{ff} L \tag{2.116}$$

since in the case of a fully ionised hydrogen gas, we have $n_i = n_e$.
Detailed calculations show that $g_{ff} \propto T^{0.15}\nu^{-0.1}$. We can therefore plot the brightness as a function of the frequency, given that :

$$I_\nu = B_\nu(T)(1 - e^{-\tau_\nu}) \tag{2.117}$$

In the two limits of an optically thin or optically thick region we have :

$$I_\nu = \begin{cases} \tau_\nu B_\nu & \tau_\nu \ll 1 \\ B_\nu & \tau_\nu \gg 1 \end{cases}$$

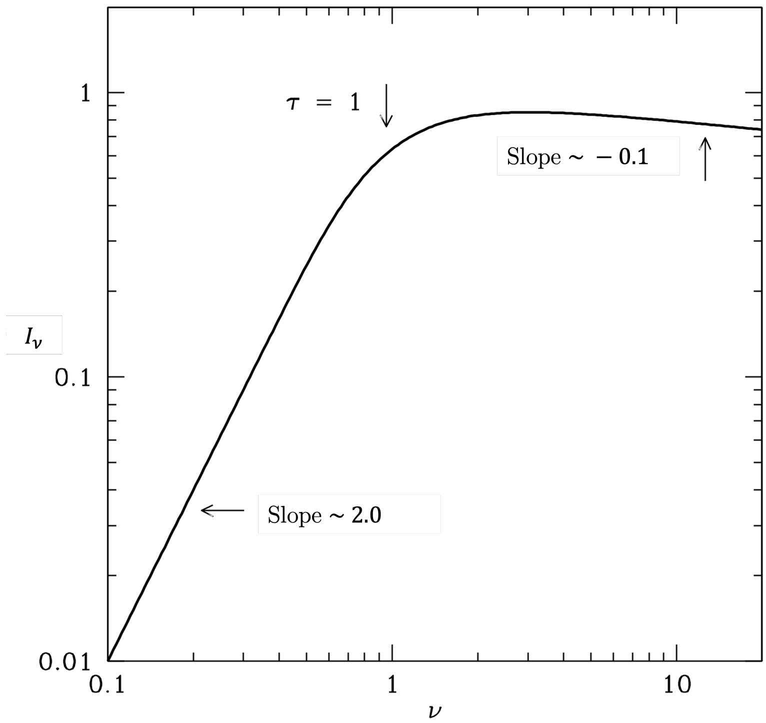

Figure 2.6 : Spectrum of thermal bremsstrahlung. The indicated slopes are for two regimes : optically thin and optically thick.

and

$$I_\nu \propto \begin{cases} \nu^{-0.1} & \tau_\nu \ll 1 \\ \nu^2 & \tau_\nu \gg 1 \end{cases}$$

Figure 2.6 shows an example of Bremsstrahlung for the hydrogen atom.
The cooling rate is found by integrating the quantity :

$$\Lambda_{ff} = \int \epsilon_\nu^{ff} \mathrm{d}\nu \qquad (2.118)$$

at all frequencies up to the frequency of the cutoff given by the exponential term in eq.2.107, which depends on T. Therefore we can rewrite the previous equation as :

$$\Lambda_{ff} \propto n_e n_i Z^2 T_e^{1/2} \approx a_{ff} n_e n_i Z^2 T_e^{1/2} = 1.435 \times 10^{-40} n_e n_i Z^2 T_e^{1/2} W m^3 \qquad (2.119)$$

Given that $d\Lambda/dT > 0$, then if the heating is constant, this results into a stable cooling process.

2.3.5 Cooling of molecular gas

In the cool molecular phase of the ISM, the excitation conditions for *rotational transitions* of molecules [3] are well-matched to the typical temperature in molecular clouds. From your Quantum Mechanics course, you have demonstrated that the rotational states have energy levels given by :

$$E_J = J(J+1)\frac{\hbar^2}{2I} = J(J+1)B \tag{2.120}$$

where I is the moment of inertia of the molecule.
The Einstein A coefficient for a rotational transition can be written as :

$$A_{nm} = \frac{8\pi^2}{3\hbar c^2} Z_0 \nu^3 | < n|\widehat{d}|m > |^2 \tag{2.121}$$

The molecule must have a permanent dipole $\widehat{d} = \mu$, then for the $J+1 \rightarrow J$ transition, we can quote the QM result :

$$A_{J+1,J} = \frac{8\pi^2}{3\hbar c^2} Z_0 \nu^3 | < J+1|\mu|J > |^2 \tag{2.122}$$

$$= \frac{8\pi^2}{3\hbar c^2} Z_0 \nu^3 \mu^2 \frac{J+1}{2J+1} \tag{2.123}$$

with selection rules $\Delta J = \pm 1$, $\Delta m_J = 0, \pm 1$, or $\Delta J = 0, \pm 1$.
The energy spacing between the level is :

$$h\nu_{J+1,J} = 2B(J+1) \tag{2.124}$$

Although H_2 is by far the most abundant molecule, it has *no permanent dipole moment* hence it cannot have transitions $\Delta J = \pm 1$, it can only undergoes quadrupole transitions, i.e. $\Delta J = \pm 2$. This causes H_2 rotational transitions possible only between levels with high ΔE (corresponding to excitation temperature larger than 500K). Hence it is observable in the mid-IR only in rare warm molecule regions. Therefore H_2 transitions are not a good coolant of the bulk of the molecular gas. However there are molecules which do have dipole transitions and with lower B, hence transitions that can be excited at much lower temperature, typical of the bulk of the ISM :

$$^{12}\text{CO} \quad J = 1 \rightarrow 0 \quad 115.27 \text{ GHz} \tag{2.125}$$

$$^{12}\text{CO} \quad J = 2 \rightarrow 1 \quad 230.54 \text{ GHz} \tag{2.126}$$

$$\text{CS} \quad J = 1 \rightarrow 0 \quad 48.99 \text{ GHz} \tag{2.127}$$

$$\text{NCN} \quad J = 1 \rightarrow 0 \quad 86.63 \text{ GHz} \tag{2.128}$$

and at $\sim$100 GHz, $T = h\nu/k_B \sim 5$K. These rotational transitions are excellent coolant of the cold molecular phase of the ISM.

[3] by definition a rotational transition is an abrupt change in angular momentum

2.3.6 Cooling by dust

Dust is one of the key element of the ISM, it accounts for ∼50% of the heavy elements. Even though dust is ∼1% of the baryonic mass of the Galaxy, it accounts for ∼40% of the luminosity. Dust grains are mainly produced by the stars at the end of their life. Originally discovered by the fact that it provides significant absorption at optical wavelength (due to the size of dust grains, ranging from 1nm to 1μm with a mean size of about 0.1μm), it is now clear that dust is crucial to the physics of structure formation. Dust grains are the solid phase of the ISM. The smallest particles are just large molecules such as the family of polycyclic aromatic hydrocarbons (PAHs). The larger particles are amorphous grain principally of silicates and carbon, but with a more complex icy surface layer (mantle).
The main consequences of dust in the ISM are :

- Dust grains provides significant absorption at optical wavelengths referred to, in this context, as *extinction*
- The grains absorb short wavelength photons (typically UV and optical), which excite phonon[4] modes within the grain giving the grain a characteristic temperature, *the grain then radiates thermally in the far- and mid-infrared*
- The surface of dust grains acts as a catalysts for chemical reaction in the ISM and the formation of large molecule (for example, most of the molecular hydrogen H_2 is formed on the surface of dust grains).
- The grains can also scatter photons elastically

Consider a grain of a single radius a and assume that scattering does not contribute to dust heating (we only consider absorption of photons), the absorption coefficient of a spherical grain is then given by :

$$\alpha_{ext}(\nu) = n_g \sigma_{ext}(\nu, a) = n_g Q_{ext}(\nu, a)\sigma_a = n_g Q_{ext}(\nu, a)\pi a^2 \tag{2.129}$$

where $\sigma_a = \pi a^2$ is the cross section[5] of the grain, n_g is the number density of grains and $Q_{ext}(\nu, a)$ is the efficiency of extinction (relative to the cross section) at a frequency ν. This extinction coefficient can be divided into terms, such as :

$$Q_{ext} = Q_{abs} + Q_{sca} \tag{2.130}$$

[4]by definition, a phonon is a collective excitation in a periodic, elastic arrangement of atoms or molecules in condensed matter, specifically in solids and some liquids

[5]by definition, the cross section is a measure of the probability that a specific process will take place when some kind of radiant excitation (e.g. a particle beam, sound wave, light, or an X-ray) intersects a localized phenomenon

where Q_{abs} is the efficiency of absorption (radiation absorbed by the grain) and Q_{sca} is the efficiency of scattering (radiation scattered in other direction). The power absorbed by a single grain from an incident radiation field F_ν is given by :

$$\int_0^{+\infty} F_\nu \sigma_a Q_{abs}(\nu, a)\,\mathrm{d}\nu \tag{2.131}$$

The grain reaches an equilibrium temperature, T_g, in this radiation field. According to the Kirchoff law [6], the emissivity is given by $\epsilon = 4\pi j_\nu = 4\pi\alpha_\nu B_\nu(T_g)$. Hence for a single grain the power radiated is :

$$4\pi \int_0^{+\infty} \sigma_a Q_{abs}(\nu, a) B_\nu(T_g)\,\mathrm{d}\nu \tag{2.132}$$

and in equilibrium the power absorbed must be equal to the power radiated, hence if we know Q_{abs} we can find T_g for a given incident radiation field. It can be shown that for typical dust properties :

$$T_g^{eq} \propto F_{UV}^{1/5} \tag{2.133}$$

in particular, recalling that $F = \frac{L}{4\pi R^2}$:

$$T_g^{eq}[K] \approx 40 L_{39}^{1/5} R_{pc}^{-2/5} \tag{2.134}$$

where L_{39} is the luminosity of the UV-optical radiation source in units of 10^{39} $\mathrm{erg\,s^{-1}}$ and R_{pc} is the distance from the source in parsec.
Typically dust heated by UV emission reaches temperatures of order 100K in star forming regions. If there is a significant amount of dust then the interstellar medium is optically thick to the optical and UV radiation, but the resulting far-infrared emission is completely optically thin and escapes resulting in efficient cooling (Figure 2.7). This process is certainly extremely important in very dusty star-forming galaxies.

2.3.7 Radiative heating and cooling by recombination

Photons with energies greater than the ionisation potential (I_i) of a species lead to the ejection of an electron with energy $h\nu - I_i$. This electron can then heat the gas via collision processes. Some of the electron kinetic energy will lead to excitation of electronic levels and subsequent re-radiation and hence no heating of the gas.
To illustrate this phenomena, we consider a cloud of pure hydrogen (typically a nebula or a HII region). The ionisation rate is given by :

$$n_{H^0} S_\star \sigma_i \tag{2.135}$$

[6]For an arbitrary body emitting and absorbing thermal radiation in thermodynamic equilib-

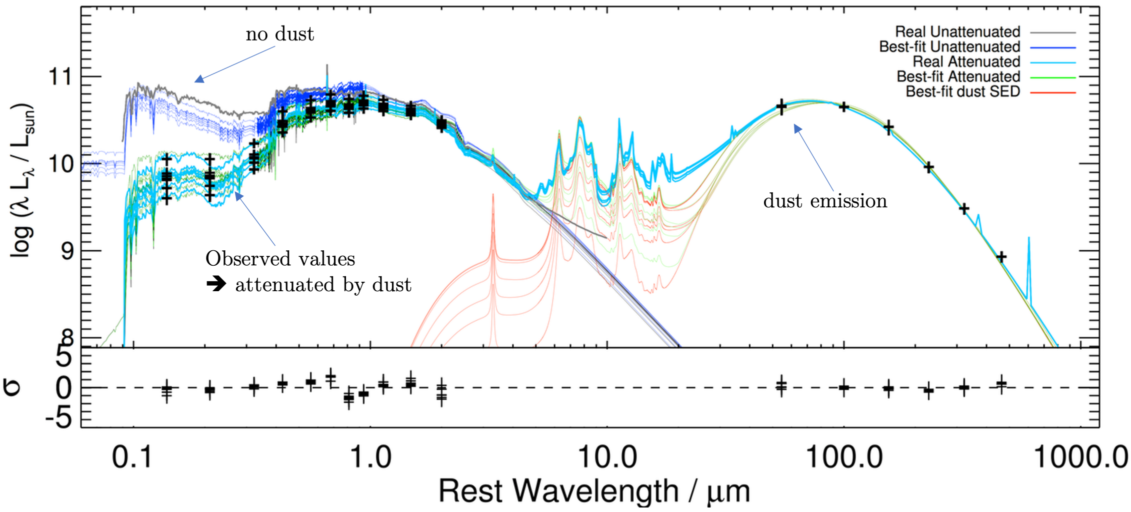

Figure 2.7 : Spectral Energy Distribution (SED) of a galaxy showing the unattenuated emission (short wavelength) and the emission of dust (large wavelength). From Hayward & Smith (2015)

where $S_\star$ is the flux of ionising photons, n_{H^0} the density of neutral hydrogen and σ_i the ionisation cross section. In equilibrium, the ionisation rate must equal the recombination rate, therefore :

$$n_{H0} S_\star \sigma_i = n_e^2 \alpha_B \tag{2.136}$$

here α_B is the net recombination coefficient and we have assume $n_e = n_i$.
The heating rate is given approximately by :

$$\Gamma = n_{H^0} S_\star \sigma_i (h\overline{\nu} - I_H) \tag{2.137}$$

since $h\overline{\nu} - I_H$ is the mean energy of ejected electrons (I_H is the ionisation energy of hydrogen, i.e. 13.6 eV). We assume the electrons are characterised by a temperature T_e with mean energy per electron of $\frac{3}{2}kT_E$. This kinetic energy is lost on recombination, hence the cooling rate is :

$$\Lambda = n_e^2 \alpha_B \frac{3}{2} kT_e \tag{2.138}$$

Equating heating and cooling rate gives :

$$n_{H^0} S_\star \sigma_i (h\overline{\nu} - I_H) = n_e^2 \alpha_B \frac{3}{2} kT_e \tag{2.139}$$

rium, the emissivity is equal to the absorptivity therefore :

$$T_e = \frac{2}{3}\frac{h\overline{\nu} - I_H}{k_B} \tag{2.140}$$

If the ionisation radiation is from a central star, then we can approximate the emission is approximately that of a thermal emitter of temperature $T_\star$ and a reasonable approximation is that $h\overline{\nu} - I_H \sim k_B T_\star$ implying :

$$T_e \sim \frac{2}{3}T_\star \tag{2.141}$$

Typically $T_\star \sim 3 \times 10^4 - 6 \times 10^4$ giving $T_e \sim 4 \times 10^4 - 8 \times 10^4$.
In a gas with a high ionisation fraction, all of the electron energy is available to heat the gas via collisions. In a mainly neutral gas, inelastic collisions give rise to hydrogen (and metal) emission lines which escape the cloud and therefore not all the electron energy is available for heating the gas. In cool mostly neutral clouds where ionisation of metals (e.g. C, Si, Fe) occurs, this heating by starlight can then be efficient.
Similar arguments also apply to heating by X-rays and cosmic rays (energetic ionised particles), which ionise principally hydrogen, and the emission of photoelectrons from dust particles.

2.3.8 Mechanical heating

Mechanical heating can occur in many different ways provided there is some mechanisms to dissipate kinetic energy and transform it into heat. The main processes we will come across will be heating in shocks (which you met last year) and heating in viscous accretion discs (from last term course). As a very good estimate, in a strong shock all of the kinetic energy of the upstream gas is converted into internal energy (heat) downstream of the shock.

2.4 The multi-phase ISM

We start off by considering the two phases of neutral hydrogen. Equating the net cooling rate to the net heating rate gives the equilibrium temperature as a function of density (eq. 2.139). Figure 2.8 shows an ionised calculation where the heating is principally due to photoelectrons from grains. From this, we can construct the equilibrium pressure as a function of the number density :

- For the pressure shown, there are three equilibrium points (A, B and C)
- The middle point is however unstable ; if the gas is compressed slightly, the pressure drops and the gas compresses further

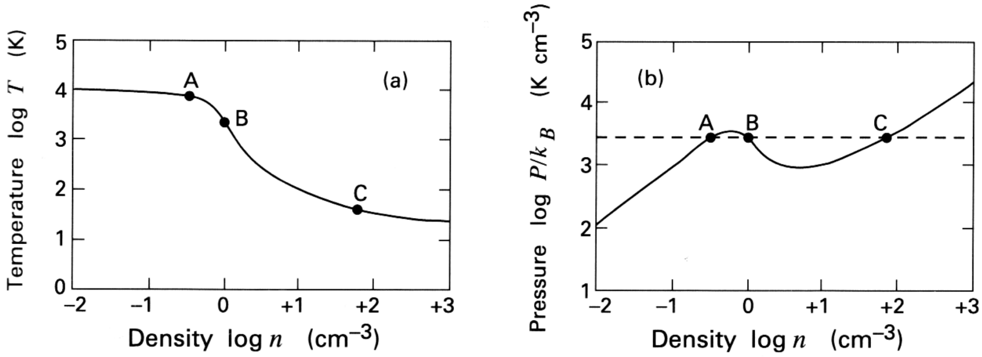

Figure 2.8 : Left : Equilibrium temperature versus number density ; Right : equilibrium pressure versus number density.

- There are therefore *two* stable equilibrium points : one hot at low density (A) and one cold at high density (B).

A model of this type explains the warm and cold hydrogen phases. Indeed, the gas in the ISM is a *multi-phase medium* in which different phases are in approximate pressure balance.
Can we explain molecular clouds in this way ? The answer is NO. The masses of a so-called *giant molecular clouds* is $M \sim 10^5 M_\odot$ with a radius of ∼50 pc. The gravitational potential energy is of order $\sim GM^2/r$ while the thermal energy is of order $\sim \frac{3}{2}(M/m_H)k_BT$, hence :

$$\frac{E_{grav}}{E_{thermal}} \sim \frac{GMm_H}{rk_BT} \approx 100 \tag{2.142}$$

Such clouds are self-gravitating and are the sites for massive star-formation.
For the hot ionised phase it is useful to calculate the cooling time. Cooling at $\sim 5\times10^5$ K is dominated by line emission from collisionnaly-excited ions and :

$$\Lambda \approx 1.6 \times 10^{-35} n_e n_i \left(\frac{T}{10^6}\right)^{-0.6} \text{W m}^{-3} \tag{2.143}$$

and the cooling time for the gas with $T \sim 5\times10^5$K and $n_e \sim 3\times10^3$m^{-3} :

$$\tau_c = \frac{\frac{3}{2}n_e k_B T}{\Lambda} \sim 4 \times 10^6 \left(\frac{T}{5\times10^5}\right)^{1.6} \left(\frac{n_e}{3\times10^3}\right)^{-1} \text{yr} \tag{2.144}$$

The gas does not need a constant heating source, but still cools quickly on timescale much shorter than those over which a galaxy evolves. The heating source is in

Phase	n_{tot} (10^6 m^{-3})	$T(K)$	M / 10^9M$_\odot$	f
Molecular	> 300	10	4.0	0.01
Cold neutral	50	80	3.0	0.04
Warm neutral	0.5	$\sim$5000	2.0	0.3
Warm ionised	0.3	10 000	$\sim$0.2	0.15
Hot ionised	3×10^{-3}	3×10^5	$<$0.02	0.5

Table 2.2: The different phase of the neutral hydrogen in the ISM

fact shocks produced by supernova remnants ; the gas can however cool rapidly especially in any region in which the density is slightly higher than typical value and then condense to one of the denser phases.