Formation of Structure in the Universe

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

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.

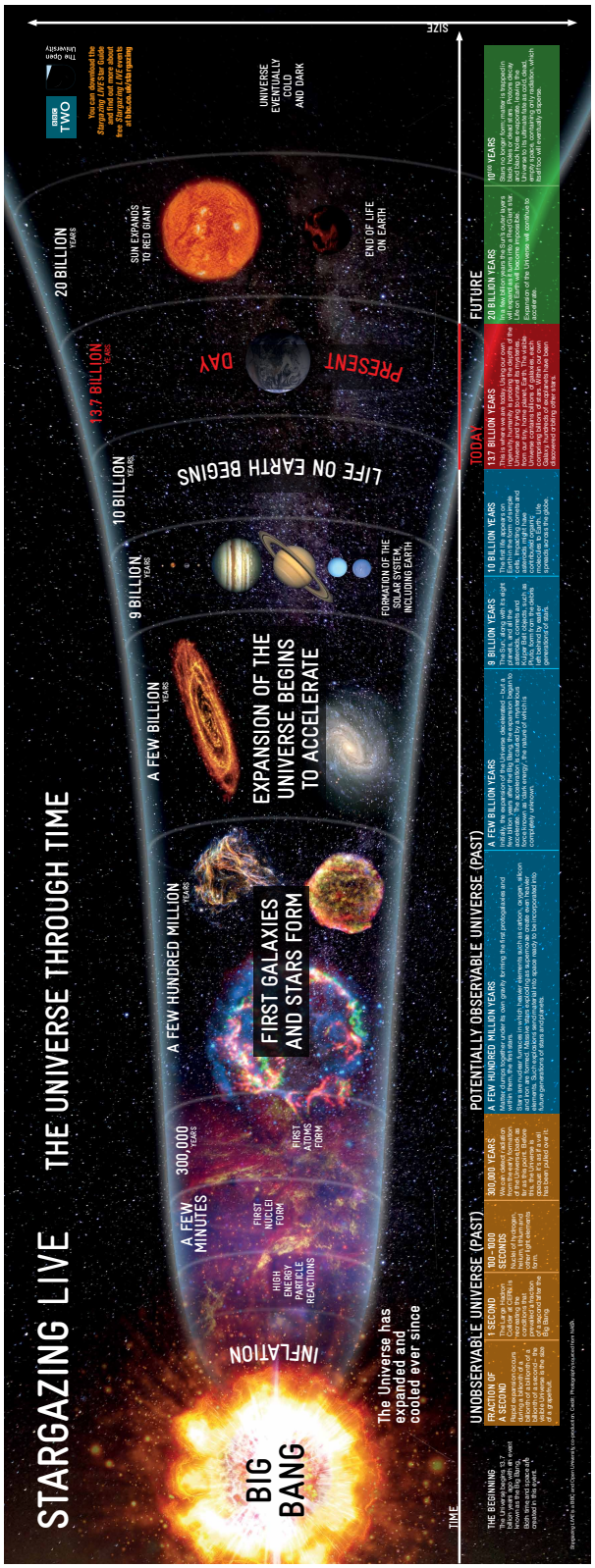


Figure 1.1 : The history of the Universe: (bottom) timeline of the history of the Universe with specific events, such as the inflation and the formation of neutral hydrogen. Source: BBC

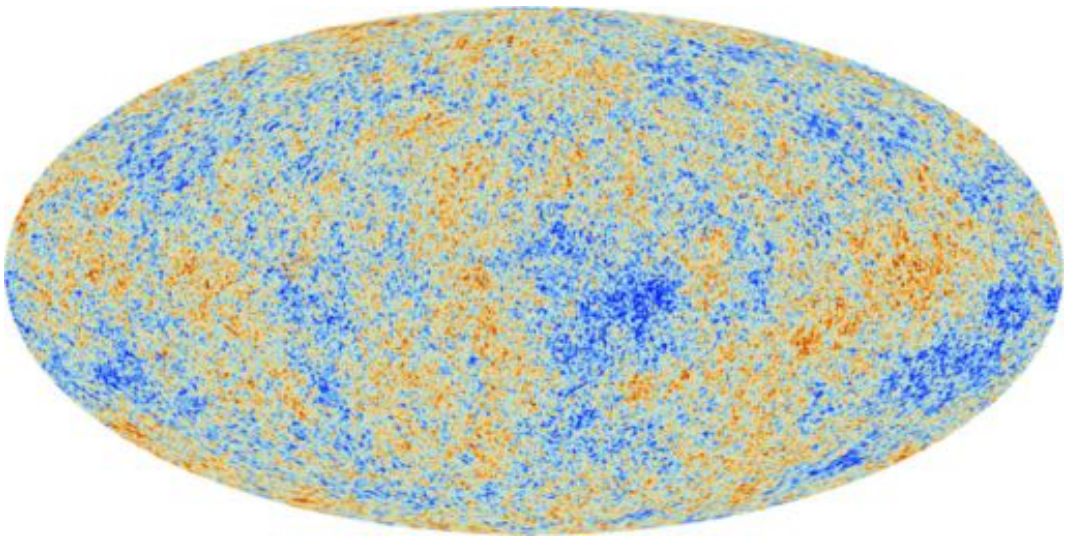


Figure 1.2 : The Cosmic Microwave Background observed by the Planck satellite. The colours trace the temperature of the radiation. It demonstrates the non-uniformity of the universe 370 000 years after the Big-Bang. Source : Planck collaboration

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).

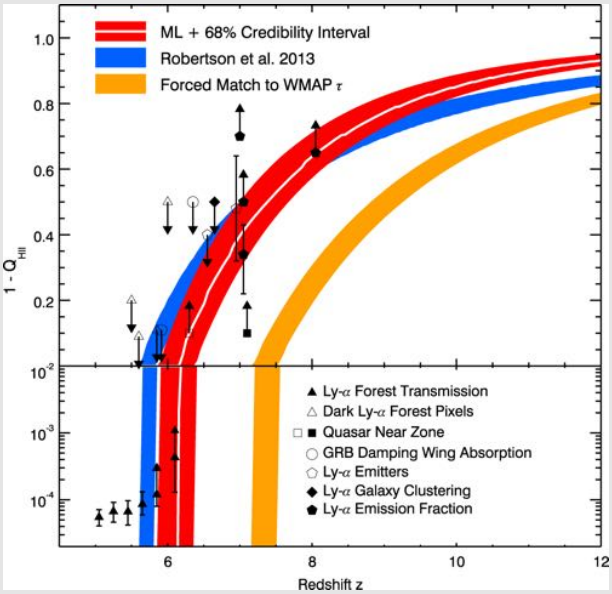


Figure 1.3 : Evolution of the neutral fraction of hydrogen (1-Q_{HII}) as a function of redshift. The data points show the constrain from galaxies/quasars observations. The coloured regions are from CMB observations with different assumptions (blue and red used *Planck* data and the yellow curve is based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

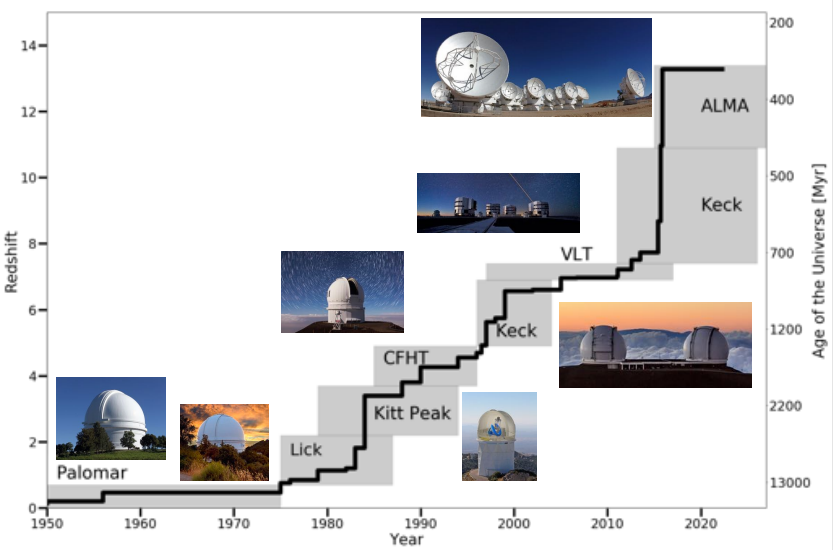
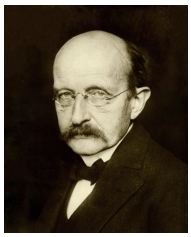


Figure 1.4 : Evolution of the redshift of the most distant galaxy known as a function of time. The name and picture of each telescope involved in the discovery is indicated. This evolution is exponential and follows the opening of new observatory.

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 13^{th} 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 \geq$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 :

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

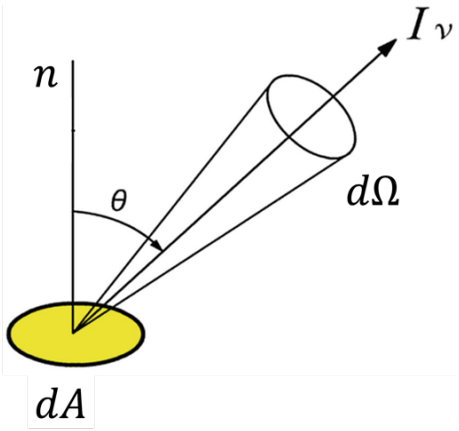


Figure 2.1 : Geometry for obliquely incident rays

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 :

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

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

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 $\mathrm{d}F$ 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 $\mathrm{d}V$ 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 $\mathrm{d}t$ 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

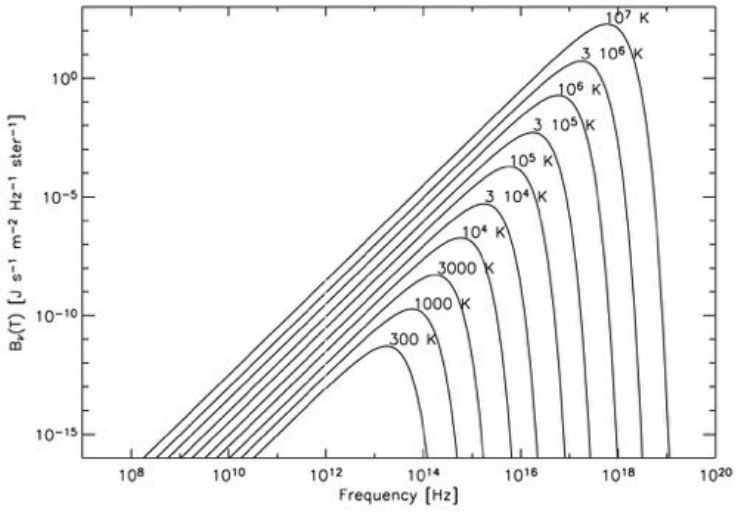


Figure 2.2 : Spectrum of blackbody radiation at various temperature (from Kraus J.D. 1966)

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 $\mathrm{d}S$ 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_ν=B_ν 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 $\mathrm{d}t$ and volume $\mathrm{d}V$ is :

$$\mathrm{d}V\,\mathrm{d}t\,\frac{h\nu_0}{4\pi}n_1B_{12}\int\mathrm{d}\Omega\int\mathrm{d}\nu\phi(\nu)I_\nu \tag{2.75}$$

Therefore the energy absorbed out of a beam in frequency range $\mathrm{d}\nu$ solid angle $\mathrm{d}\Omega$ time dt and volume $\mathrm{d}V$ is :

$$\mathrm{d}V\,\mathrm{d}t\,\mathrm{d}\Omega\,\mathrm{d}\nu\frac{h\nu_0}{4\pi}n_1B_{12}\phi(\nu)I_\nu \tag{2.76}$$

Assuming the volume element is $\mathrm{d}V = \mathrm{d}A \times \mathrm{d}s$, the absorption coefficient is given by :

$$\alpha_\nu = \frac{h\nu}{4\pi}n_1B_{12}\phi(\nu) \tag{2.77}$$

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 :

$$\alpha_\nu = \frac{h\nu_0}{4\pi}\phi(\nu)(n_1B_{12} - n_2B_{21}) \tag{2.78}$$

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

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

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 : $n_0 \sim n_e \propto n_{\text{atoms}} \propto n$, where n is the total gas density. Also, $n_1, n_2 \propto n_{\text{atoms}} \propto n$. 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$\times 10^7$ 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

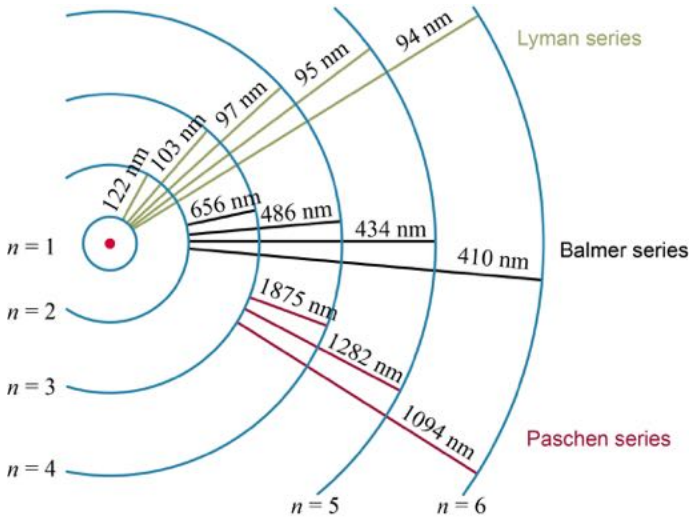


Figure 2.3 : Level transition in an Hydrogen atom : transition to the fundamental state $(n = 1)$ are the Lyman transition, to level $n = 2$ the Balmer transition, etc...

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 $10^{11}\mathrm{cm}^{-3}$. 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

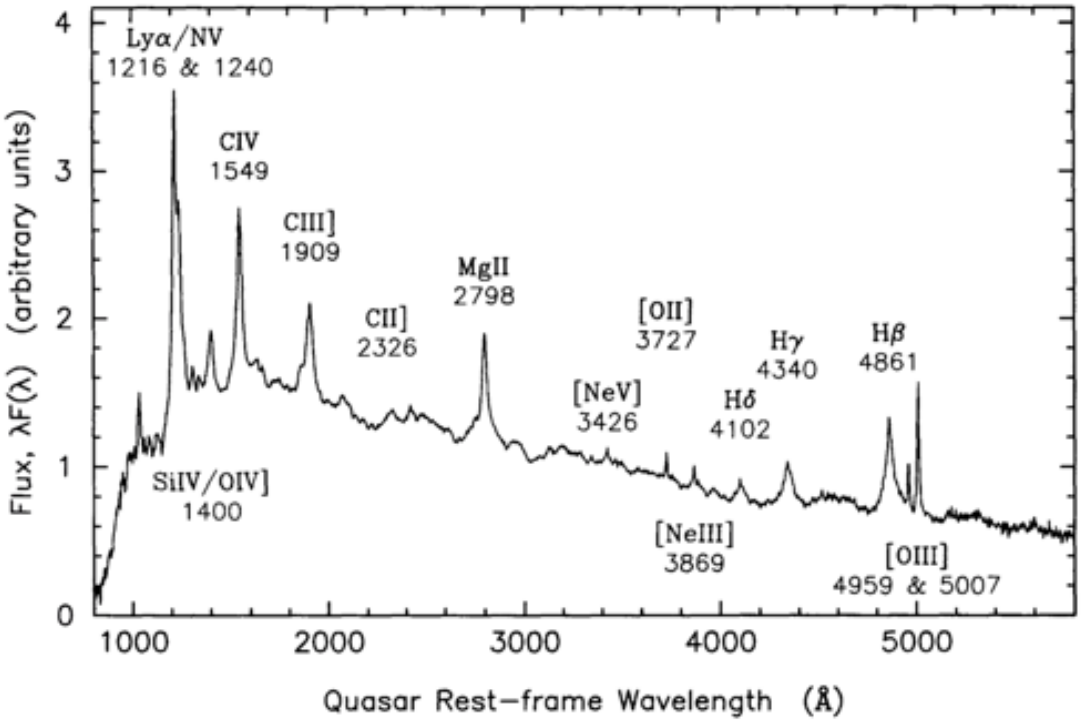


Figure 2.4 : Composite spectrum obtained by stacking 718 quasars spectra plotted as $\lambda F(\lambda)$ vs. rest-frame wavelength with the principal emission features identified. The flux scale is in arbitrary units. From Francis et al. (1991).

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 :

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

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 :

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

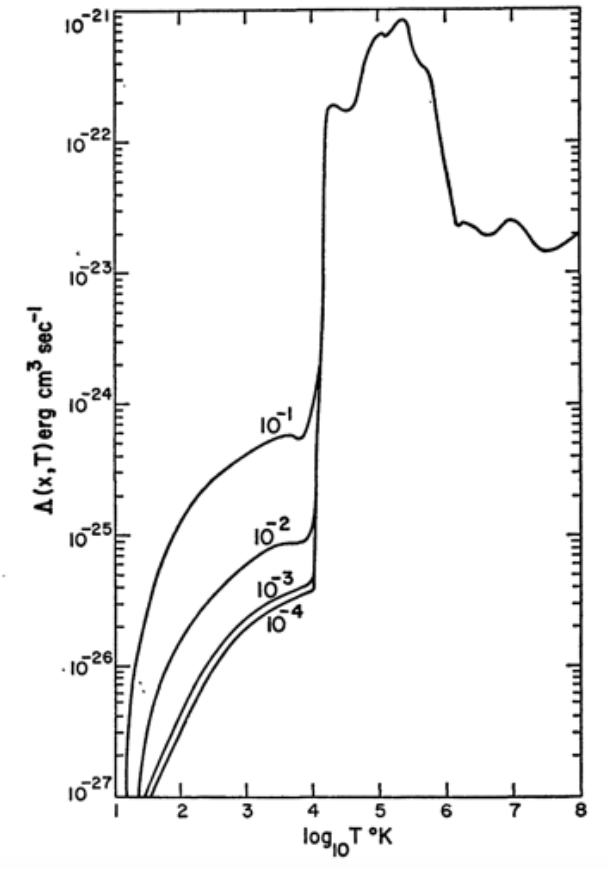


Figure 2.5 : Net cooling rate as a function of the temperature, also referred to the *cooling curve*. The curve is plotted for several density ranging from 10^{-4} to 0.1 cm^{-3}. From Dalgarno & McCray 1972.

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 :

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

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 :

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

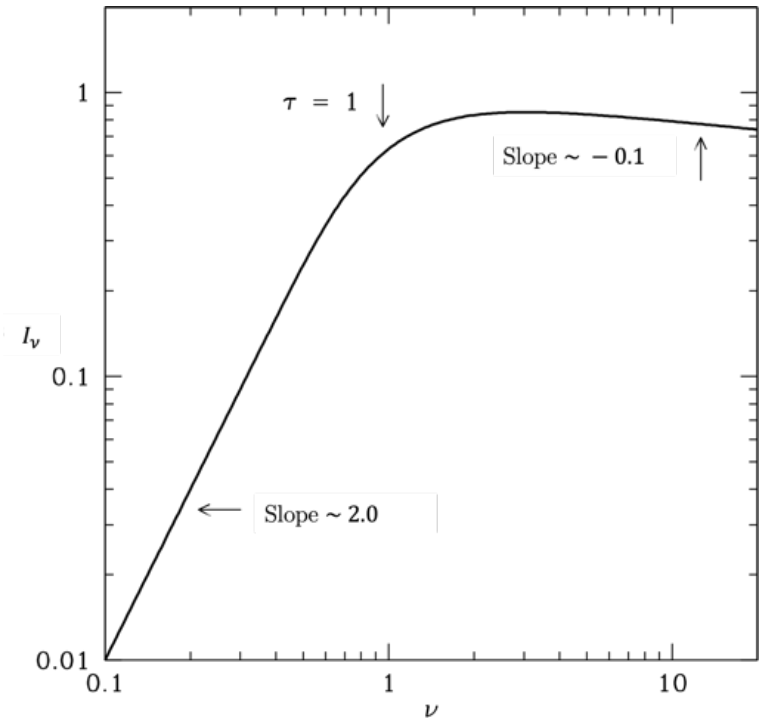


Figure 2.6 : Spectrum of thermal bremsstrahlung. The indicated slopes are for two regimes : optically thin and optically thick.

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 :

$$\Lambda_{ff} = \int \epsilon_\nu^{ff} \mathrm{d}\nu \tag{2.118}$$

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 :

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

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}CO	$J = 1 \rightarrow 0$	115.27 GHz	(2.125)
^{12}CO	$J = 2 \rightarrow 1$	230.54 GHz	(2.126)
CS	$J = 1 \rightarrow 0$	48.99 GHz	(2.127)
NCN	$J = 1 \rightarrow 0$	86.63 GHz	(2.128)

and at $\sim$100 GHz, $T = h\nu/k_B \sim$ 5K. 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} 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-

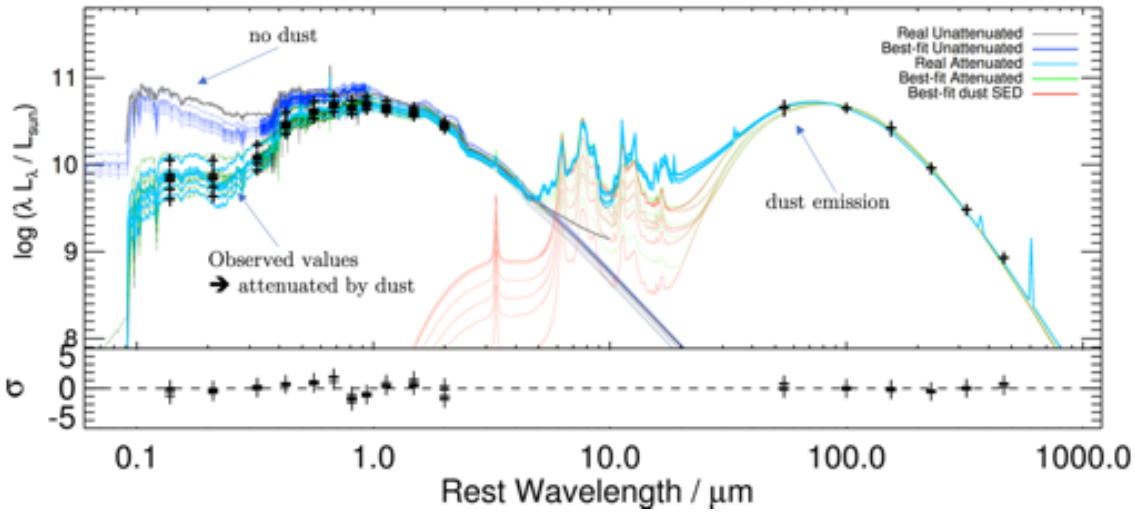


Figure 2.7 : Spectral Energy Distribution (SED) of a galaxy showing the unattenuated emission (short wavelength) and the emission of dust (large wavelength). From Hayward & Smith (2015)

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

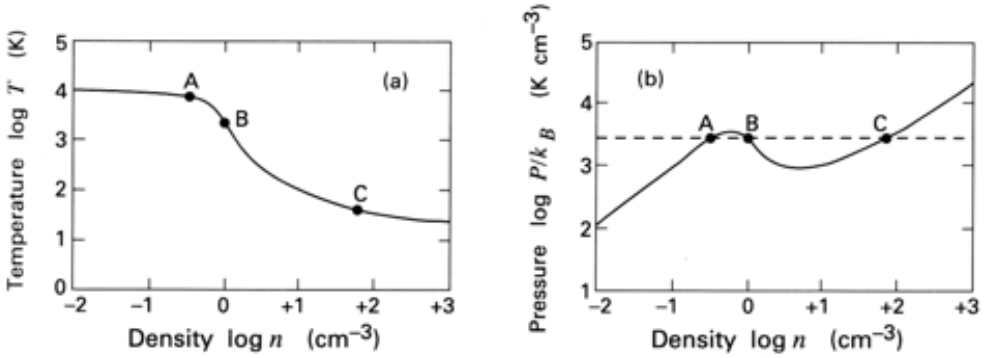


Figure 2.8 : Left : Equilibrium temperature versus number density ; Right : equilibrium pressure versus number density.

- 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 $n_e \sim 3\times10^3 \text{m}^{-3}$:

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

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.

3. Gravitational stability and instability

Examples ... which might be multiplied ad libitum, show how difficult it often is for an experimenter to interpret his results without the aid of mathematics.

— John William Strutt Rayleigh

In this chapter, we consider the physics of isolated self-gravitating gaseous systems (clouds). We will look at the stability of these systems as a first step to understand gravitational collapse and star formation. You met some of this material last year in the Astrophysics fluids course, but we will now consider the physics in more details.

3.1 Equations of hydrodynamics and hydrostatic equilibrium

The equations describing the hydrostatic equilibrium [1] of a cloud of gas are essentially the same equations which describe stellar structure.
In fluid dynamics, the *Euler equation* is a quasilinear hyperbolic equation governing adiabatic and inviscid flow (viscosity equal to 0):

$$\boxed{\rho\frac{\partial\vec{v}}{\partial t}+\rho(\vec{v}\cdot\nabla)\vec{v}=-\nabla P-\rho\nabla\Phi_g} \tag{3.1}$$

where v is the flow velocity, ρ the density of the fluid, P the mechanic pressure and Φ_g the gravitational potential. In equilibrium ($v=0$), the Euler equation reduces

[1] By definition, hydrostatic equilibrium is the condition of a fluid or plastic solid at rest, which occurs when external forces, such as gravity, are balanced by a pressure-gradient force.

to :

$$-\nabla P - \rho \nabla \Phi_g = 0 \tag{3.2}$$

The *Poisson's equation* for gravity gives the gravitational potential :

$$\boxed{\nabla^2 \Phi_g = 4\pi G\rho} \tag{3.3}$$

The equation of state for the gas is that for an ideal gas :

$$\boxed{p = \frac{\rho k T}{\mu}} \tag{3.4}$$

where μ is the mean mass per particle.
Finally, we also note the equation of continuity :

$$\boxed{\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0} \tag{3.5}$$

In general, we also need an equation describing the energy flux : the most general form is complicated and in the following sections we will make some approximations which avoid this complication.

3.2 The isothermal sphere

The simplest example of a system in which thermal pressure and self gravity give rise to a stable configuration is the *isothermal* sphere. We will use this as the equilibrium model in discussion of gravitational collapse.
We assume :

- a spherical symmetry
- the gas has a uniform temperature T
- the equation of state is given by :

$$\boxed{P = \rho \frac{k_B T}{\mu} = a_T^2 \rho} \tag{3.6}$$

  where $a_T^2 = k_B T/\mu$ is the *isothermal sound speed*[2]

With these assumptions, the spherically symmetry equations of hydrostatic equilibrium become :

$$-\frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}r} - \frac{\mathrm{d}\Phi_g}{\mathrm{d}r} = 0 \tag{3.7}$$

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 \frac{\mathrm{d}\Phi_g}{\mathrm{d}r}\right) = 4\pi G\rho \tag{3.8}$$

[2]the speed of sound is the speed at which pressure disturbances travel in a medium

3.2.1 Singular isothermal sphere

In this first section, we will try to get the solution for a singular isothermal sphere.

The gravitational force is given by :

$$-\rho\nabla\Phi_g = -\frac{GM\rho}{r^2} \tag{3.9}$$

Therefore, we can find one solution to equation 3.7 in terms of mechanical pressure :

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{GM\rho}{r^2} \tag{3.10}$$

where the mass, M, is the mass within the radius r such as :

$$M = M(r) = 4\pi \int_0^r \rho(r')r'^2\,\mathrm{d}r' \tag{3.11}$$

or in differential form :

$$\frac{\mathrm{d}M}{\mathrm{d}r} = 4\pi r^2\rho \tag{3.12}$$

We can therefore rewrite the isothermal equation of state (eq:3.6), we obtain :

$$\frac{\mathrm{d}\rho}{\mathrm{d}r} = -\frac{GM}{a_T^2}\frac{\rho}{r^2} \tag{3.13}$$

It is straightforward to verify that the following form of the density profile satisfies these equations :

$$\rho(r) = \frac{a_T^2}{2\pi G r^2} \tag{3.14}$$

From the previous equations, we can determine several properties of the cloud as follow :

1. its total mass can be found by integrating out to the radius of the cloud r_0 :

$$M_0 = \int_0^{r_0} \frac{a_T^2}{2\pi G r^2} 4\pi r^2\,\mathrm{d}r = \frac{2a_T^2 r_0}{G} \tag{3.15}$$

2. the boundary of the cloud has to be in equilibrium, there must be an external pressure, P_0, which equals the pressure at the surface of the cloud :

$$P_0 = a_T^2\rho(r_0) = \frac{a_T^4}{2\pi G r_0^2} \tag{3.16}$$

3. The cloud radius and the isothermal sound speed can be estimated from the mass of the gas and the external pressure. Given the mass of the gas, M_0, and external pressure p_0, the cloud radius r_0 and isothermal sound speed, a_T^2 (or equivalently temperature) can be found.

4. Although the density and pressure of the cloud diverge as $r \to 0$ the total mass, internal energy, etc... are bounded.

3.2.2 General solution

We now seek the general solution for the isothermal sphere.

Returning to the original spherically symmetric force equation (eq. 3.7), we have:

$$-\frac{1}{\rho}\frac{\mathrm{d}p}{\mathrm{d}r} - \frac{\mathrm{d}\Phi_g}{\mathrm{d}r} = 0 \qquad \Rightarrow \qquad -\frac{a_T^2}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}r} = \frac{\mathrm{d}\Phi_g}{\mathrm{d}r} \tag{3.17}$$

and integrating:

$$-\ln\rho = \frac{\Phi_g}{a_T^2} + \text{constant} \qquad \text{or} \qquad \rho = \rho_c e^{-\Phi_g/\alpha_T^2} \tag{3.18}$$

where ρ_c is a constant.

Introducing the dimensionless variables $\psi = \Phi_g/a_T^2$ and $\xi = (4\pi G\rho_c/a_T^2)^{1/2}r$, eq.3.8 becomes :

$$\frac{1}{\xi^2}\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\xi^2\frac{\mathrm{d}\psi}{\mathrm{d}\xi}\right) = e^{-\psi} \tag{3.19}$$

The singular solution of the previous equation is:

$$\psi = \ln(\xi^2/2) \tag{3.20}$$

At the center of the cloud ($r = 0$, then $\xi = 0$), the gravitational force must vanish, therefore :

$$\left.\frac{d\psi}{d\xi}\right|_{\xi=0} = 0 \tag{3.21}$$

We also require $\psi(\xi = 0) = 0$ which fixes ρ_c to be the core density.
Unfortunately there is no analytic solution which matches these boundary conditions and the equation must be integrated numerically.
We can estimate the total mass within a radius $r_0(\xi_0)$ as follow :

$$M_0 = 4\pi\int_0^{r_0}\rho r^2\,\mathrm{d}r \tag{3.22}$$

$$= 4\pi\rho_c\Big(\frac{a_T^2}{4\pi G\rho_c}\Big)^{3/2}\int_0^{\xi_0} e^{-\psi}\xi^2 d\xi \tag{3.23}$$

$$\tag{3.24}$$

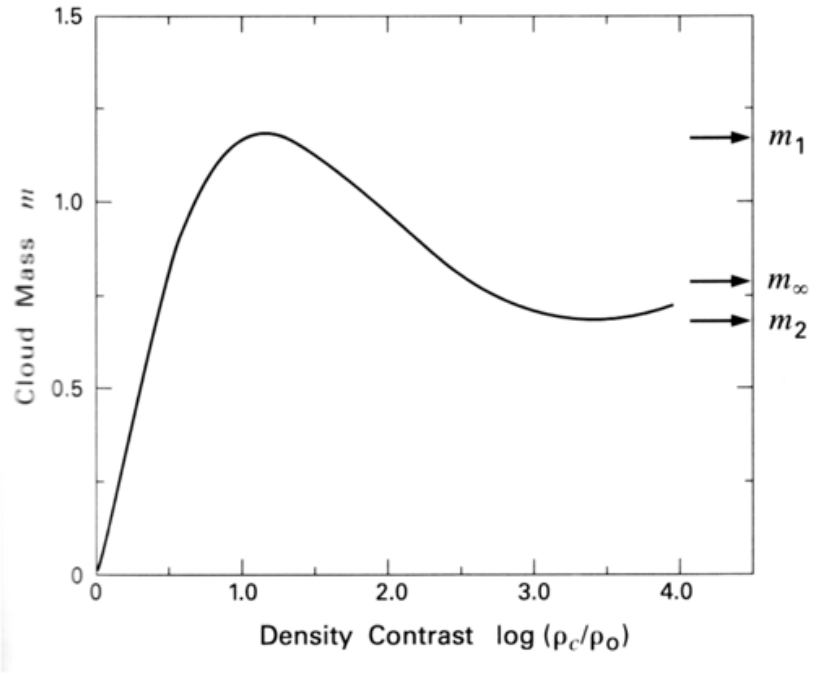


Figure 3.1 : Evolution of the dimensionless mass of a clouds ($m = P_0^{1/2} G^{3/2} M_0 / a_T^4$) as a function of the density contrast. It clearly shows there is a mass (m_1) at which the cloud is in equilibrium.

From eq:3.19, we deduce that

$$M_0 = 4\pi\rho_c \left(\frac{a_T^2}{4\pi G\rho_c}\right)^{3/2} \left(\xi^2 \frac{d\psi}{d\xi}\right)_{\xi=\xi_0} \tag{3.25}$$

We still have the boundary condition $P_0 = a_T^2 \rho(r_0)$.

There is an additional parameter ρ_c compared to the singular solution, hence the solution is not fully constrained by M_0 and P_0, and we must additionally specify ρ_c, a_T^2, or r_0.
It could also be useful to consider a whole family of solutions described by the ratio ρ_c/ρ_0 ("density contrast") where $\rho_0 = \rho(r_0)$. The variation of mass as a function of the density contrast is shown in Figure3.1, where it is clear that for a given external pressure p_0 and temperature (a_T) there is a maximum mass (m_1) for which the cloud can be in equilibrium.

3.3 The polytropic sphere

Polytropes are self-gravitating gaseous spheres that were, and still are, very useful as crude approximation to more realistic stellar models. Last year, in the astrofluids course you studied the polytropic sphere with an equation of state of the form:

$$\boxed{P = K\rho^{1+\frac{1}{n}} = K\rho^{\Gamma}} \tag{3.26}$$

where n is the polytropic index ($n = 0$ for rocky planets, $n = 1.5$ is a good model for fully convective star cores - like those of red giants- , brown dwarfs, giant gaseous planets - like Jupiter). In general as the polytropic index increases, the density distribution is more heavily weighted toward the center ($r = 0$) of the body. Case with $n = \infty$ is the isothermal case we studied at the previous section.

For the general polytropic case, it is easy to demonstrate that the temperature always follows the gravitational potential :

$$k_B T = \frac{1-\Gamma}{\Gamma}\mu\Phi_g \tag{3.27}$$

3.4 Virial equilibrium for the self-gravitating sphere

In the following section, we will look at the same problem but with a different view which throws away the internal detail of the cloud but which enables us to more easily examine whether the solutions we find are stable.

The equations of hydrostatic equilibrium in terms of mass are :

$$\frac{\mathrm{d}P}{dr} = -\frac{GM\rho}{r^2} \tag{3.28}$$

$$\frac{\mathrm{d}M}{\mathrm{d}r} = 4\pi r^2\rho \tag{3.29}$$

We can rewrite these equations treating mass as an independent variable, such as:

$$\mathrm{d}r = \frac{\mathrm{d}M}{4\pi r^2\rho} \tag{3.30}$$

hence,

$$4\pi r^3 \mathrm{d}P = -4\pi r GM\rho\,\mathrm{d}r = -\frac{GM}{r}\mathrm{d}M \tag{3.31}$$

We know that $4\pi r^3 = 3V$, then :

$$\int_{V=0,p=p_c}^{V=V_0,p=p_0} 3V\,\mathrm{d}P = -\int_0^{M_0}\frac{GM}{r}\mathrm{d}M \tag{3.32}$$

A useful mathematical method to solve the previous equation is "integrating by parts" such as :

$$\int_a^b u(x)v'(x)\mathrm{d}x = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)\,\mathrm{d}x \tag{3.33}$$

Therefore, integrating the left hand side by parts give :

$$3\,[PV]_{0,p_c}^{V_0,p_0} - 3\int_{0,p_c}^{V_0,p_0} P\,\mathrm{d}V = -\int_0^{M_0}\frac{GM}{r}\,\mathrm{d}M \tag{3.34}$$

The first term of the previous equation evaluates to $3P_0V_0 = 4\pi r_0^3 P_0$; the right hand side is just the gravitational potential energy of the cloud:

$$\Omega = -\int_0^{M_0}\frac{GM}{r}\,\mathrm{d}M \tag{3.35}$$

and the final term can be re-written treating the mass as an independent variable, hence $\mathrm{d}V = \mathrm{d}M/\rho$, and finally :

$$\boxed{3\int\frac{p}{\rho}\,\mathrm{d}M + \Omega = 4\pi r_0^3 p_0} \tag{3.36}$$

This equation is always valid since we have made no assumption about the internal structure of the cloud.

3.5 Stability of an isothermal cloud

In the following section, we will consider the case of an isothermal sphere. We can therefore simplify the previous equation, considering that : $P = a_T^2\rho$:

$$3\int\frac{P}{\rho}\,\mathrm{d}M = 3a_T^2 M_0 = 3\frac{k_B T}{\mu}M_0 \tag{3.37}$$

Also $\Omega \sim -\frac{3}{5}GM_0^2/r_0$, the virial equilibrium then becomes:

$$\boxed{3\frac{k_B T}{\mu}M_0 - \frac{3}{5}\frac{GM_0^2}{r_0} - 4\pi r_0^3 P_0 = 0} \tag{3.38}$$

We can use this equation to discuss the stability of a general cloud : if this equation is $= 0$, the cloud is in equilibrium. However if it is <0 the clouds will collapse, and if it is positive, the cloud will expand.
For a cloud of mass M_0 and temperature T in equilibrium in an environment of

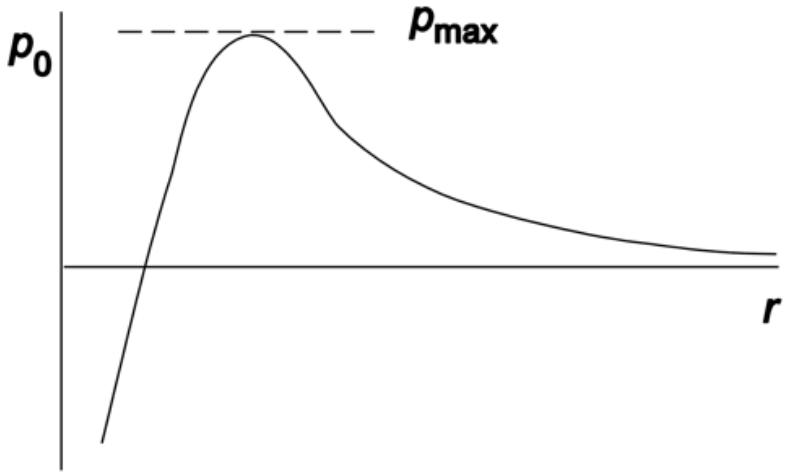


Figure 3.2 : Equilibrium pressure as a function of the cloud radius.

pressure p_0, we can find the relationship between p_0 and the cloud radius, here written as r since we are treating it as an independent variable :

$$P_0 = \frac{3k_BTM_0}{4\pi r^3\mu} - \frac{3}{20\pi}\frac{GM_0^2}{r^4} \tag{3.39}$$

The maximum in the curve (obtained by derivating the previous equation) occurs at a radius :

$$r_{max} = \frac{4}{15}\frac{GM_0\mu}{kT} \tag{3.40}$$

with a maximum pressure of :

$$P_{max} = c_g\left(\frac{k_BT}{\mu}\right)^4\frac{1}{G^3M_0^2} \tag{3.41}$$

where c_g depends on the precise value of the gravitational potential energy term.

We are now able to discuss the stability of a self-gravitating isothermal cloud:

- with a given mass: Suppose that we have a cloud of mass M_0 which is in equilibrium with $r > r_{max}$, if the external pressure is increased by a small amount, the system will now lie above the equilibrium line in Figure 3.2 ; the virial equation shows that the cloud must shrink. Provided $p_0 < p_{max}$, the cloud can collapse to a new equilibrium radius. If $p_0 > p_{max}$, the cloud is not stable and thermal pressure can not halt the collapse.

- with a given internal pressure:, a cloud will become unstable to collapse when its mass exceed :

$$\boxed{M_{BE} = c_g^{1/2} \left(\frac{kT}{\mu}\right)^2 \frac{1}{G^{3/2} P_0^{1/2}} = c_g^{1/2} \frac{a_T^3}{\rho_0^{1/2} G^{3/2}}} \tag{3.42}$$

  The previous equation gives the *Bonnor-Ebert* mass, M_{BE}.

On Figure 3.1, there is a maximum near $\rho_c/\rho_0 \sim 14.1$ when the dimensionless mass is equal to 1.18 :

$$M = 1.18 a_T^4 \frac{1}{G^{3/2} P_0^{1/2}} \sim M_{BE} \tag{3.43}$$

This is also a point limiting stability. If we consider a cloud of given ρ_c/ρ_0 and with an increasing external pressure : this results in an increase of m. For stability the internal pressure of the cloud must increase, but at constant T this requires ρ in the cloud to increase and hence ρ_c since the density is a monotonically decreasing function of radius. This occurs when we are to the left of the peak in Figure 3.1, i.e. we have $\rho_c/\rho_0 < 14.1$.

3.6 Jeans instability

In the following section, we will describe another way of examining the stability of gravitational collapse. Here we start with the equations of hydrodynamics and do a perturbation analysis.
Our initial condition (subscript '0') is to assume that the fluid is stationary with constant density and pressure, i.e.:

$$v_0 = 0 \qquad \rho_0 = \text{constant} \qquad P_0 = \text{constant} \tag{3.44}$$

The unperturbated equations are just the equations of hydrodynamics.
Now we introduce perturbated quantities (subscript '1') :

$$\rho = \rho_0 + \rho_1 \qquad \Phi_g = \phi_0 + \phi_1 \tag{3.45}$$

$$\vec{v} = \vec{v}_0 + v_1 \qquad P = P_0 + P_1 \tag{3.46}$$

As previously, the unperturbated potential is assumed to satisfy :

$$\nabla^2 \phi_0 = 4\pi G \rho_0 \tag{3.47}$$

Unfortunately no solution exists when ρ_0 is a constant. Then, to first order, in small quantities we have :

- Continuity equation:

$$\rho_0(\nabla \cdot \vec{v}_1) = -\frac{\partial \rho_1}{\partial t} \tag{3.48}$$

- Euler equation :

$$\frac{\partial v_1}{\partial t} = -\nabla\phi_1 - \frac{\nabla P_1}{\rho_0} \tag{3.49}$$

- Gravity

$$\nabla^2\phi_1 = 4\pi G\rho_1 \tag{3.50}$$

- We further assume isothermal behaviour :

$$P_1 = a_T^2\rho_1 \tag{3.51}$$

Differentiating the continuity equation with respect to time, we obtain :

$$\frac{\partial}{\partial t}(\nabla \cdot \vec{v}_1) = -\frac{1}{\rho_0}\frac{\partial^2 \rho_1}{\partial t^2} \tag{3.52}$$

and taking the divergence of the Euler equation :

$$\frac{\partial}{\partial t}\nabla \cdot v_1 = -\nabla^2\phi_1 - \frac{a_T^2}{\rho_0}\nabla^2\rho_1 \tag{3.53}$$

Combining the two previous equations gives :

$$\boxed{\left(\nabla^2 - \frac{1}{a_T^2}\frac{\partial^2}{\partial t^2} + \frac{4\pi G\rho_0}{a_T^2}\right)\rho_1 = 0} \tag{3.54}$$

This is similar to the wave equation, and therefore we look for wave-like solutions of the form :

$$\rho_1 \propto e^{i(\vec{k}\cdot\vec{r}-\omega t)} \tag{3.55}$$

which gives a dispersion relation :

$$a_T^2k^2 - \omega^2 = 4\pi G\rho_0 \tag{3.56}$$

N.B.: in the two previous equation k is the wave number, and not the Boltzmann constant
The system is unstable when the modes grow, i.e. $\omega^2 < 0$ and therefore we have a condition for *limiting stability* when ω^2=0 and this defines a critical wave number :

$$k_J^2 = \frac{4\pi G\rho_0}{a_T^2} = \frac{4\pi G\mu}{k_BT}\rho_0 \tag{3.57}$$

or a characteristic wavelength of $\lambda_J = 2\pi/k_J$.

The total mass within a sphere of diameter equal to the Jeans Wavelength of $M_J = \frac{4}{3}\pi\left(\frac{\lambda}{2}\right)^3 \rho_0$ and solving :

$$\frac{\lambda_J}{2} = \frac{3}{\pi^2}\frac{GM_J\mu}{k_B T} \tag{3.58}$$

This is almost exactly the same form as the expression of r_{max} we obtained from the virial analysis, except we now have what we call the Jeans Mass : $M_J \sim M_{BE}$. For $\lambda > \lambda_J$ or $M > M_J$ the modes grow exponentially.
The Jeans Mass is usually defined as :

$$\boxed{\frac{M_J}{M_\odot} = 1.0 \times \left(\frac{T}{10K}\right)^{3/2} \times \left(\frac{n_H}{2 \times 10^{10}\mathrm{m}^{-3}}\right)^{-1/2}} \tag{3.59}$$

Note the strong dependence of the Jeans mass on temperature, hence the importance of gas cooling. The latter can reduce the temperature hence allowing the collapse of less massive clouds and the formation of low mass stars.
It is also interesting to note that in the early Universe, the lack of metals and dust, and (as a consequence) the much reduced molecular gas content, implies much reduced cooling of the gas clouds, which is expected to lead to the formation of hypermassive stars.

3.7 Magnetic fields

One important constituent of the ISM are magnetic fields: these can provide additional forces which can act to stabilise clouds against gravitational collapse. The solution of the Euler equation in the case of a magnetic field is given by :

$$P_0 = \frac{3k_B T M_0}{4\pi r_0^3 \mu} + \frac{1}{4\pi r_0^4}\left(\beta\frac{\Phi_M^2}{2\mu_0} - \frac{3}{5}GM_0^2\right) \tag{3.60}$$

where β is a constant, and Φ_M is the magnetic density field. This relation is similar to the non-magnetic one, but where the effective mass is decreased by the magnetic field. The pressure described by the previous equation will be a monotonically decreasing function of r if :

$$\beta\frac{\Phi_M^2}{2\mu_0} > \frac{3}{5}GM_0^2 \tag{3.61}$$

and the cloud will always be stable if Φ_M is a constant. If the reverse condition is true we have a similar situation to that we have already considered but with an effectively reduced gravitational mass for the cloud.

Cloud type	n_{tot} (10^6 m^{-3})	L (pc)	T (K)	M $M_\odot$	B (nT)
Giant molecular cloud	100	50-500	15	10^5	1?
Dark cloud complex	500	10	10	10^4	1?
Individual dark cloud	10^3	2	10	30	2-10
Dense core	10^4	0.1	10	10	2-10

Table 3.1: Properties of several Cloud types

3.8 Application to molecular clouds

We have already seen that the molecular phase of the ISM has much higher densities and lower temperatures than either the atomic or ionised phases (Tab. 2.2) and discussed the structure of Giant Molecular Clouds. In the following, we apply the idea we just developed in this chapter to analyse the stability of molecular clouds (Tab. 3.1).

We described Giant Molecular Clouds as swarms of more coherent clumps. The Jeans mass for gas with n_H ∼2000 and T ∼10K is ∼3M$_\odot$. This is two orders of magnitude below the masses of the individual clouds and of order the mass of typical dense cores. There must be an additional form of support. Zeeman splitting [3] provides a method for measuring the magnetic fields although this has only proved possible towards a handful of dark clouds at the present time, possible fields strengths are also shown in Tab.3.1.

From the magnetic virial equation we can find the maximum cloud mass which could be supported against its own self-gravity by magnetic pressure alone. We have :

$$M^2 = \frac{5}{3G}\beta\frac{\pi^2 r^4 B^2}{2\mu_0} \tag{3.62}$$

or

$$M \approx \left(\frac{B}{nT}\right)\left(\frac{r}{pc}\right)^2 M_\odot \tag{3.63}$$

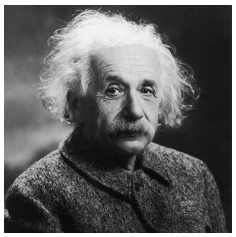
Inserting the numbers from Tab.3.1 it suggests that most dark clouds can be stabilised by magnetic effects :

- for dense cores M ∼M$_J$
- the density ratio in dense cores can also be measured an ρ_c/ρ_0 ∼10 typically.

The analyse ignores one further possible source of cloud support which is bulk random motions of gas-turbulence.

[3] Effect of splitting spectral line into several components with a static magnetic field

4. Gravitational collapse



We can not solve our problems with the same level of thinking that created them

— Albert Einstein

In the following chapter, we will consider the collapse of self-gravitating clouds under gravity. We will first consider this in the context of a cloud (or dense core) within a galaxy, but we shall see later that this gives a great deal of insight to the collapse of gas on galactic-scales.

4.1 Free-fall time

By definition, the free-fall time (t_{ff}) is the characteristic time that would take a body to collapse under its own gravitational attraction, if no other forces existed to oppose the collapse. As such, it plays a fundamental role in setting the timescale for a wide variety of astrophysical processes—from star formation to helioseismology to supernovae—in which gravity plays a dominant role.

Consider an initially homogeneous spherical cloud of gas of density ρ_0, radius R and mass M_0. The cloud has no internal pressure throughout the collapse. Gas molecules initially at radius r_0 will have a mass of gas M within this radius and during collapse this remains constant, therefore from Newton gravity the equation of motion for these molecules is :

$$\frac{\partial^2 r}{\partial t^2} = -\frac{GM_r}{r^2} \tag{4.1}$$

Therefore :

$$\frac{\partial}{\partial t}\left(\frac{\partial r}{\partial t}\right) = -\frac{GM_r}{r^2} \tag{4.2}$$

Integrating over dt, gives :

$$\frac{1}{2}\left(\frac{\partial r}{\partial t}\right)^2 = \left[\frac{GM}{r}\right]_{r_0}^{r} = \frac{GM}{r_0}\left(\frac{r_0}{r} - 1\right) = \frac{4\pi}{3} r_0^2 \rho_0 G \left(\frac{r_0}{r} - 1\right) \tag{4.3}$$

By making the substitution $r = r_0 sin^2\theta$, we can integrate this equation and find that the gas arrives at the centre at the free-fall time :

$$\boxed{t_{ff} = \left(\frac{3\pi}{32 G \rho_0}\right)^{1/2}} \tag{4.4}$$

This is the same result for every starting point in the cloud and therefore can be applied to any radius. In terms of the initial mass and radius of the whole cloud, we get :

$$t_{ff} = \left(\frac{\pi^2}{8}\frac{R^3}{GM_0}\right)^{1/2} \approx \left(\frac{R^3}{GM_0}\right)^{1/2} \tag{4.5}$$

4.2 Inside-out collapse

We begin our discussion of collapse and protostellar formation by considering the collapse of an isolated cloud without the presence of magnetic fields. We further simplify the problem by assuming a spherical symmetry and isothermal behaviour. We assume that there is a central sink for inflowing material - the growing central condensed object (e.g. a protostar). The equations governing the dynamics of the problem are just the Euler equation and equation of continuity. The radial equations are:

$$\frac{\partial v_r}{\partial t} + v_r \frac{\partial v_r}{\partial r} = -\frac{a_T^2}{\rho}\frac{\partial \rho}{\partial r} - \frac{GM_r}{r^2} \tag{4.6}$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{r^2}\frac{\partial r^2 \rho v_r}{\partial r} = 0 \tag{4.7}$$

where $\rho = \rho(r,t)$ and

$$M_r(t) = \int_0^r 4\pi r^2 \rho(r,t)\, \mathrm{d}r \tag{4.8}$$

Differentiating eq.4.8 and using the equation of continuity eq.4.7, we find :

$$\frac{\partial M_r}{\partial t} = -4\pi r^2 \rho v_r \tag{4.9}$$

4.2.1 Similarity Analysis

We now transform these equations to dimensionless form. The independent variables are r and t, the constants are G and a_T, and the variables are $\rho(r,t)$, $v(r,t)$ and $M(r,t)$.

- The only way to form a dimensionless length is :

$$x = \frac{r}{a_T t} \tag{4.10}$$

- The variables are then given by :

$$M_r(r,t) = \frac{a_T^3 t}{G} m(x) \tag{4.11}$$

$$\rho(r,t) = \frac{1}{4\pi G t^2} \alpha(x) \tag{4.12}$$

$$v(r,t) = a_T \beta(x) \tag{4.13}$$

- We are effectively doing a transformation of coordinates from (r,t) to (x,t) in which we find the time dependence disappears

- Since $\mathrm{d}x = \frac{1}{a_T t}\mathrm{d}r - \frac{r}{a_T t^2}\mathrm{d}t$, we have :

$$\left.\frac{\partial}{\partial r}\right|_t = \frac{1}{a_T t}\frac{\partial}{\partial x} \tag{4.14}$$

$$\left.\frac{\partial}{\partial t}\right|_r = \left.\frac{\partial}{\partial t}\right|_x + \left.\frac{\partial x}{\partial t}\right|_r \frac{\partial}{\partial x} = \left.\frac{\partial}{\partial t}\right|_x - \frac{x}{t}\frac{\partial}{\partial x} \tag{4.15}$$

  Using these definitions the above equations reduce to

$$m = x^2 \alpha (x - \beta) \tag{4.16}$$

$$[(x-\beta)^2 - 1]\frac{1}{\alpha}\frac{d\alpha}{dx} = \left[\alpha - \frac{2}{x}(x-\beta)\right](x-\beta) \tag{4.17}$$

$$[(x-\beta)^2 - 1]\frac{d\beta}{dx} = \left[\alpha(x-\beta) - \frac{2}{x}\right](x-\beta) \tag{4.18}$$

  These equations must be solved numerically, but we can learn a great deal from their form.

- an exact solution of these equations is the singular isothermal sphere, as we expect :

$$\beta = 0 \qquad \alpha = \frac{2}{x^2} \qquad m = 2x \tag{4.19}$$

  This is a static, but unstable solution.

- There is a critical point for the flow when both side of the previous differential equations vanish, this occurs when :

$$x - \beta = 1, \qquad \alpha = \frac{2}{x} \tag{4.20}$$

- This critical point condition is satisfied by the singular isothermal sphere for $x = 1$:
  - A possible form of solution is that there is a change in character at $x = 1$
  - For $x > 1$ the solution is the singular sphere
  - For $x < 1$ the solution has explicit infall
  - This requires that the initial condition for the cloud is that of the singular isothermal sphere ; if the initial density exceeds that of the singular isothermal solution then we expect some infall at all radii, but with a significant increase at $x \sim 1$
- this characteristic form of solution leads to the concept of *inside-out collapse* with the inner regions collapsing and the outer regions stationary until the transition point expands and gas joins the infall : the transition point expands into the cloud with $r = a_T t$, i.e. at the speed of sound.

The full equations have to be solved numerically. The initial condition we choose: (i) we consider an equilibrium isothermal sphere and (ii) we use the boundary conditions appropriate to that initial state - for example constant external pressure. We also assume that there is a sink for matter reaching the origin - this will turn into a protostar. A physically interesting case to consider is that of a cloud which is marginally unstable to begin, i.e. with a mass slightly larger than the Bonnor-Ebert / Jean Mass and perturb the initial state slightly and follow the evolution (Figure 4.1).

4.2.2 Physics Analysis

Using the above analysis as a guide, we can develop a good physical model for the collapse. Important is the concept of the critical point at which the nature of the solution changes at $x \sim 1$.
We consider an initial condition for the cloud equal to the singular isothermal sphere and consider the propagation of the transition point into the cloud :

- The transition point moves outwards as a rarefaction wave with only the gas inside of the radius $R_{ff} \approx a_T \times t$ moving inward. We have *inside-out collapse*.

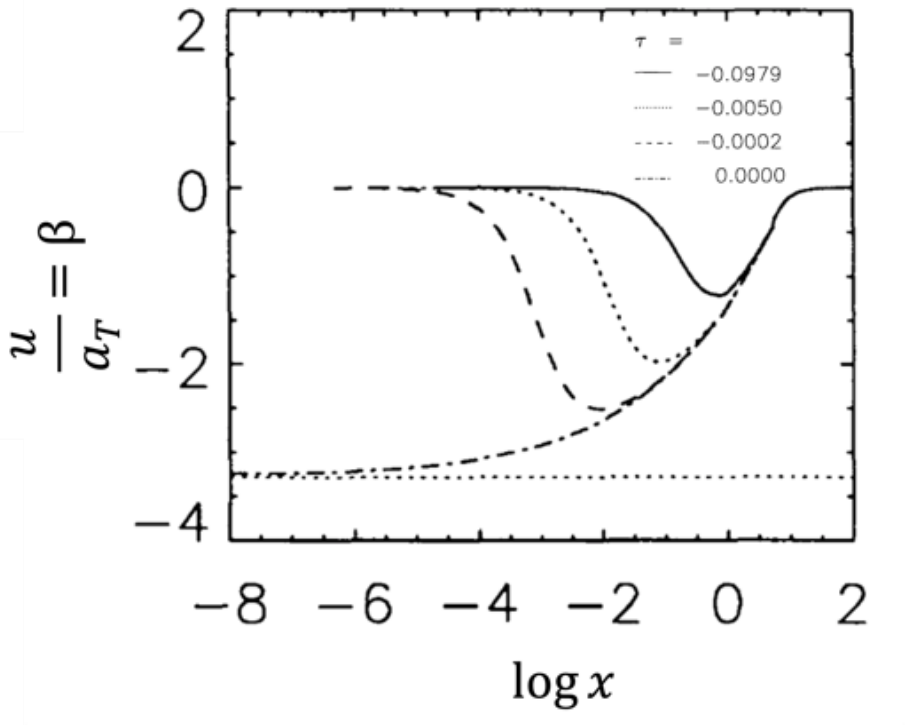


Figure 4.1 : Velocity during the collapse of an isothermal sphere with mass slightly above the Jeans mass. $\tau = 0$ (the dotted-dashed line) signifies the start of the creation of the protostar as mass starts to flow into the sink. The time τ is given as a function of t_{ff}. From Foster & Chevallier 1993

- After a short fraction of a free-fall time a large fraction of the gas within this radius is moving supersonically with the velocity v increasing to the center - $r/|v|$ is less than the sound crossing time
- Gas is falling onto a growing central object - the protostar - mass $M_\star$; close to this protostar gas is approximately in free fall with $v_{FF} \approx (\frac{2GM_\star}{r})^{1/2}$
- At the transition point, the gas moves approximately sonically :

$$v_{ff} \approx a_t \text{ and } a_T \sim \frac{M_\star G}{R_{ff}} \tag{4.21}$$

- The rate of growth of $M_\star$ is determined by accretion at a rate :

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \lim_{r\to 0} -4\pi r^2 v\rho \tag{4.22}$$

- Transition to the free-fall collapse occurs at radius R_{ff} such that $v_{ff} \approx a_T$, i.e. $a_T^2 \sim \frac{M_\star G}{R_{ff}}$.
- The solution shows near constant accretion with $\frac{\mathrm{d}M}{\mathrm{d}t} \approx \frac{a_T^3}{G}$

indeed if we assume a constant accretion rate then $M_\star = \frac{\mathrm{d}M}{\mathrm{d}t} \times t$ and therefore :

$$\frac{\mathrm{d}M}{\mathrm{d}t} \approx \frac{M_\star}{t} \approx \frac{a_T^2}{G}\frac{R_{ff}}{t} \approx \frac{a_T^3}{G} \tag{4.23}$$

The density profile in this region must satisfy :

$$\rho = \frac{\frac{\mathrm{d}M}{\mathrm{d}t}}{4\pi r^2|v|} = \frac{\dot{M}}{4\pi r^2 v_{ff}} = \frac{\dot{M}}{4\pi r^{3/2}\sqrt{2GM_\star}} \tag{4.24}$$

A distance $R \gg R_{ff}$ the solution must approach that of the isothermal sphere with a large density contrast, i.e. $\rho \propto 1/r^2$, the singular sphere solution.
The change in slope from the outer singular isothermal sphere ($\rho \propto r^{-2}$) to the inner free-fall region ($\rho \propto r^{-3/2}$) implies a "rarefaction wave".
Inserting values, the accretion rate for the growth of the protostar is :

$$\dot{M} \approx 2 \times 10^{-6} \left(\frac{T}{10^6 K}\right)^{3/2} M_\odot yr^{-1} \tag{4.25}$$

5. From gas cloud to collapsed object

That is one of the things that has come out of the discovery of pulsars - more knowledge about the space between the stars.

— Dame Susan Jocelyn Bell Burnell

In previous chapter, we have considered the equilibrium and then collapse of an isolated gas cloud. We now apply this physics to the formation of a central collapsed object forming at the centre of the cloud. Although we will use the formation of star, or more strictly a protostar as our example in this chapter, many of the ideas will be used again when we consider the formation of other collapsed objects in particular galaxies.

5.1 Basic physics of object formation

If the cooling time in the gas is much shorter than the collapse time ($\tau_c < t$), the system will evolve approximately isothermally. The shortest collapse time is the free-fall time, therefore we can say that if the cooling time is shorter than the free-fall time ($\tau_c < t_{ff}$), then the cloud will collapse approximately isothermally. From the way the Jeans mass depends on the temperature T and density ρ ($M_J \propto T^{3/2}\rho^{-1/2}$) we can see that the Jeans Mass decreases. This results in smaller regions of the cloud having the ability to collapse in a process called fragmentation, which we will discuss in the next chapter.

The increasing density has another consequence however :

- The optical depth is given by $\tau = \alpha R = \rho\kappa R$ (eq. 2.28). Assuming that the mass in the collapsing object is constant : $R \propto \rho^{-1/3}$ and hence $\tau \propto \rho^{2/3}$.

- At some point, the collapsed object, or as we will now call it the core, becomes optically thick ($\tau > 1$), cooling is then very inefficient and the core ceases to cool.
- It will in general continue to collapse as more mass is added due to accretion from the surrounding cloud because of the inside-out collapse. The temperature of the core now rises as the collapse proceeds adiabatically.

5.2 Evolution of the first core

From the discussion in the previous section, we see that the collapse of an isolated gas cloud results in what we call the first core, which is when the initial collapse ceases to be isothermal. We can apply the virial theorem to this first core, which is mainly composed of molecular hydrogen. Taking the gravitational potential energy to be $\sim -\frac{3}{5}\frac{GM^2}{R}$ and neglecting external pressure and magnetic fields in comparison to the thermal and gravitational terms we can estimate the temperature of the core:

$$T \approx \frac{\mu}{5k_B}\frac{GM}{R} \approx 850\left(\frac{M}{5\times 10^{-2}M_\odot}\right)\left(\frac{R}{5\,\mathrm{AU}}\right)^{-1} K \tag{5.1}$$

The accretion rate onto this core is, as we previously calculated:

$$\dot{M} \approx 2\times 10^{-6}\left(\frac{T}{10^6 K}\right)^{3/2} M_\odot\,\mathrm{yr}^{-1} \tag{5.2}$$

Further mass addition, together with decreasing R, means that the temperature soon exceeds 200 K, and collisional dissociation of H_2 begins (although k_BT is much less than the dissociation potential). This phase change is crucial ; further mass addition happens with a very much slower rise in temperature. The central region of atomic gas has a significant density gradient and eventually undergoes collapse to form a very dense central core (c.f. instability of an isothermal sphere with large density contrast).

5.3 Structure around the protostar

The collapse of the first core gives a protostar of mass $\sim 0.1M_\odot$ and radius of several $R_\odot$. From the virial theorem this gives T $\sim 10^5$K, the density is now $\sim 10^{-2}$ g cm^{-3} and this really can be considered a protostar. We will now consider the environment of this protostar.

5.3.1 Accretion luminosity

Accretion occurs onto the protostar via the inside-out collapse of the cloud. We start by considering the energetics of the protostar in more detail. Since the protostar forms from cold, low density gas, we can take the initial thermal and mechanical energy of the gas, then energy balance requires :

$$U_{gr} + U_{th} + E_p + L_r t = 0 \tag{5.3}$$

where U_{gr} is the gravitional potential energy of the protostar ($U_{gr} \approx -GM_\star^2/R_\star$) ; its internal energy from the virial theorem is $U_{th} = -U_{gr}/2$; E_p is the enery necessary to change the phase of the gas (i.e. dissociate and ionise the original molecular gas), and L_r is the mean energy radiated over the formation time of the protostar t.
For an initial hydrogen fraction X and helium fraction Y, we have :

$$E_p = \frac{XM_\star}{m_H}\left[\frac{E_d(H)}{2} + E_i(H)\right] + \frac{YM_\star E_i(He)}{4m_H} \tag{5.4}$$

where E_d(H) is the dissociation energy of H_2 (E_d(H) =4.2 eV), E_i(H) =13.6 eV and E_i(He)=75 eV are the total ionisation potential of H and He respectively.
If L_r=0, the protostar would have a radius of :

$$R_{max} = \frac{GM_\star^2}{2E_p} = 60\left(\frac{M_\star}{M_\odot}\right) R_\odot \tag{5.5}$$

which is much larger than the observed size of T Tauri stars which are the immediate descendants of the protostars we are considering. This implies that $L_r \times t$ is comparable to the first term, i.e. most of the excess energy is radiated away ; L_r must be close to the accretion luminosity :

$$L_r \approx L_{acc} = \frac{GM_\star \dot{M}}{R_\star} = 61\left(\frac{\dot{M}}{5\times10^{-5}M_\odot \mathrm{yr}^{-1}}\right)\left(\frac{M_\star}{M_\odot}\right)\left(\frac{R_\star}{5R_\odot}\right) \tag{5.6}$$

5.3.2 Accretion shock and dust envelope

The accreting gas is infalling at close to the free-fall velocity :

$$v_{ff} = \left(\frac{2GM_\star}{R_\star}\right)^{1/2} = 280\left(\frac{M_\star}{M_\odot}\right)^{1/2}\left(\frac{R_\star}{5R_\odot}\right)^{-1/2} \mathrm{km\ s^{-1}} \tag{5.7}$$

This greatly exceeds the sounds speed in the gas and a strong shock forms with $v_s \approx v_{ff}$ giving a post-shock temperature of order 10^6 K ; at this temperature

there is significant UV and soft X-ray production. This whole region is optically thick and radiates with an effective temperature such that :

$$L_{acc} \approx 4\pi R_\star^2 \sigma_B T_{eff}^4 \tag{5.8}$$

or

$$T \approx \left(\frac{GM_\star \dot{M}}{4\pi\sigma_B R_\star^3}\right)^{1/4} = 7300 \left(\frac{\dot{M}}{5\times10^{-5} M_\odot yr^{-1}}\right)^{1/4} \left(\frac{M_\star}{M_\odot}\right)^{1/4} \left(\frac{R_\star}{5R_\odot}\right)^{-3/4} K \tag{5.9}$$

Surrounding the star, the gaseous envelope also contains dust. The UV flux produced at the stellar surface is able to vaporize dust grains within a region called the *opacity gap* out to a radius known as the dust destruction front.
Outside of this radius the dust absorbs the radiation and re-radiates. In the dusty layer, the dust temperature must drop until the layer becomes optically thin to the re-emission of infrared radiation from the dust. This occurs at a radius R_{phot} such that :

$$\rho\kappa R_{phot} = 1 \tag{5.10}$$

and

$$L_{acc} = 4\pi R_{phot}^2 \sigma_B T_{phot}^4 \tag{5.11}$$

5.4 Evolution of the protostar

We now turn our attention to the structure and evolution of the protostar itself, also referred as *stellar evolution*. This could be the topic of an entire course, and we will only discuss it briefly in the following. Few analytical calculations are possible, but we must rely largely on the results of numerical simulations to guide our understanding of the physics.

5.4.1 Protostellar structure during accretion

The structure of the protostar will be governed by the equations of hydrostatic equilibrium plus equations describing the thermal structure of the protostar - these are just the equations of stellar structure you met last year, although the boundary conditions are now different as discussed later.
The first equations are identical to those we met in our analysis of hydrostatic equilibrium of a cloud except that everything is now a function also of time :

$$\frac{\partial P}{\partial r} = -\frac{GM\rho}{r^2} \tag{5.12}$$

$$\frac{\partial M}{\partial r} = 4\pi r^2 \rho \tag{5.13}$$

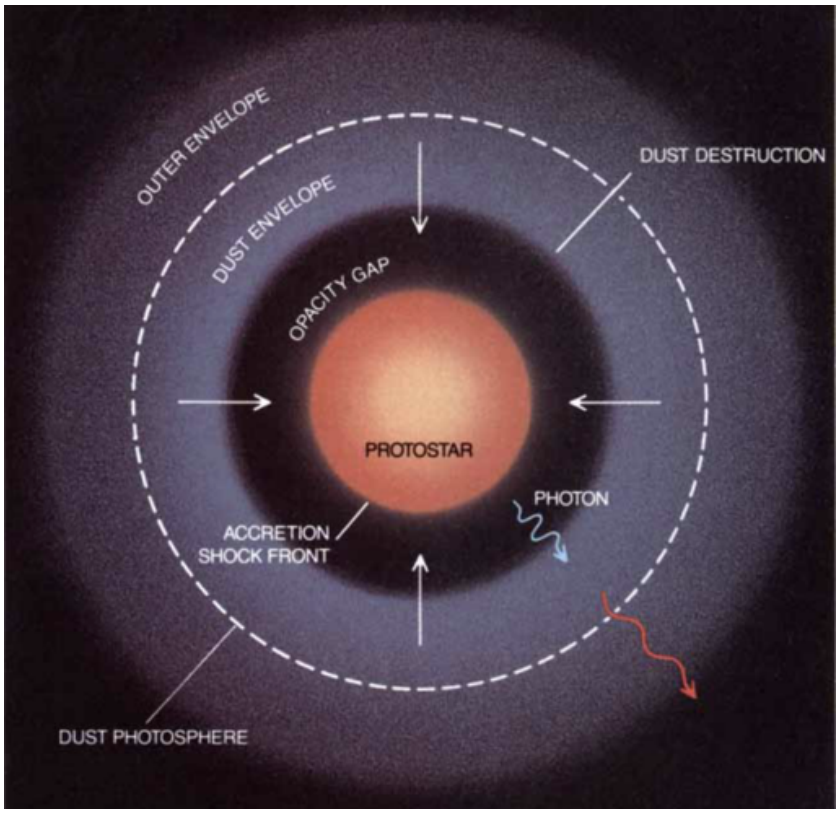


Figure 5.1 : The environment of a protostar. From Stahler 1991

we again assume an ideal equation of state :

$$P = \frac{\rho k T}{\mu} \tag{5.14}$$

We then apply the first law of thermodynamics to obtain an equation for the heat production :

$$\rho T \frac{\partial S}{\partial t} = \rho\epsilon - \nabla \cdot \vec{F} \tag{5.15}$$

where F is the radiative flux defined as :

$$F = \frac{L}{4\pi r^2} \tag{5.16}$$

hence :

$$\frac{\partial L}{\partial r} = 4\pi r^2 \rho\epsilon - 4\pi r^2 T \frac{\partial S}{\partial t} \tag{5.17}$$

This differs from the equation you had last year by the inclusion of the time derivative of the entropy. If the heat is transported by radiative diffusion we also have:

$$\frac{L}{4\pi r^2} = -\frac{4acT^3}{3\rho\kappa}\frac{\partial T}{\partial r} \tag{5.18}$$

These equations must be combined with a set of boundary conditions. In outline these are:

- The mass and luminosity go to zero as $r \to 0$
- The surface luminosity of the protostar itself is given by $L_\star - L_{acc}$
- The surface pressure of the protostar must balance the momentum flux, or *ram pressure* of the infalling gas which is $\sim \rho v_{ff}^2$ giving :

$$P(r_0) = \frac{\dot{M}}{4\pi}\left(\frac{2GM_\star}{R_\star^5}\right)^{1/2} \tag{5.19}$$

  The results of a numerical integration give for the accretion rate $\dot{M} = 1 \times 10^{-5} M_\odot \mathrm{yr}^{-1}$. This is shown in Figure 5.2.

- As parcels of gas fall onto the protostar they add an extra layer of material with both mass and entropy
- As the gas falls onto the star the energy density of $\sim -GM_\star\rho/R$ is thermalised at the shock and the entropy of this gas is just proportional to $\sim -GM_\star/R_\star$

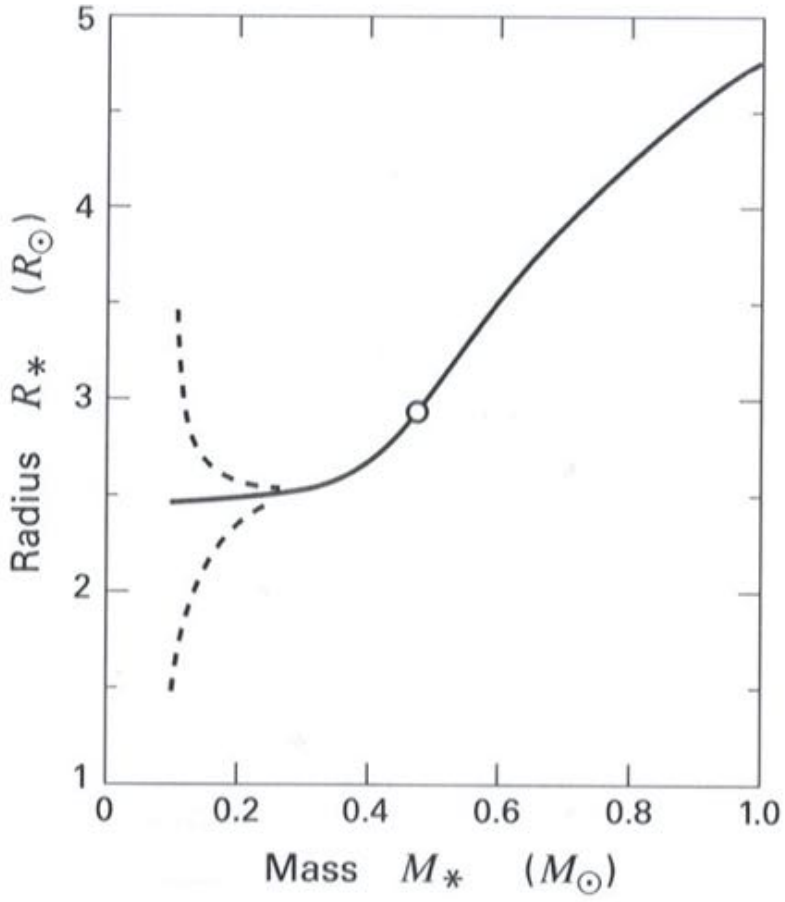


Figure 5.2 : Mass-radius relation. The three curves have different initial values of $R_\star$ for a starting mass of $0.1M_\odot$

- if initially $R_\star$ is large, the entropy of the gas added to the protostar is low and the protostar shrinks under gravity
- The converse is true if $R_\star$ is initially small
- The protostar is characterised by its entropy profile $S(r)$
- The results show that $M_\star/R_\star$ is an increasing function in the early stages which gives an increasing entropy distribution with radius.

5.4.2 Onset of deuterium burning and convection

Convective stability

Consider a fluid element which moves a small distance through the atmosphere Δr so that it remains in pressure balanced with the surrounding gas :

- The element expands adiabatically to a lower density
- For stability, the density of this element must be greater than the density of the surrounding gas
- If the entropy is increasing with radius the element is of lower entropy than the surrounding
- for an ideal gas :

$$S = c_v \log(p/\rho^\gamma) \tag{5.20}$$

where $\gamma = \frac{c_p}{c_v}$ is the specific heat ratio of a gas.

hence ρ increases as the entropy decreases. The atmosphere is convectively stable therefore if :

$$\frac{\partial s}{\partial r} > 0 \tag{5.21}$$

Deuterium burning

At a temperature of about 10^6 K the first nuclear fuel to ignite is deuterium :

$$^2\mathrm{H} + {}^1\mathrm{H} \rightarrow {}^3\mathrm{He} + \gamma \tag{5.22}$$

which releases 5.5 MeV. The heating rate due to this process is very dependent on temperature with $\epsilon_D \propto T^{11.8}$! The protostar is unable to effectively transport the large luminosity produced in the core via radiative transport, the core heats up, reversing the entropy gradient and the protostar becomes convective.
We can calculate the maximum energy flux which can be carried by a pure radiative

flux. This occurs when the entropy is constant. From the equations of protostellar structure :

$$L_{crit} = 4\pi r^2 \frac{4acT^3}{3\rho\kappa}\left(\frac{\partial T}{\partial r}\right)_s = -\frac{GM16\pi acT^3}{3\rho\kappa}\left(\frac{\partial T}{\partial p}\right)_s \tag{5.23}$$

- Although the amount of deuterium is small, convection helps to bring new fuel to the core from the accreting gas
- This deuterium burning phase acts as a thermostat - the *deuterium thermostat*
- Any rise in $M_\star/R_\star$ increases the stellar entropy which, via convection, increases T_c ; this leads to a substantial increase in ϵ_D which inflates the star, reducing $M_\star/R_\star$.

5.4.3 Deuterium shell burning

As the protostar mass continues to grow via accretion the energy production from the deuterium burning remains approximately constant determined by the rate of supply of new fuel from the accreting gas.

- L_{crit} however rises
- we can show that L_{crit} scales as $M_\star^{11.2}R_\star^{-1/2}$; eventually $L_{crit} = L_D$ and radiative energy transport can again remove energy from the core. The results of more detailed calculations are shown on Figure 5.3
- Without convection new deuterium accreted onto the protostar accumulates in a shell
- Effectively the radiative transport acts as a barrier preventing deuterium reaching the core
- deuterium in the core is quickly depleted
- Eventually the temperature of this shell reaches 10^6K and the shell ignites
- The hot outer shell leads to a substantial increase in the stellar radius (Figure 5.4)

5.4.4 Contraction and hydrogen burning

The final stage of protostellar evolution we shall follow is the contraction of the star. Without deuterium burning in the core, the self-gravity of the protostar drives

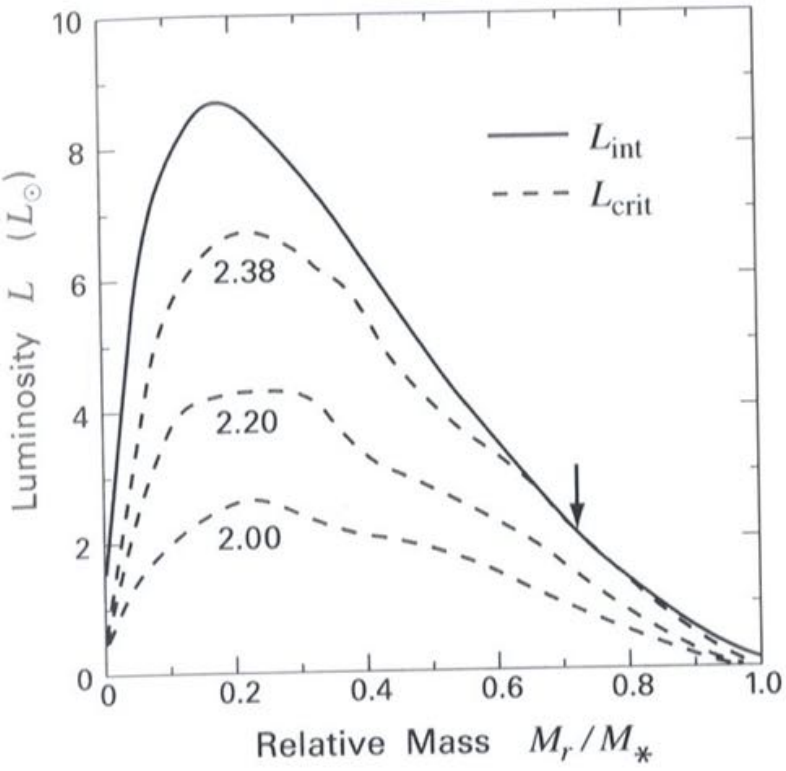


Figure 5.3 : End of the convection phase for different stellar masses. The black arrow shows the radiative barrier

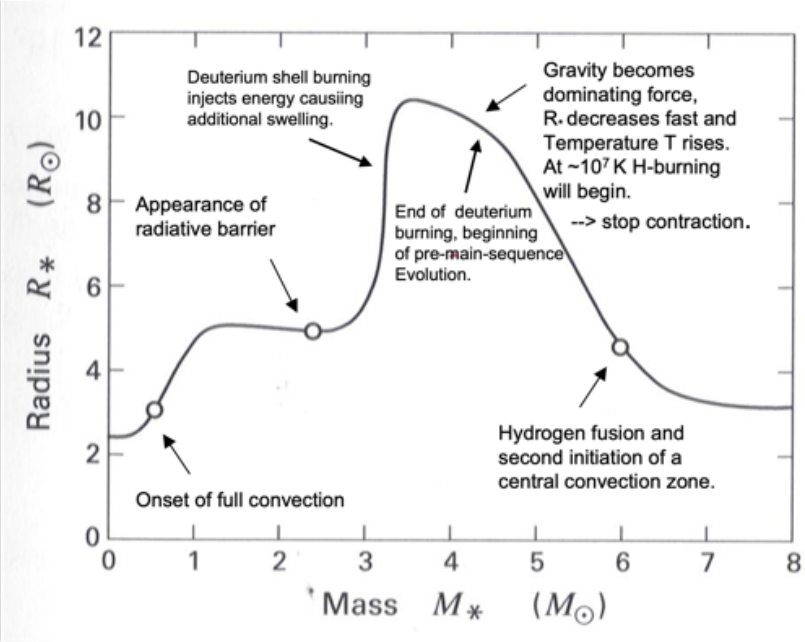


Figure 5.4 : Radius mass relation accreting at $1 \times 10^{-5} M_{\odot} \mathrm{yr}^{-1}$. The circles mark the onset of full convection, the appearance of the radiative barrier and the onset of hydrogen burning.

the gravitational contraction of the star. The rate at which the star contracts is determined by the rate at which the star loses internal energy due to radiation. This is the *Kelvin-Helmholtz timescale:*

$$t_{KH} = \frac{GM_\star^2}{R_\star L_\star} = 3 \times 10^7 \left(\frac{M_\star}{M_\odot}\right)^2 \left(\frac{R_\star}{R_\odot}\right)^{-1} \left(\frac{L_\star}{L_\odot}\right)^{-1} \text{yr} \tag{5.24}$$

As the contraction proceeds the core temperature continues to rise until eventually 10^7K, hydrogen burning commences and halts the contraction. Further temperature rise enables the CNO cycle. At this stage the protostar is regarded as a pre-main-sequence star.

6. Galaxies and star-formation on galactic scales

Equipped with his five senses, man explores the universe around him and calls the adventure Science

— Edwin Hubble

In this chapter, we will review the properties of galaxies in the local Universe, and will discuss the star-formation on galactic scales.

6.1 Properties of Galaxies in the Local Universe

In this section, we will build on the material type you met in relativistic astrophysics and cosmology course.

6.1.1 The Galaxy Zoo

By observing the galaxies in the 1920's, Edwin Hubble identified several morphology: elliptical, spiral, barred spiral and irregular (Figure 6.1). In the following, we will briefly give the properties of each type of galaxies.

Ellipticals - or early type galaxies

The elliptical galaxies are classified following the ratio between their major (a) and minor (b) axis, and named as En, with n defined as :

$$n = 10\frac{a-b}{a} \tag{6.1}$$

They are gas poor galaxies, with no (or little) star formation. Their stellar population is generally dominated by old stars, and their stellar mass is ranging from $10^{11}\mathrm{M}_\odot$ to $10^{13}\mathrm{M}_\odot$. However, dwarf elliptical galaxies have masses of a few $10^{10}\mathrm{M}_\odot$ or less.

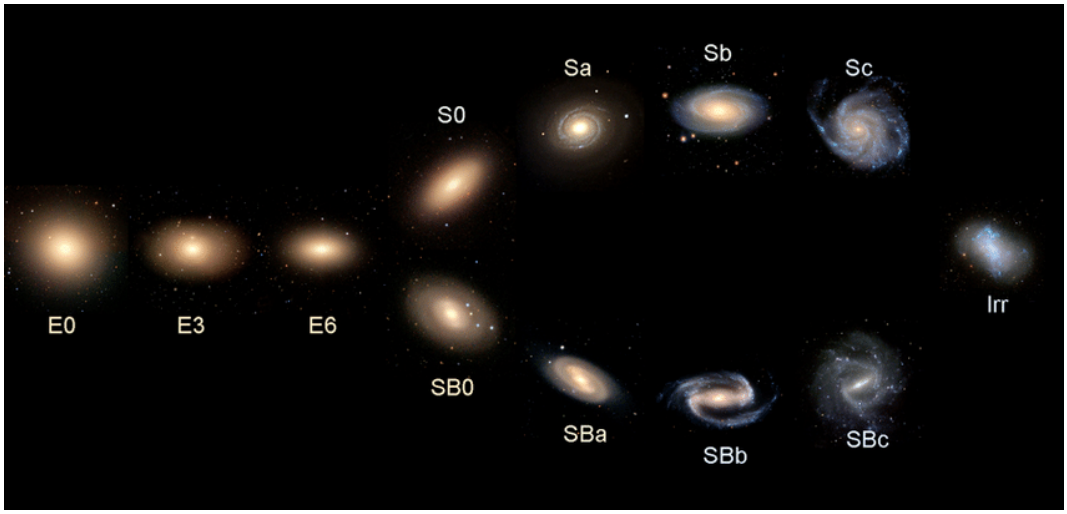


Figure 6.1 : The Hubble sequence with the different types of galaxies : elliptical, spiral, barred spiral and irregular. From Cui et al. (2014)

The surface brightness of an elliptical galaxy is given by :

$$I(R) = I_o \exp\left[-\left(\frac{R}{a}\right)^{1/4}\right] \tag{6.2}$$

Spirals - or late type galaxies

A spiral galaxy is composed of a bright bulge and a disk. They could be either spiral with no bar (named S) or barred spiral (named SB). From Sa to Sc (or SBa to SBc), the openness of arms increases, the prominence of the bulge decreases and the gas content increases. They are more gas rich than ellipticals with at least some star-formation on-going. The gas is both low density neutral hydrogen and dense molecular hydrogen. The fractional mass of neutral hydrogen gas to total (i.e. M(HI)/M) is less than 0.03 for a Sa and goes up to 0.1 for Sc. In the description of a spiral galaxy, we can also include the presence of rings identified with "r" after the name of the spiral (Figure 6.2) and the luminosity of the arms ("I" for well defined arms to "V" to less luminous fuzzy arms ; Figure 6.3).

Firstly we describe the structure of spiral (or barred spiral) galaxy in terms of a near spherical bulge which has similar properties to an elliptical, plus a disc. The simplest description of the system requires a two dimensional model. A spiral seen face-on has a surface brightness profile which follows an exponential distribution

Figure 6.2 : Color image of the galaxy NGC 6782 classified as SB0(r)I. Source : NASA

Figure 6.3 : Color image of the galaxy M83 classified as SBc(s)II. Source : NASA

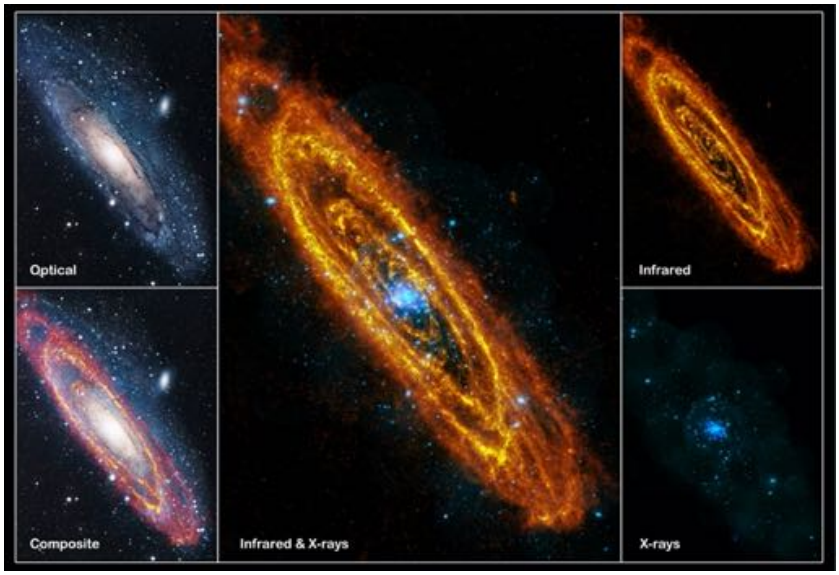


Figure 6.4 : Color image of the galaxy M31 as seen by the Hubble Space Telescope at optical wavelength(top left), by the Herschel telescope in infrared (top right), by the Chandra telescope in X-ray (bottom right). Source : ESA

such as :

$$I(R) = I_o \exp(-\frac{R}{a}) \tag{6.3}$$

The vertical structure in the disk also follows an exponential distribution with a scale height[1] we will call h. This implies an overall mass distribution in the disc of:

$$\rho(R, z) = \rho_0 \exp(-R/a) \exp(-|z|/h) \tag{6.4}$$

The most obvious feature of spiral galaxies, their spiral arms, are a complicated dynamical feature which we will return to later in the course. Here we note that the spiral arms are delineated in many different ways : they are seen in the stellar distribution, atomic and molecular gas, young stars, magnetic fields, etc... (Figure 6.4)

6.1.2 The galaxy Luminosity Function

By definition, the galaxy Luminosity Function is the distribution in luminosity (L) of the number density of galaxies ($\phi(L)$) at a given redshift. Its form has been

[1]bu definition, the scale height is a measure of the decrease of something that falls off exponentially by height, specifically, the height over which it falls by a factor of e (∼2.718)

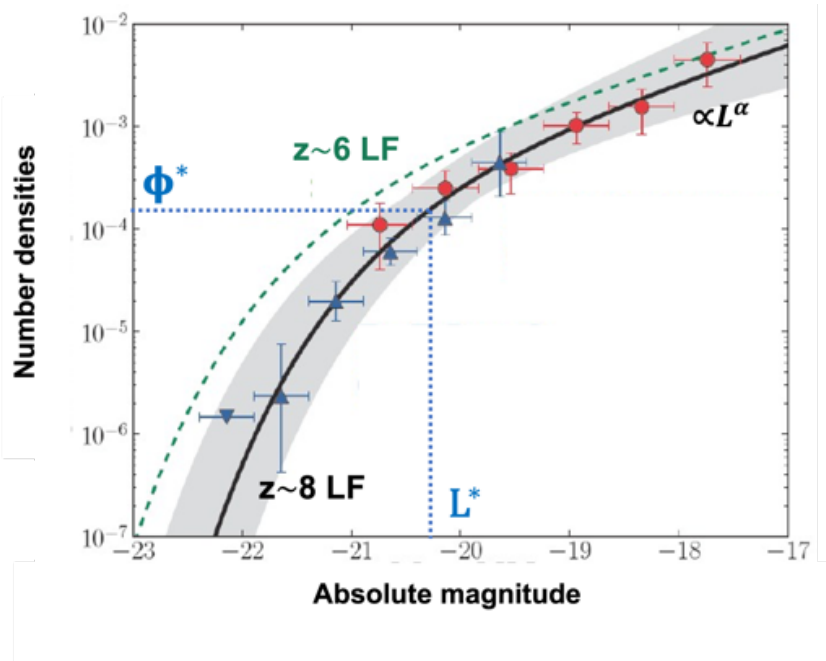


Figure 6.5 : UV Luminosity Function estimated at $z \sim 8$ (black line) compared to the UV Luminosity Function at $z \sim 6$ (green dashed line). The three parameters of the LF are also indicated. Adapted from Schmidt et al. (2014)

described empirically by Schechter (1976) :

$$\Phi(L/L^\star)d(L/L^\star) = \Phi^\star(L/L^\star)^\alpha \exp(-(L/L^\star))d((L/L^\star) \tag{6.5}$$

that we can re-write as :

$$\Phi(x)dx = \Phi^\star x^\alpha e^{-x} dx \tag{6.6}$$

where $\Phi^\star$ and $L_\star$ are the density and luminosity where there is a change in the shape of the function (from exponential to linear), and α is the slope at the faint-end of the Luminosity Function.

To determine the number density of galaxies, $\Phi(L)$ we need a redshift survey complete to L_{lim} at a given redshift z. The number of galaxies per given luminosity is not similar to $\Phi(L)$ because of the Malmquist bias which is an effect in observational astronomy which leads to the preferential detection of intrinsically bright objects. In statistics, this bias is referred to as a selection bias or data censoring. It affects the results in a brightness-limited survey, where stars below a certain apparent brightness cannot be included. Since observed stars and galaxies appear

dimmer when farther away, the brightness that is measured will fall off with distance until their brightness falls below the observational threshold. Objects which are more luminous, or intrinsically brighter, can be observed at a greater distance, creating a false trend of increasing intrinsic brightness, and other related quantities, with distance. This effect has led to many spurious claims in the field of astronomy. Properly correcting for these effects has become an area of great focus.

To test the integrity of a survey regardless of the luminosity bias, we can use the volume-luminosity test (also called the V/V_{max} method) develop by Schmidt et al. (1968). The $< V/V_{max} >$ method tests whether the distribution of objects is uniform within the volume of space defined by the observational selection criteria. Among other advantages, it is suitable for samples containing few objects and allows to combine samples of sources obtained with different selection criteria. Historically, it has been employed to study the space distribution of quasars and to assess the cosmic evolution of their population. For a uniform population of sources with measured fluxes S, V/V_{max} are the ratios of the volume V within which each source is distributed to the maximum volume V_{max} within which each source could still be detected (which is individually defined by the sample selection flux limit). In an Euclidean space V/V_{max} should be uniformly distributed between 0 and 1 with an average value $< V/V_{max} >$ =0.5.

Currently the most accurate determination of the UV Luminosity Function of galaxies from $z \sim$4 to 10 has been done by Bouwens et al. (2021) using data from the deepest surveys obtained with the *Hubble* Space Telescope (*Hubble* Ultra Deep Field, CANDELS, Frontier Fields, etc... - Figure 6.6)

6.1.3 Stellar population

Stars are mainly characterised by their luminosity (or mass) and surface temperature. They are classified according to their spectral type (O, B, A, F, G, K, M) which is an ordering in terms of decreasing surface temperature. A diagram showing the luminosity of a star as a function of the temperature is called a *Hertzsprung-Russel* (HR) diagram (Figure 6.7). Most of the stars occupy the region in the diagram along the line called *the main sequence*. During the stage of their lives in which stars are found on the main sequence line, they are fusing hydrogen in their cores. The next concentration of stars is on the horizontal branch (helium fusion in the core and hydrogen burning in a shell surrounding the core). Another prominent feature is the Hertzsprung gap located in the region between A5 and G0 spectral type and between +1 and -3 absolute magnitudes (i.e. between the top of the main sequence and the giants in the horizontal branch). RR Lyrae

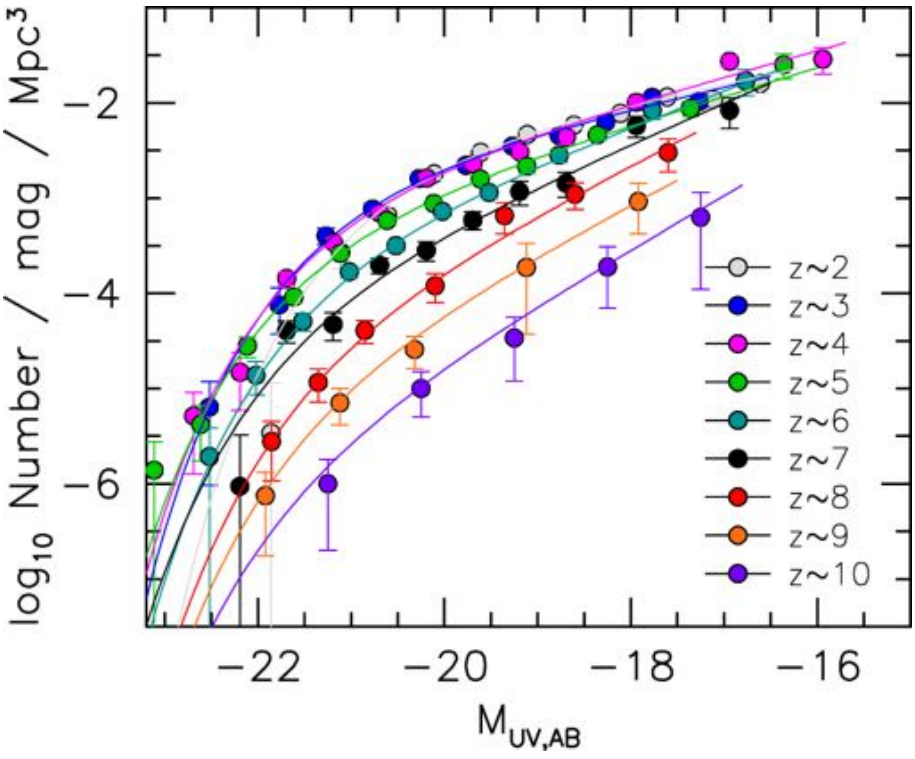


Figure 6.6 : Evolution of the UV Luminosity Function obtained from the analysis of the deepest HST surveys from $z \sim 2$ to $z \sim 10$. A clear decrease in number density if observed from $z \sim 2$ to $z \sim 10$, suggesting a strong decrease in the number of galaxies. We can also see a clear evolution of the characteristic luminosity towards the bright luminosities from $z \sim 10$ to $z \sim 2$ suggesting an increase of the number of bright galaxies with cosmic time. Source : Bouwens et al. (2021)

variable stars can be found in the left of this gap on a section of the diagram called the instability strip. Cepheid variables also fall on the instability strip, at higher luminosities.

The H-R diagram can be used to roughly measure how far away a star cluster or galaxy is from Earth. This can be done by comparing the apparent magnitudes of the stars in the cluster to the absolute magnitudes of stars with known distances (or of model stars). The observed group is then shifted in the vertical direction, until the two main sequences overlap. The difference in magnitude that was bridged in order to match the two groups is called the distance modulus and is a direct measure for the distance (ignoring extinction). This technique is known as main sequence fitting and is a type of spectroscopic parallax.
Stellar luminosity scales approximately as :

$$L \propto M^{\alpha} \tag{6.10}$$

with $\alpha \sim 3$ for stars with mass below $0.5 M_{\odot}$ and $\alpha \sim 4$ for higher masses.
We can also estimate the time a star can remain on the Main Sequence, since the amount of hydrogen is scaled to the stellar mass :

$$\tau_{ms} \propto \frac{M}{L} \propto M^{1-\alpha} \tag{6.11}$$

After this time, stars will evolve off of the main sequence and detailed calculations exist for how this evolution occurs. By applying the previous equation to the most massive stars, it is clear that their lifetime is shorter compared to solar-like stars. Therefore the most massive stars (typically O and B stars) are excellent tracer of recent star-formation. As a consequence of this, UV radiation is generally a good tracer of recent star-formation.

The number of newly formed stars with masses in the range $M \rightarrow M + dM$ is given by the *Initial Mass Function*(IMF), which is commonly taken to have a power-law form (also called Salpeter IMF) :

$$\frac{dN}{dM} \propto M^{\theta} \tag{6.12}$$

where for a standard Salpeter IMF : $\theta = -2.35$
Note that as a consequence of the IMF steeply declining with mass, the bulk of the mass budget is in low mass stars. On the contrary, despite being much less numerous, high mass stars dominate the luminosity, as a consequence of the steep stellar luminosity dependence on the stellar mass.

The Cepheids to probe the distance of nearby galaxies

Cepheids are the first rung in the extragalactic distance ladder and hold a special place in the subject because they were used by Edwin Hubble in his first radial velocity vs. distance plot that led to the discovery of the expanding Universe. Cepheids are a class of variable stars located in the upper H-R diagram (see Figure 6.7); they are evolved, core helium-burning stars whose progenitors are thought to have been B- or late O-type main sequence stars. Their visual magnitudes vary in a regular fashion with amplitudes of between a few tenths of magnitude and ∼2 magnitudes, with periods ranging from a few days to a few weeks. The key role that Cepheids have played in the determination of the extragalactic distance scale stems from the existence of a tight period-luminosity relation: the longer the period of their variability, the brighter their absolute magnitude. Cepheids are supergiant stars with $R \sim 50 R_\odot$ and $L \geq 10^3 L_\odot$, bright enough to be seen over intergalactic distances. Thus by measuring the period of an extragalactic Cepheid star, it is possible to deduce its distance modulus by comparing its observed magnitude with the absolute magnitude, provided the period-luminosity relationship has been calibrated with the known distances of nearby Cepheids in the Milky Way. A later refinement of the calibration includes a colour term, giving the period-luminosity-colour (PLC) relation:

$$M_V = \alpha \log P + \beta (B-V)_0 + \gamma \tag{6.7}$$

where α, β and γ are constants, P the variability period, $(B-V)_0$ the intrinsic $B-V$ colour obtained after correcting for variable extinction Then, including the relation between absolute magnitude and distance :

$$m - M = 5 \log d_{pc} - 5 \tag{6.8}$$

where m is the observed magnitude and d_{pc} is the distance in parsec, therefore

$$d_{pc} = 10^{0.2(\beta[(B-V)_0 + \frac{m}{\beta}] - \alpha \log P + \gamma')} \tag{6.9}$$

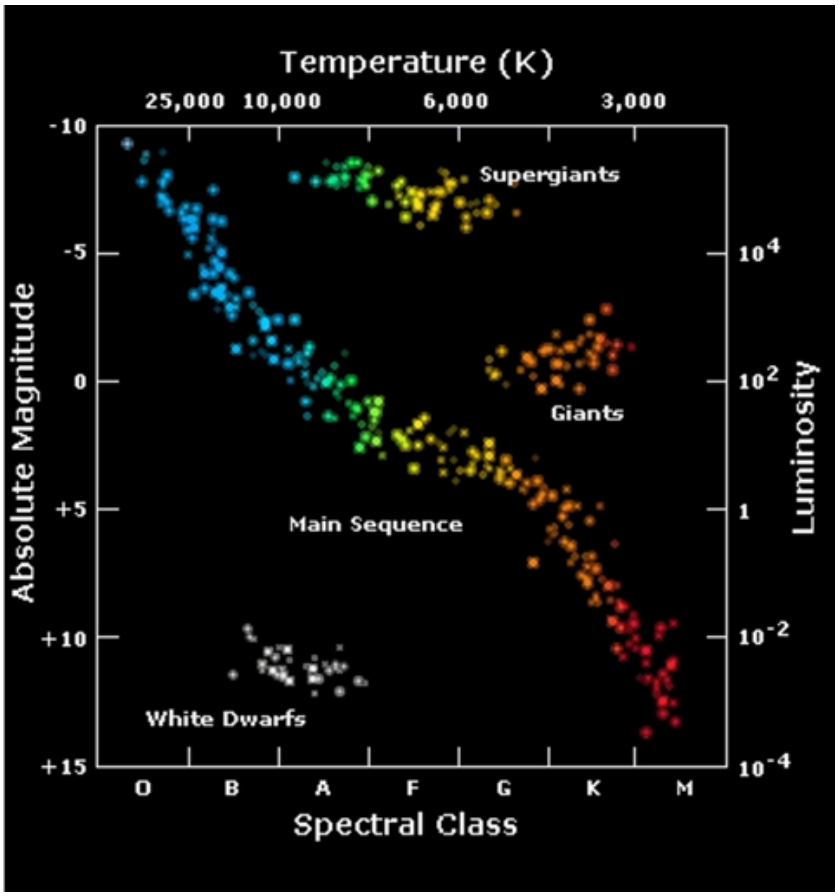


Figure 6.7 : The HR diagram. The position of a star in this diagram gives its nature (giant, dwarfs, etc...). The main sequence is the line on which are most of the stars during their lifetime. Source : NSO

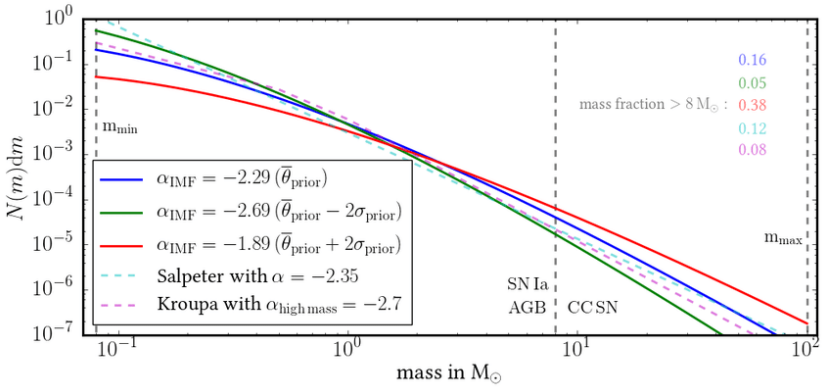


Figure 6.8 : Different Initial Mass function, including the standard Salpeter IMF (in dashed blue line). From Rybizki et al. (2017)

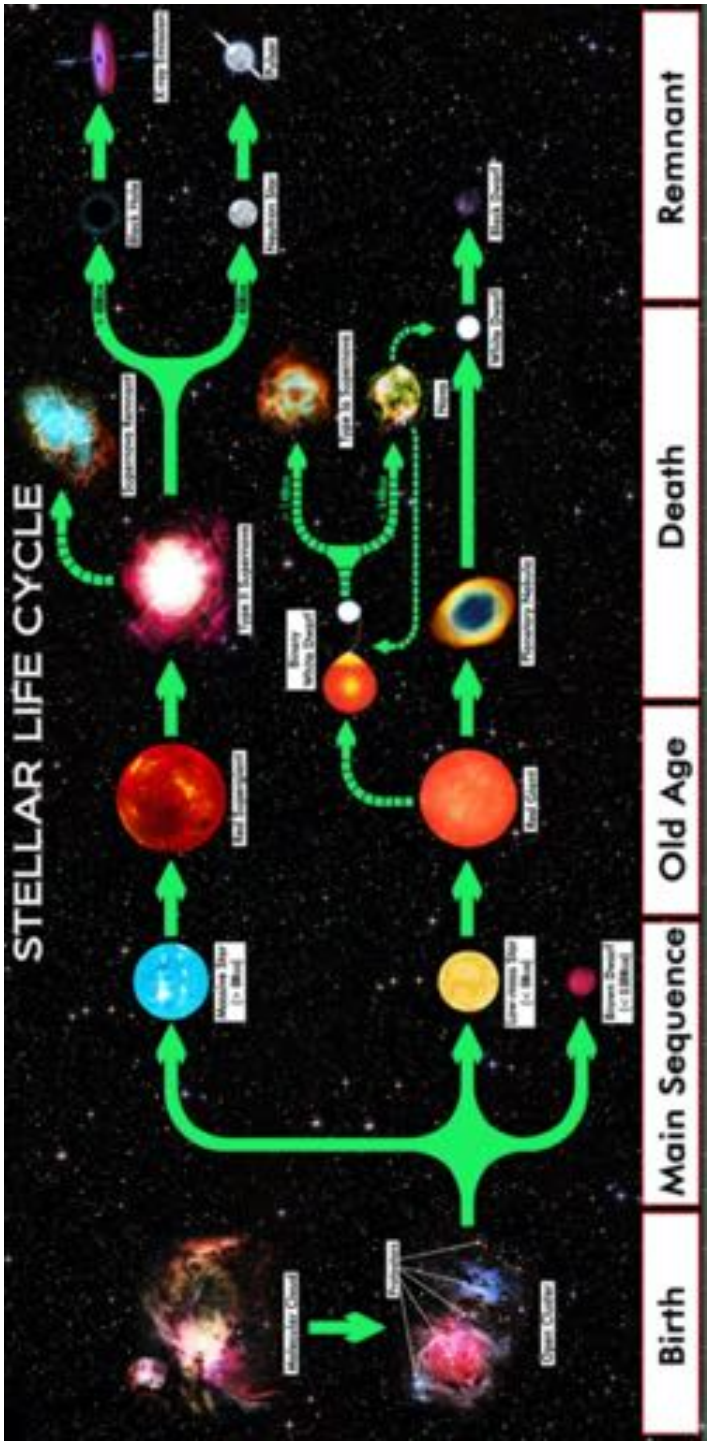


Figure 6.9 : Summary of the life of a star depending on its original mass. Source : R. N. Bailey

6.2 Cloud fragmentation

In the previous chapters, we studied the stability of clouds, but we skipped over an important problem. We saw that the Jeans mass in the ISM was of order the mass of a star. This clearly implies that any more massive cloud also exceeds the Jeans mass and is therefore unstable. Why do we not produce therefore a few very massive collapsed objects (stars ?) from the collapse of a giant molecular cloud ? What we see observationally is that giant molecular clouds are composed of many small dense clouds.

Therefore the scenario that must occurs is the fragmentation of a single massive cloud as it starts to collapse. The current model considers perfectly spherical collapse, anything breaking this symmetry will greatly complicate the physics. To get some insight into how a large cloud might collapse, we return to our analysis of the free-fall timescale.

For a cloud which greatly exceeds its Jeans Mass, it is clear that the gravitational energy is much larger than the thermal energy, and therefore the pressure-free collapse model we used to derive the free-fall time is - at least initially - a reasonable model. Consider again an initially homogeneous spherical cloud of gas of density ρ_0, radius R, and mass M_0 with no internal pressure throughout the collapse. Gas molecules initially at a radius r_0 will have a mass of gas M_r within the radius, and during the collapse this mass remains constant. Therefore the equation of motion for these molecules is :

$$\frac{\partial^2 r}{\partial t^2} = -\frac{GM_r}{r^2} \tag{6.13}$$

Multiplying by $\dot{r} = \frac{dr}{dt}$ and integrating with respect to time gives :

$$\frac{1}{2}\dot{r}^2 = \left[\frac{GM}{r}\right]_{r_0}^{r} = \frac{GM}{r_0}\left(\frac{r_0}{r} - 1\right) = \frac{4\pi}{3} r_0^2 \rho_0 G \left(\frac{r_0}{r} - 1\right) \tag{6.14}$$

hence

$$\frac{dr}{dt} = -\sqrt{\frac{8\pi G \rho r_0^2}{3}} \left(\frac{r_0}{r} - 1\right)^{1/2} \tag{6.15}$$

where the negative sign is chosen to indicate the collapse. This is what we had previously.

Now let's follow this solution by introducing the dimensionless length $\xi = r/r_0$ and the characteristic time $t_0 = \sqrt{3/8\pi G\rho_0}$, and introducing $\tau = t/t_0$. Then :

$$\frac{d\xi}{d\tau} = -\left(\frac{1}{\xi} - 1\right)^{1/2} \tag{6.16}$$

we make the substitution $\xi = cos^2\alpha$, then

$$\frac{d\alpha}{d\tau} = \frac{1}{2cos^2\alpha} \tag{6.17}$$

separating variables and integrating gives :

$$\alpha + \frac{1}{2}sin2\alpha = \tau \tag{6.18}$$

Several conclusions arise from the previous equations :

- The end of collapse occurs when $\xi = 0$, i.e. when $\alpha = \pi/2$ giving :

$$t_{ff} = \frac{\pi}{2}t_0 = \left(\frac{3\pi}{32G\rho_0}\right)^{1/2} \tag{6.19}$$

- The mass within each collapsing shell is conserved, and since the density is initially uniform, it must remain uniform (because t_{ff} does not depend on r_0) and scales such as $r_0^3\rho_0 = r^3\rho(t)$ (constant mass within $r(t)$) or :

$$\frac{\rho(t)}{\rho_0} = \frac{r_0^3}{r^3} = \frac{1}{cos^6\alpha} \tag{6.20}$$

- For a time t close to collapse, we have :

$$\tau = \frac{t}{t_0} = \frac{t_{ff} - (t_{ff} - t)}{t_0} = \frac{\pi}{2} - \epsilon \qquad \text{At this time } r \sim 0 \rightarrow \alpha = \frac{\pi}{2} - \beta \tag{6.21}$$

  Inserting this into eq.6.18 gives, on expanding the "sin" :

$$\beta^3 = \frac{3\epsilon}{2} \tag{6.22}$$

- Then the density close to collapse must be :

$$\frac{\rho(t)}{\rho_0} = \frac{1}{cos^6(\pi/2 - \beta)} = \frac{1}{sin^6\beta} \approx \left(\frac{2}{3\epsilon}\right)^2 = \left(\frac{2t_0}{3(t_{ff} - t)}\right)^2 \tag{6.23}$$

  Note that $\rho(t)/\rho_0$ depends only on t_{ff}, i.e. only on ρ_0, and not on the initial radius r_0.

Now consider that towards the center of the initial sphere the density was perturbed to have a slightly higher density - an overdensity $\rho' = \rho_0 + \delta_0$. This overdensity will have a slightly shorter free-fall time given by :

$$t'_{ff} \approx t_{ff}\left(1 - \frac{\delta_0}{2\rho_0}\right) \tag{6.24}$$

Towards the end of the collapse, this will have grown relative to the mean density of the cloud to :

$$\frac{\rho(t)'}{\rho(t)} \approx \left(\frac{t_{ff} - t}{t'_{ff} - t}\right)^2 \tag{6.25}$$

$$\approx 1 + \frac{\delta_0 t_{ff}}{\rho_0 (t_{ff} - t)} \tag{6.26}$$

Therefore the overdensity grows as :

$$\frac{\delta(t)}{\rho(t)} \approx \frac{\delta_0 t_{ff}}{\rho(t_{ff} - t)} \tag{6.27}$$

This implies that all overdensities (perturbations) grow on the same timescale (i.e. about simultaneous fragmentation). This is different to what we might expect from the Jeans analysis. In the latter, the growth of instabilities of a given mass was exponential and the timescale for growth depended on the wavenumber and hence the mass. Moreover the growth time depended on the wavelength with the largest modes growing fastest.

As a conclusion, we have demonstrated that a small inhomogeneity in the pressure-free case will grow algebraically with time and that all perturbations grow at the same rate. Qualitatively :

- A cloud which is initially very large compared to the Jeans mass will start to undergo approximately pressure-free collapse.
- Many factors will break the symmetry we have considered in our ideal models such as : the initial shape of the cloud, large-scale rotation, small scale velocity variations - turbulence
- Any initial inhomogeneities will grow with time and they all grow on similar timescales
- Eventually we expect the densest of these to become self-gravitating in their own right.

How do we form the Initial Mass Function of stellar masses ?

This is very still uncertain. What we do know observationally is that there is good correspondence between the cloud-mass spectrum and the shape of the IMF. However how does this mass spectrum come about ? Input physics almost certainty includes :

- Turbulence - energy input drives random motions in the gas giving rise to a turbulent cascade. The standard result is that the spectrum of energy in turbulent motion satisfies:

$$E(k)dk \propto k^{-5/3}dk \tag{6.28}$$

- The most successful models invoke scale-free, or fractal, structures within the cloud
- Competitive accretion - for example the denser cores grow by faster than the less dense cores by competing more strongly for the low density gas.

To make further progress requires numerical simulation (Figure 6.10).

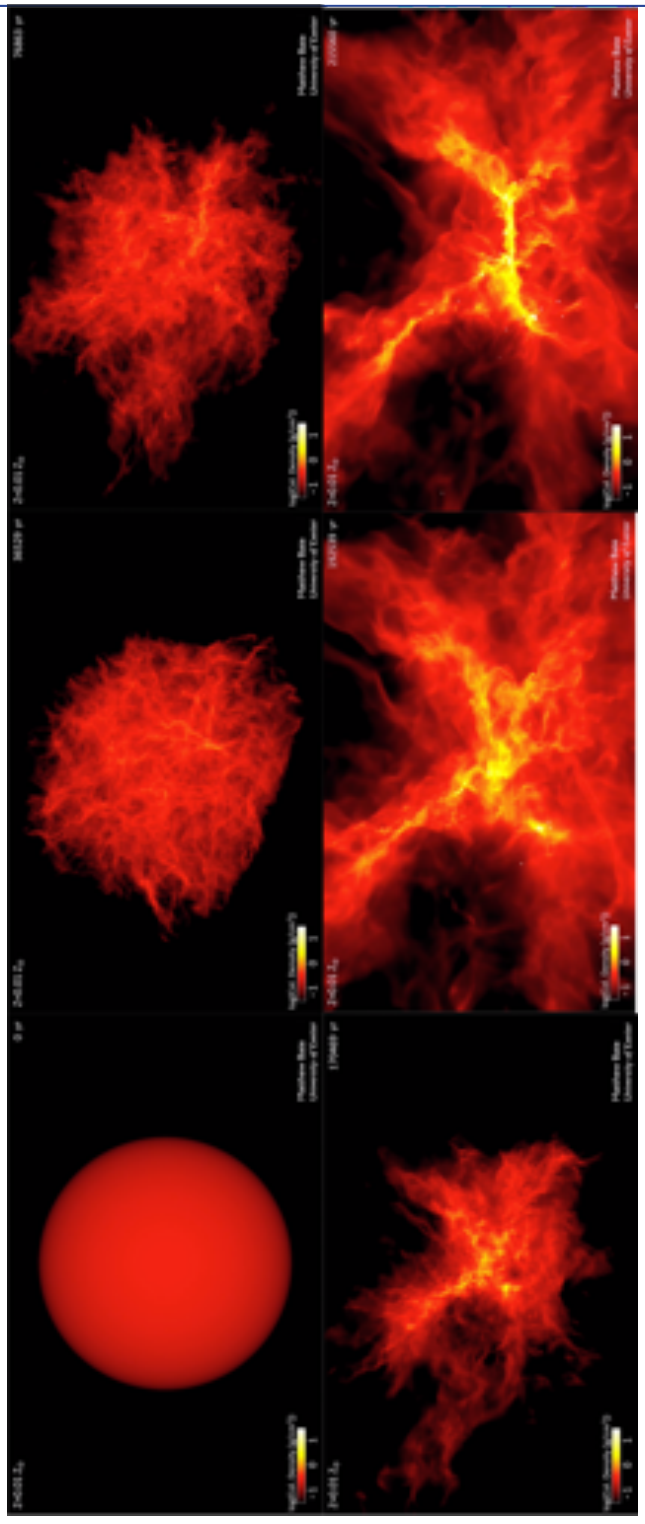


Figure 6.10 : Simulation of the fragmentation of a giant molecular cloud leading to the formation of several stars. Key times are : $t = 0$ initial gas cloud ; t =36kyr clouds of interstellar gas are seen to be very turbulent with supersonic motions ; t =77kyr the turbulent motions in the cloud form shock waves that slowly damp the supersonic motions ; t =170kyr when enough energy has been lost in some regions of the simulation, dense core form ; t =190kyr the formation of stars and brown dwarfs begin in the dense core ; t =210kyr as the stars and brown dwarfs interact with each other, many are ejected from the cloud.

6.3 Galactic-wide star formation

In this section, we consider what factors controls star-formation on a galactic scale. Our main goal is to determine the *star-formation rate* (SFR) for the galaxy (mass of stars formed per unit time). Before going any further, we need to define the following quantities :

- ψ : SFR per unit volume (or possible mass) of the galaxy
- Ψ : SFR for the whole galaxy
- Σ_{SFR} : SFR per unit projected area of a galaxy (a useful observational definition)

We first start by considering observational results. We certainly expect the star formation rate to depend on the amount of available fuel. This is characterised observationally by assuming what is called the *Schmidt law* with :

$$\Sigma_{SFR} \propto \sigma^n \tag{6.29}$$

where σ is the surface gas density. For a constant disc thickness this can be written as :

$$\psi \propto \rho^n \tag{6.30}$$

The best observational results give rise to what we call the *Schmidt-Kennicutt law* (see also Figure 6.11):

$$\frac{\Sigma_{SFR}}{M_\odot yr^{-1}kpc^{-2}} = 2.5 \times 10^{-4} \left(\frac{\sigma}{M_\odot pc^{-2}} \right)^{1.4} \tag{6.31}$$

This result is for the averaged properties of galaxies.

We can now consider some possible simple models for the star formation rate. In all cases, we need to assemble clouds which exceed a Jeans mass and then allow them to collapse.

Collisional assembly

Here we assume that the ISM consists of a large number of small clouds, each less massive than the Jeans mass. Larger clouds are constructed by collisions between the clouds. In this simple case, the collision rate will be of order $n^2 \times v \times A$, where n is the number density of clouds, A their cross section and v the RMS mean (random) velocity in the disc. Also, we assume the collision time is long compared to the free-fall time of the clouds once they exceed their Jeans mass. This suggests that $\psi \propto \rho^2$ which is not in agreement with the Schmidt-Kennicutt law.

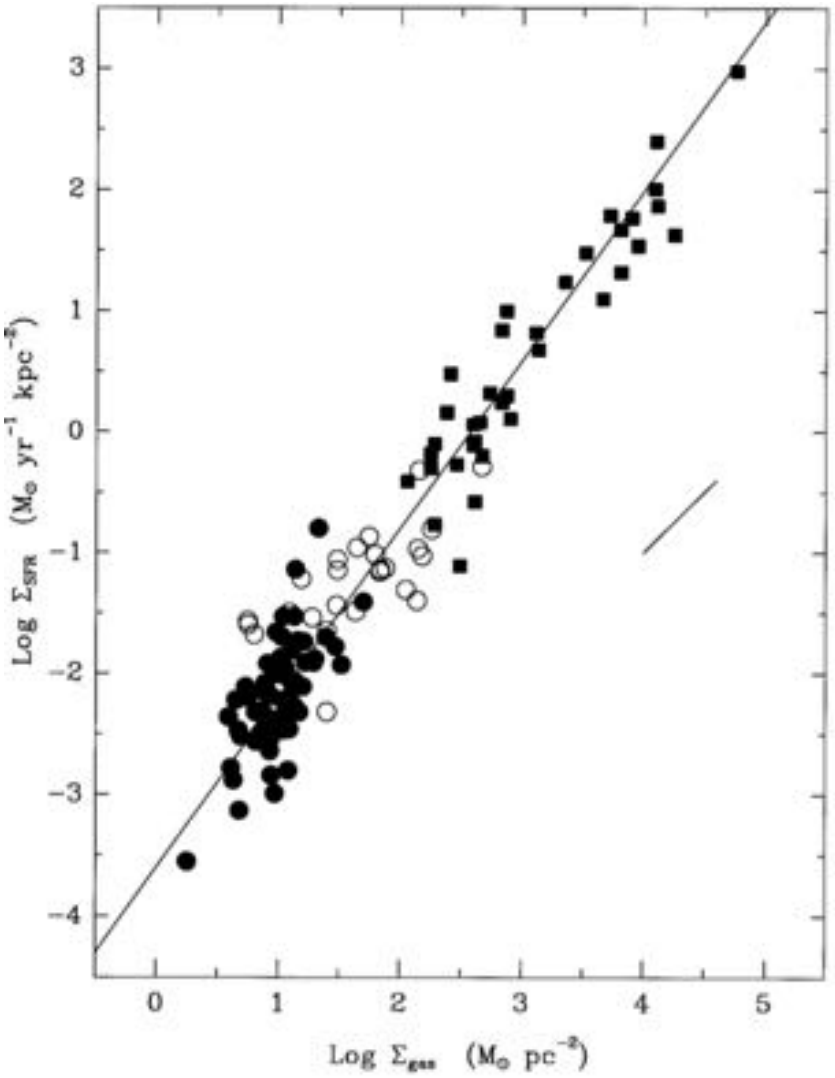


Figure 6.11 : Composite star-formation law for the normal disk (filled circles) and starburst (squares) samples. Open circles show the SFRs and gas densities for the centres of the normal disk galaxies. The line is a least-squares fit with N=1.4. Source : Kennicutt (1998)

Collapse-time limited

An even simpler model is to assume we have large clouds already in place, which exceeds the Jeans mass by a large factor. These will collapse and fragment on a free-fall time as we discussed previously. The SFR will then be determined by the amount of gas which can form stars, which will just be proportional to the density divided by the collapse time scale, such as :

$$\psi \propto \frac{\rho}{t_{ff}} \tag{6.32}$$

but $t_{ff} = (3\pi/32G\rho)^{1/2}$ therefore we get :

$$\psi \propto \rho^{3/2} \tag{6.33}$$

which is close to the Schmidt-Kennicutt law.
However, recent studies have demonstrated that this model is certainly too simplistic and additional effects have to be taken into account.

From the observational point of view, it has been shown that the bulk of the Schmidt-Kennicutt relation is driven by the molecular component of the gas : if the gas surface density is divided into atomic and molecular components, then it becomes observationally clear that the star formation rate (or Σ_{SFR}) does not depend on Σ_{HI} while it depends strongly on Σ_{H2}. This clearly indicates that star-formation is associated with the molecular phase of the gas (where cooling and, therefore, fragmentation, can occur more efficiently). Furthermore, if one considers only the dense component of the molecular gas (i.e. gas with density of about $10^6 \mathrm{cm}^{-3}$ or higher), then the relation with the SFR becomes tighter and linear. Therefore, observations suggest that there is a 1:1 relation between the star formation rate and the mass of the dense component of the molecular gas. The super-linear relation with the total H_2 mass is probably a consequence of the fact that in low density environments, there is an additional diffuse component of the molecular gas that does not participate to the star formation process.

6.4 Simple models of gas and star formation evolution in galaxies

In the following section, we will develop a very simple model of gas in a galaxy to trace the evolution of the star-formation rate. We first need to make the following assumptions :

- The initial total gas mass is M_0

- The mass in the gas is a function of time : $g(t)$
- The mass in stars is given by : $s(t)$
- The star formation rate is given by $\Psi(t)$

Gas is returned from the stars to the ISM via supernovae. We make the approximation that this is an instantaneous process and that the fraction of mass locked up in old stars is α. Gas is therefore returned to the ISM from supernovae at a rate : (1-α)Ψ. This phenomena is called *feedback* and we will discuss it later.

6.4.1 Closed-box model of star formation in a galaxy

For this model, we assume that there is no gas inflow or outflow, i.e. the galaxy evolves as a "closed box". The mass of gas therefore evolves as follows :

$$\frac{dg}{dt} = -\Psi + (1-\alpha)\Psi = -\alpha\Psi \tag{6.34}$$

For simplicity, we can assume a linear relation :

$$\Psi(t) = \epsilon g(t) \tag{6.35}$$

where ϵ can be seen as *the star formation efficiency*, i.e. the mass of stars formed per unit gas mass. We therefore get for the evolution of the mass of gas :

$$\frac{dg}{dt} = -\alpha\epsilon g \tag{6.36}$$

integrating :

$$\ln(g(t)) - \ln(M_0) = -\alpha\epsilon t \tag{6.37}$$

hence :

$$g(t) = M_0 e^{-\alpha\epsilon t} \tag{6.38}$$

Therefore for a galaxy evolving like a closed box, the gas mass decreases exponentially with time.
For the stellar mass :

$$s(t) = M_0 - g(t) = M_0(1 - e^{-\alpha\epsilon t}) \tag{6.39}$$

6.4.2 The effect of inflows and the gas regulator (or "bathtub") model

If we want to develop a more realistic model, we need to assume that galaxies are not evolving as closed boxes : inflows and outflows of gas characterize the life of most galaxies.

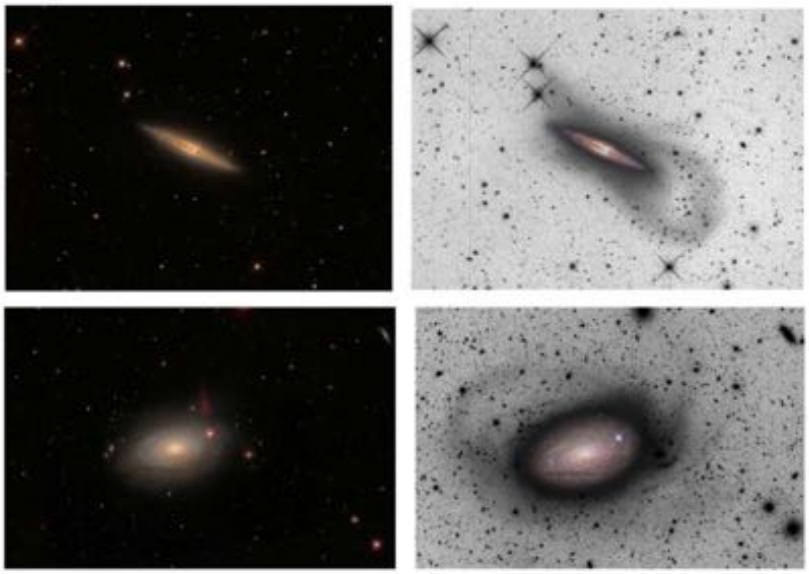


Figure 6.12 : Accreting gas (inflowing gas) in two nearby galaxies. Images on the left from the Sloan Digital Sky Survey of the nearby galaxies NGC 4013 (top) and M63 (bottom) show regular disks while the much deeper images on the right reveal streams of stars accreted from smaller galaxies. From Carlin et al. (2016)

We first focus on gas inflows. The presence of gas inflows was first inferred directly from the analysis of the stellar metallicities in our Galaxy (we will look at this later on). Direct evidence for accreting gas onto the Milky Way was then found through the detection of high velocity gas clouds in the halo of the Milky Way, detected trough HI 21cm observations. Direct evidence for gas accretion has been observed also in other local galaxies (Figure 6.12). According to galaxy evolutionary models, gas inflows are even more prominent and important in high-redshift galaxies.
Suppose now, a galaxy subject to a constant gas inflow rate Φ. Then the evolution of the gas mass becomes :

$$\frac{dg}{dt} = -\alpha\Psi + \Phi \tag{6.40}$$

One may naively expect that a very large inflow rate Φ may produce a galaxy extremely rich in gas, with a total mass completely dominated by the gas mass. However, for the bulk of the galaxies that is not the case. This is because the star formation rate Ψ is linked to the total gas mass through the Schmidt-Kennicutt relation, and it acts as a "valve" that regulates the total amount of gas in the

galaxy, by transforming the excess inflowing gas into stars. More specifically :

$$\frac{dg}{dt} = -\alpha\epsilon g + \Phi \tag{6.41}$$

As the inflow deposit gas onto the galaxy, the gas mass g increases until the point where the right hand term of eq.6.41 becomes 0. At this point the gas content of the galaxy is in equilibrium, i.e. any inflowing gas is transformed into stars. Therefore equilibrium occurs when the gas mass in the galaxy is :

$$g = \frac{\Phi}{\alpha\epsilon} \tag{6.42}$$

This scenario is often dubbed "bathtub" model, where the gas inflow can be seen as water flowing from the tab, the gas mass can be seen as the water in the bathtub, while the star formation rate is the water flowing out of the drain. The rate at which water flows out of the drain is proportional to the water pressure, hence proportional to the amount of water in the bathtub. The amount of water in the bathtub reaches a level where its pressure onto the drain makes the rate of outflowing water equal to the rate of inflowing water. If the rate of inflowing water from the tap is increased or decreased, the level of water in the bathtub increases or decreases to reach a new equilibrium point, where the associated pressure makes the outflow rate again in equilibrium with the inflow rate. Conceptually, the amount of gas in a galaxy works in a similar way, where the water pressure is replaced by the Schmidt-Kennicutt law.

It is often useful to define the *gas fraction*, which is the mass of the gas relative to the total baryonic content (i.e. gas and stars), which is often an indicator of evolutionary stage of a systems :

$$f_{gas} = \frac{M(gas)}{M(baryons)} = \frac{M(gas)}{M(gas) + M(star)} = \frac{g}{g+s} \tag{6.43}$$

We have seen that for a constant inflow rate, at equilibrium, the gas mass is constant, and given by eq.6.42. However, the stellar mass keeps growing and it is given by :

$$s = \Phi t - g \tag{6.44}$$

N.B.: This is true only when the galaxy has reached equilibrium.
Therefore, the gas fraction steadily decreases with time, hence making galaxies "gas poor" as :

$$f_{gas} = \frac{1}{\alpha\epsilon t} \tag{6.45}$$

Once more, this is correct only after the equilibrium has been reached and neglecting the stellar mass produced before reaching equilibrium, i.e. when $t >> 1/\alpha\epsilon$. This implies that the gas fraction is independent of the inflow rate.

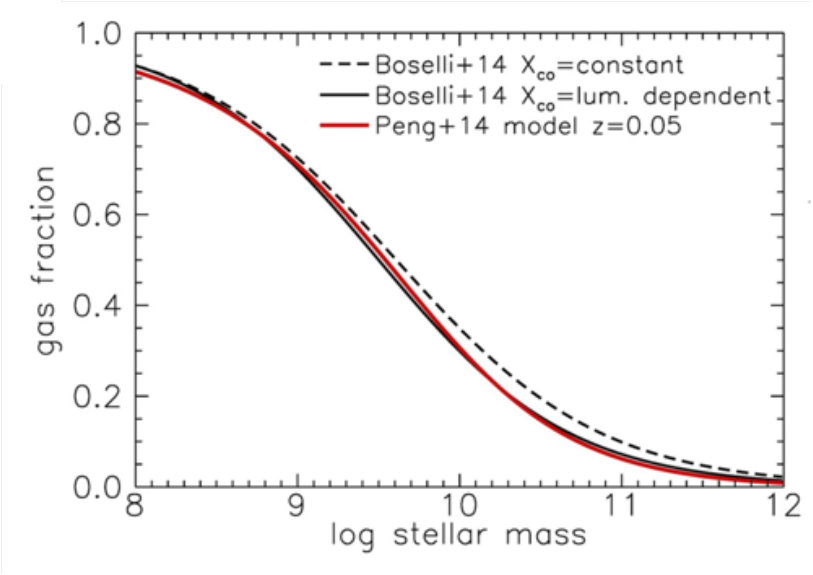


Figure 6.13 : Evolution of the gas fraction as a function of the stellar mass observed in local galaxies. From Peng et al. (2015)

Note that the gas fraction is also indirectly related to the stellar mass. Indeed also the stellar mass depends on time, according to equation 6.44. The gas fraction can also be expressed as :

$$f_{gas}(t) = \frac{\Phi + 1}{\alpha \epsilon s(t)} \tag{6.46}$$

which highlights the (indirect) relation between stellar mass and gas fraction. In particular, it is expected that the gas fraction should decrease with stellar mass, which is indeed observed in local galaxies, as shown in Figure 6.13
Note that these equations are also highlighting that the gas fraction is a good tracer of the galaxy evolutionary stage, i.e. galaxies with low gas fraction are typically more evolved than galaxies with high gas fraction.

6.4.3 The effect of outflows

There is a clear observational evidence that prominent outflows eject large mass of gas out of galaxies. This will be seen in more detail in the next lectures, especially when introducing the concept of "feedback". Here we only mention that one of the primary mechanisms responsible for driving outflows is associated to supernova explosions. Radiation pressure (on dust) from the light emitted by young, luminous stars is another possible mechanism. Since both the supernova rate and radiation

from young stars are proportional to the star formation rate, the outflow rate Λ is often conveniently expressed as proportional to the star formation rate :

$$\Lambda = \lambda\Psi \tag{6.47}$$

where λ is often called the "outflow loading factor" (observational constraints generally give $\lambda \sim 1$ for actively star forming galaxies).
By introducing the effect of outflows, the variation of gas with time becomes :

$$\frac{dg}{dt} = -\alpha\Psi + \Phi - \lambda\Psi \tag{6.48}$$

or, by replacing the linear Schmidt-Kennicutt relation :

$$\frac{dg}{dt} = -\alpha\epsilon g + \Phi - \lambda\epsilon g \tag{6.49}$$

which gives an equilibrium gas mass (right hand term to zero) :

$$g = \frac{\Phi}{(\alpha + \lambda)\epsilon} \tag{6.50}$$

So the effect of outflows is similar to that of varying the value of the constant α. The stellar mass at equilibrium (neglecting the stellar mass formed before reaching equilibrium) is given by $s \approx \alpha\Psi t = \alpha\epsilon g t$. Therefore, at equilibrium, the gas fraction is still given by :

$$f_{gas} \approx \frac{1}{\alpha\epsilon t + 1} \approx \frac{1}{\alpha\epsilon t} \tag{6.51}$$

The gas outflows seems not to have an effect in explaining the lower gas fraction in massive galaxies. However, as we shall see later on, in massive galaxies Active Galactic Nuclei (i.e. supermassive accreting black holes) can greatly contribute to enhance the outflow rate, hence effectively increasing the value of λ, even by a factor of several, in massive galaxies, hence contributing to greatly reduce the gas content in massive galaxies.

6.5 Metallicity evolution of galaxies

With the term "metals" astronomers refer to all elements heavier than helium. The mass fraction of heavy elements, also called the metallicity, is indicated with Z and defined by :

$$Z = \frac{M_{metals}}{M_{tot}} \approx \frac{M_{metals}}{M_H} \tag{6.52}$$

It is generally important to differentiate between stellar metallicity (mass fraction of metals in the stellar atmospheres) and gas metallicity (mass fraction of metals in the interstellar medium). The solar metallicity is $Z_{\odot}$=0.014
Except for some lithium and beryllium, metals are produced by stellar (and explosive) nucleosynthesis and released into the Inter Stellar Medium (ISM) at the end of the stellar lifetime trough supernova explosion and stellar winds. The metals injected into the ISM are then used by the formation of the next generation of stars. Different supernovae (hence stars with different masses) enrich the ISM with different elements. Stars with mass larger than $8M_{\odot}$ leave the Main Sequence in less than about 30 Myr and finish their life as type II (core-collapse) supernovae, which enrich the ISM mostly with α-elements[2] (O, Ne, Mg, Si, S, Ca, ...). Stars with mass less than $8M_{\odot}$ take much longer to leave the Main Sequence (e.g. 1 Gyr or longer, depending on their mass). These stars evolve as Asymptotic Giant Branch stars (AGBs) and then Planetary Nebulae, which enrich the ISM mostly with carbon and nitrogen, and then can yield type Ia Supernovae, which inject into the ISM mostly Fe-peak elements.

Modelling the metallicity evolution of galaxies is a very complex field of astrophysics. Multiple effects have to be taken into account (e.g. production of different elements on different timescales, effects of gas inflows diluting the gas metallicity, effect of gas outflows ejecting metals out of the galaxy, variation of the star formation efficiency). The reward is that the comparison between the metallicity predicted by models and observations can provide tight constraints on the galaxy evolutionary scenarios.

It is beyond the scope of this course to investigate in detail the galaxy metallicity evolutionary models and scenarios. In the following, we will only focus on the metallicity evolution in the case of a closed box system that, although relatively simple, provide some interesting information on galaxy evolution.

Metallicity evolution in a closed box system

We recall the basic definitions and assumptions that we introduced for the closed box model :

- The system is closed with no mass loss or gain :
  - Total mass M_0
  - Mass in gas : $g(t)$

[2]The α process, also known as the α ladder, is one of two classes of nuclear fusion reactions by which stars convert helium into heavier elements, the other being the triple-alpha process. The

  - Mass in stars : $s(t)$
  - Star Formation Rate : $\Psi(t)$

- Gas is returned from stars to the ISM via supernovae at a rate (1-α)Ψ
- We also follow the metallicity of the gas, i.e. the mass fraction of heavy elements in the gas :
  - The production of new metals per mass of stars is p
  - The rate of mass of new metals returned to the ISM via supernovae is therefore $p(1-\alpha)\Psi$
  - The total mass of metals returned to the ISM via supernovae is $(p+Z)(1-\alpha)\Psi$

We recall that the mass of gas in a closed box system evolves as follows :

$$\frac{dg}{dt} = -\Psi + (1-\alpha)\Psi = -\alpha\Psi \tag{6.53}$$

The production of metals is given by :

$$\frac{d(gZ)}{dt} = (p+Z)(1-\alpha)\Psi - Z\Psi = p(1-\alpha)\Psi - \alpha Z\Psi \tag{6.54}$$

Combining these two equations gives :

$$g\frac{dZ}{dt} = p(1-\alpha)\Psi = -p\frac{1-\alpha}{\alpha}\frac{dg}{dt} = -P\frac{dg}{dt} \tag{6.55}$$

where P is called the "yield" in this context. Note that this is a differential equation relating the metallicity and gas mass and in this simple case does not depend on the star formation rate which could have any form ; the previous equation can therefore be rewritten as :

$$dZ = -P\frac{dg}{g} \tag{6.56}$$

We can easily integrate this equation using the boundary conditions $t = 0$, $Z = 0$; $g = M_0$ and for a general time t we get :

$$Z(t) = -P\ln(g/M_0) \tag{6.57}$$

$$g(t) = M_0\exp(-Z/P) \qquad \text{and} \qquad s(t) = M_0 - g(t) = M_0(1-\exp(-Z/P)) \tag{6.58}$$

Models of stellar evolution suggest a value for the yield of about $0.5Z_\odot$ where $Z_\odot$ is the solar metallicity.

Stars formed at a time $t < t_1$ must have a metallicity less than $Z(t_1)$ since they formed out of gas which was less enriched in the past (keep in mind that the metallicity observed on the photospheres of Main Sequence stars reflect the metallicity of the gas out of which these were formed since during the Main Sequence there is no mixing between the stellar interior, where nucleosynthesis occurs, and the photosphere). Therefore, for example, the fraction of stars with metallicities less than 0.1 of the solar value ($Z_\odot$) is given by :

$$\frac{s(< Z_\odot/10)}{M_0} = 1 - \exp(-Z_\odot/10P) \approx 1 - \exp(-1/5) \approx 0.2 \qquad (6.59)$$

The problem is that this is much larger than the number of old low-metallicity stars observed in the disc of our galaxy. This problem is commonly called *G-dwarfs problem*. Our star is itself an old G-star with a relatively high metallicity.

There are of course many ways to improve the above model and solve this problem. The most likely solution (broadly accepted), which also hints at our cosmological models for galaxy evolution, as already discussed, is to allow for inflow and outflow of gas. Gas inflows in particular are thought to be the key ingredient to explain the G-dwarfs problem. Even if the inflowing gas has very low metallicity, the G-dwarf problem can be solved because the additional gas can prolong the duration of star formation (relative to the closed box system), hence there are more stars formed at late times, when the gas has already been enriched with metals.

6.6 Stellar orbits and spiral structure

In the previous sections, we have seen that star-formation should occur in regions of overdensities, i.e where it is more likely that gas clouds are compressed, perturbed and collapse to form stars. Observationally we see that this happens in the spiral arms of disk galaxies or at the edges of stellar bars. In the following we investigate the dynamics behind the formation of spiral arms and bars, and the physical conditions that lead to disk instabilities. We will first start by recalling the basic properties of the rotation curves in spiral galaxies.

6.6.1 Rotation curves in galaxy disks

By using spectral lines, such as the neutral hydrogen (HI) at 21cm, or optical nebular emission lines (e.g. Hα), we can measure the rotation curve (the rotation velocity as a function of radius) with accuracy. Disk rotation curves are characterized by a central region (a few kpc in radius at most) where the rotation velocity

triple-alpha process consumes only helium, and produces carbon.

Figure 6.14 : Color image of a spiral galaxy (M101). The star formation is traced by the blue light, and occurs mainly in the arms. Source : STFC

scales linearly with radius ($v(r) \propto r$, i.e. the angular velocity $\Omega_r \sim$ constant). We can therefore estimate the mass within a radius r by applying Gauss's theorem, assuming a spherical mass distribution and circular orbits :

$$\frac{v^2}{r} = \frac{GM(r)}{r^2} \tag{6.60}$$

For a velocity independent of radius, this implies $M(r) \propto r$ which is clearly not consistent with the density distribution of stars, which is inferred from the observed stellar surface brightness, neither with the mass in gas. The conclusion is that there must be more mass - dark matter - than suggested by the distribution of visible matter (stars and baryonic gas). We note that $M(r) \propto 1/r^2$, which is what we expect in a simple model for a dark matter halo.

6.6.2 Stellar Orbits

To simplify the study, we consider a cylindrically symmetric model in which the potential is given by $\Phi(r, z)$ and examine orbits initially in the z =0 plane. We assume the following :

- The angular momentum per unit mass for each star is conserved :

$$l = r^2\phi = \text{constant} \tag{6.61}$$

  where ϕ is the azimuthal angle

- The energy per unit mass is also conserved, therefore :

$$E = \frac{1}{2}\dot{r}^2 + \frac{1}{2}(r\dot{\phi})^2 + \Phi(r) = \frac{1}{2}\dot{r}^2 + \frac{l}{2r^2} + \Phi(r) \tag{6.62}$$

- The equation of motion in the radial direction is just :

$$\ddot{r} - r\dot{\phi}^2 = -\frac{\partial\Phi}{\partial r} \tag{6.63}$$

- It is useful to introduce the effective potential :

$$\Phi_e = \Phi + \frac{l^2}{2r^2} \tag{6.64}$$

  to rewrite the equation of motion :

$$\ddot{r} = -\frac{\partial\Phi_e}{\partial r} \tag{6.65}$$

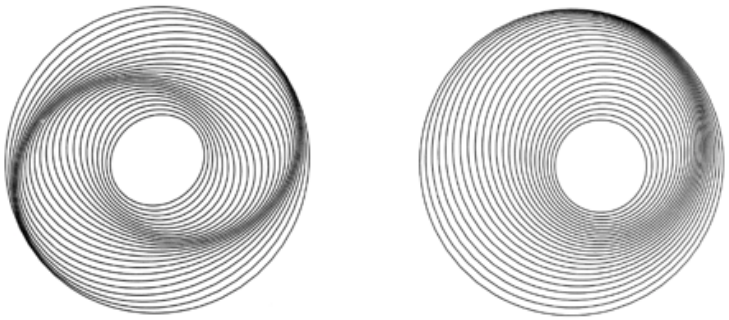


Figure 6.15 : Resonant spiral perturbation

By definition, circular orbits have $\ddot{r} = 0$ and hence are given by :

$$\frac{\partial \Phi_e}{\partial_r}\Big)_{r_0}) = 0 = \frac{\partial \Phi}{\partial r}\Big)_{r_0} - \frac{l^2}{r_0^3} \tag{6.66}$$

where r_0 is the radius of the circular orbit.

Given a $\Phi(r)$, or equivalently a mass distribution, we can calculate the properties of the circular stellar orbit at any radius. In particular at a radius r the angular velocity of the circular orbit is given by :

$$\Omega(r)^2 = \frac{l^2}{r^4} = \frac{1}{r}\frac{\partial \Phi}{\partial r} \tag{6.67}$$

A star when perturbed will undergo small motions about this circular orbit. We write $x = r - r_0$ and expand the effective potential about r_0 :

$$\Phi_e(x) = \Phi_e(r_0) + x\frac{\partial \Phi_e}{\partial r}\Big)_{r_0} + \frac{1}{2}x^2\frac{\partial^2 \Phi_e}{\partial r^2}\Big)_{r_0} + O(x^3) \tag{6.68}$$

The term in x clearly is zero and the radial equation is just :

$$\ddot{x} = -\frac{\partial \Phi_e}{\partial x} = -x\frac{\partial^2 \Phi_e}{\partial r^2}\Big)_{r_0} \tag{6.69}$$

which is the Simple Harmonic Motion (SHM) with a frequency called the *epicyclic*

frequency given by :

$$\kappa^2 = \frac{\partial^2 \Phi_e}{\partial r^2}\Big)_{r_0} = \left[\frac{\partial^2 \Phi}{\partial r^2} + \frac{3l^2}{r^4}\right]_{r_0} \tag{6.70}$$

$$= \left[r\frac{\partial \Omega^2}{dr} + 4\Omega^2\right]_{r_0} \tag{6.71}$$

$$= 4\Omega^2\left[1 + \frac{r}{2\Omega}\frac{\partial \Omega}{\partial r}\right]_{r_0} \tag{6.72}$$

Similarly, we can show that for small amplitude motion out of the plane of the disc (z direction) the star undergoes SHM with :

$$v^2 = \left.\frac{\partial^2 \Phi_e}{\partial z^2}\right|_{r=r_0, z=0} \tag{6.73}$$

6.6.3 Resonant orbits

In general spiral structure is complicated, but one important physical idea is the concept of resonant orbits. Near circular orbits will be the superposition of the pure circular orbit plus the radial motion. In the lab frame after one radial oscillation of period $T_r = 2\pi/\kappa$ the orbit will have precessed by :

$$\Delta\phi \approx \Omega T_r \tag{6.74}$$

In general the orbit will not close in this frame, but consider the situation in a frame rotating at Ω_p - the *pattern speed*. In this frame :

$$\Delta\phi_p = \Delta\phi - \Omega_p T_r \tag{6.75}$$

For the orbit to close after m radial oscillations, we require :

$$2n\pi = \Omega m T_r - \Omega_p m T_r \tag{6.76}$$

or

$$\Omega_p = \Omega - \frac{n}{m}\kappa \tag{6.77}$$

Interestingly, in many systems the form of $\Omega(r)$ means that, for $n = 1$ and $m = 2$, Ω_p is approximately constant across the disk of the galaxy.

In this case, we can arrange the phase of the orbits so that adjacent stars in certain regions of the disk have a higher density and put them on the $n = 1$, $m = 2$, perturbed orbits - these will then be long lived. We can achieved this, for example, if there is an external perturbation, which causes a rotating potential which is resonant with these orbits. There is a lot more to spiral structure than this simple analysis, but the idea of resonant orbit plays a central role.

6.6.4 Stability of a rotating disc - spiral density waves

We now consider the stability of a rotating disc in more detail. One manifestation of unstable discs are spiral density waves, but the following analysis will also provide an insight into the overall stability of the disc and star formation.
We start again with Euler's equations for an ideal fluid. We use a cylindrical coordinate system. The radial equation is :

$$\rho\frac{\partial v_r}{\partial t} + \underbrace{\rho v_r\frac{\partial v_r}{\partial r} + \rho\frac{v_\theta}{r}\frac{\partial v_r}{\partial \theta} - \frac{\rho}{r}v_\theta^2}_{[\rho(v\cdot\nabla)v]_r} = \rho\frac{\partial \phi_g}{\partial r} - \frac{\partial p}{\partial r} \tag{6.78}$$

where ϕ_g is the gravitational potential. The θ equation is :

$$\rho\frac{\partial v_\theta}{\partial t} + \underbrace{\rho v_r\frac{\partial v_\theta}{\partial r} + \rho\frac{v_\theta}{r}\frac{\partial v_\theta}{\partial \theta} - \frac{\rho}{r}v_\theta v_r}_{[\rho(v\cdot\nabla)v]_\theta} = -\rho\frac{1}{r}\frac{\partial \phi_g}{\partial \theta} - \frac{1}{r}\frac{\partial p}{\partial \theta} \tag{6.79}$$

and the equation of continuity :

$$\frac{\partial \rho}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}(\rho r v_r) + \frac{1}{r}\frac{\partial}{\partial \theta}(\rho v_\theta) = 0 \tag{6.80}$$

We are not interested here in the vertical structure in the disc and therefore we project these equations into a 2D form by integrating the pressure in the z-direction and assuming all of the mass is concentrated in a plane.
If we assume there is no dependence of the velocities on z then this integration gives identical equations except that the density and pressure are replaced by:

$$\rho(r,\theta,z,t) = \sigma(r,\theta,z)\delta(z) \qquad ; \qquad P = \int \rho \,\mathrm{d}z \tag{6.81}$$

where σ is the disk surface density. To proceed, we will use a perturbation analysis. The unperturbed solution is an axially symmetric rotating mass distribution, $\sigma_0(r)$, with v_0=$(0, r\Omega(r))$ and an unperturbed potential ϕ_{g_0}. To these we add on small perturbations, which are all functions of r, θ, and t :

$$v = (u, v + r\Omega) \tag{6.82}$$
$$\sigma = \sigma_0(r) + \sigma'(r,\theta,t) \tag{6.83}$$
$$\phi_g = \phi_{g_0} + \phi'_g(r,\theta,z,t) \tag{6.84}$$

where $\sigma_0(r)$ and $\sigma'(r,\theta,t)$ satisfy :

$$\nabla^2\phi_{g_0} = 4\pi G\sigma_0\delta(z) \qquad \nabla^2\phi'_g = 4\pi G\sigma'(r,\theta,t)\delta(z) \tag{6.85}$$

P is a force per unit length and is of course the plane (2D) analogue of pressure. We assume an isothermal-like equation of state and write :

$$P = a_0^2 \sigma \tag{6.86}$$

where a_0 is essentially the velocity dispersion of the particles – stars – in the disk. Keeping only terms to first order in small quantities, we obtain after some algebra:

$$\text{Radial } u \qquad \frac{\partial u}{\partial t} + \Omega \frac{\partial u}{\partial \theta} - 2v\Omega = -\frac{a_0^2}{\sigma_0}\frac{\partial \sigma'}{\partial r} - \frac{\partial \phi'_g}{\partial r} \tag{6.87}$$

$$\text{Angular } v \qquad \frac{\partial v}{\partial t} + \Omega \frac{\partial v}{\partial \theta} + \frac{\kappa^2 u}{2\Omega} = -\frac{a_0^2}{r\sigma_0}\frac{\partial \sigma'}{\partial \theta} - \frac{1}{r}\frac{\partial \phi'_g}{\partial \theta} \tag{6.88}$$

$$\text{Continuity} \qquad \frac{\partial \sigma'}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}(r\sigma_0 u) + \frac{\sigma_0}{r}\frac{\partial v}{\partial \theta} + \Omega \frac{\partial \sigma'}{\partial \theta} = 0 \tag{6.89}$$

κ is the epicyclic frequency we introduced earlier and comes into the equations since:

$$u\frac{\partial r\Omega}{\partial r} + u\Omega = 2\Omega u\left(1 + \frac{r}{2\Omega}\frac{\partial \Omega}{\partial r}\right) = \frac{u\kappa^2}{2\Omega} \tag{6.90}$$

We now look for spiral-like solutions writing :

$$\sigma' = \hat{\sigma} \exp\left[i(\omega t - n\theta + \Psi(r))\right] \tag{6.91}$$

with similar expressions for u, v and ϕ'_g. We make the further approximations that $\hat{\sigma}$ are approximately constant functions. The basic properties of solutions of this form are :

- The maximum in the density occurs for $\omega t - n\theta + \Psi(r) = 0$. To understand the implications of this form consider $t = 0$, then the locus of the maximum density has : $n\theta{=}\Psi(r)$ which is a spiral patter which represents n-armed spiral ; and the pattern makes an angle to the θ direction, the pitch angle, $\tan\alpha = \frac{n}{r\frac{d\Psi}{dr}} = \frac{n}{kr}$ where we define k as $k = \frac{d\Psi}{dr}$

- The radial distance between maxima is given by :

$$\Psi(r + \lambda) - \Psi(r) = 2\pi n \tag{6.92}$$

$$\text{For } \lambda << r : \Psi(r + \lambda) \approx \Psi(r) + k\lambda \text{ and } \lambda \approx 2\pi n/|k| \tag{6.93}$$

- The spiral pattern represent advances in time as the *pattern speed* :

$$\Omega_p = \dot{\theta} = \omega/n \tag{6.94}$$

Substituting these spiral-like solutions into eq.6.80 and eq.6.81 leads to a straightforward algebraic equations. Solving Poisson's equation is more difficult. The solution can be found in what is called the tight-winding approximation when $n/kr << 1$ - this implies, among other things, that the main variation in the gravitational potential is radial and not in the angular direction.
We will quote the results of this analysis :

$$\kappa^2 - n^2(\Omega_p - \Omega)^2 + k^2 a_0^2 = 2\pi G|k|\sigma_0 \tag{6.95}$$

or re-arranging :

$$n^2(\Omega_p - \Omega)^2 = (\omega - n\Omega)^2 = \kappa^2 + k^2 a_0^2 - 2\pi G|k|\sigma_0 \tag{6.96}$$

A special case is when the disc is not rotating. We can find the dispersion relation for this special case by setting $n\Omega = 0$ and $\kappa = 0$ to give:

$$\omega^2 = k^2 a_0^2 - 2\pi G|k|\sigma_0 \tag{6.97}$$

This is the dispersion relation we discussed before identifying a_0 as the isothermal sound speed a_T. Returning to the rotating disc, the limiting situation occurs when n is zero (i.e. axisymmetric disc), then:

$$\omega^2 = \kappa^2 + k^2 a_0^2 - 2\pi G|k|\sigma_0 \tag{6.98}$$

the disc is stable to the spiral perturbations provided that $\omega^2 > 0$. From the previous equation it is clear that the rotation (κ) and velocity dispersion (a_0) have both a stabilising effect, while gravity (i.e. the surface density σ_0) has a destabilising effect:

$$\frac{\omega^2}{\kappa^2} = 1 + \frac{k^2 a_0^2}{\kappa^2} - \frac{2\pi G|k|\sigma_0}{\kappa^2} \tag{6.99}$$

or

$$\frac{\omega^2}{\kappa^2} = 1 + \frac{Q^2}{4}\frac{k^2}{k_T^2} - \frac{|k|}{k_T} \tag{6.100}$$

where Q is the disc stability parameter, defined as :

$$Q = \frac{2k_T a_0}{\kappa} = \frac{\kappa a_0}{\pi G \sigma_0} \tag{6.101}$$

k_T is the *Toomre wave number* defined as :

$$k_T = \frac{\kappa^2}{2\pi G\sigma_0} \tag{6.102}$$

From this expression, the division between stable and unstable solutions occurs when $\omega^2 = 0$ or :

$$\frac{|k|}{k_T} = \frac{2}{Q^2}\left(1 \pm (1 + Q^2)^{1/2}\right) \tag{6.103}$$

This only has a solution for $|k|$ when $Q < 1$. In this case there are regions where $\omega < 0$, hence spiral perturbations grow exponentially, yielding the collapse of clouds, likely resulting into star formation. When $Q > 1 \rightarrow \omega^2 > 0 \forall |k|$ and the disc is always stable. The latter condition can be expressed in terms of a minimum velocity dispersion that makes the disk stable:

$$a_0 \geq a_{0,min} = \frac{\pi G \sigma_0}{\kappa} \tag{6.104}$$

6.6.5 Lindblad Resonances

Returning to the full dispersion relation (eq.6.96) we can write :

$$k^2 a_0^2 - \frac{|k|}{k_T}\kappa^2 + (\kappa^2 - n^2(\Omega_p - \Omega)^2) = 0 \tag{6.105}$$

This is just a quadratic in $|k|$ and, assuming the case $a_0 = a_{0,min}$ it is easy to find that to have real wave-like solutions (i.e. $|k|$ real and positive) it is necessary that :

$$1 \pm \frac{n}{\kappa}(\Omega_p - \Omega) \geq 0 \tag{6.106}$$

and hence :

$$\Omega - \frac{\kappa}{n} \leq \Omega_p \leq \Omega + \frac{\kappa}{n} \tag{6.107}$$

We only have spiral density wave solutions between the inner and outer Lindblad resonances :

- $\Omega_p = \Omega - \frac{\kappa}{n}$ inner Lindblad resonance
- $\Omega_p = \Omega + \frac{\kappa}{n}$ outer Lindblad resonance
- $\Omega_p = \Omega$ is called corotation

6.6.6 The threshold of the Schmidt-Kennicutt relation

The Schmidt-Kennicutt law (relation between the star formation rate surface density and the gas surface density) is know to have a threshold for star formation : below a minimum gas surface density, star formation becomes very inefficient ($\Sigma_{thr}(gas) \sim 5 M_\odot \mathrm{kpc}^{-2}$, but variable from galaxy to galaxy). This is illustrated in Figure 6.16. This threshold is generally responsible for the fact that, despite the gas distribution (HI) extending to large radii, star formation is confined within a certain radius.
The Toomre stability parameter gives a simple explanation for this effect. Indeed, in order to have a instability, i.e. star formation, it has to be $Q < 1$. Therefore,

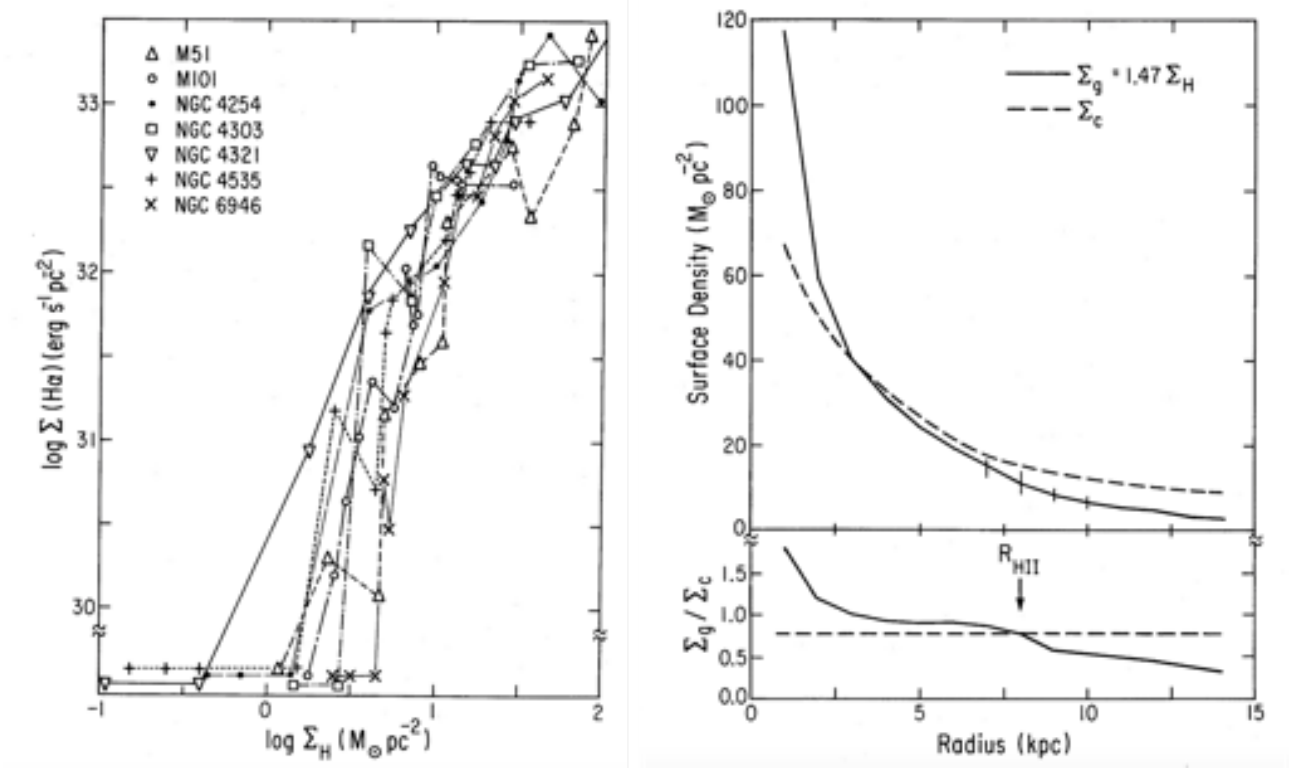


Figure 6.16 : (Left) Dependence of Hα surface luminosity on the total gas surface density - there is clear evidence for a threshold in the Hα emission which is a tracer of the star formation. (Right) Distribution of gas and critical surface density as a function of the radius. The bottom panel shows the ratio of these two quantities and the arrow indicates the edge of the region in which HII regions are found; There are no HII regions, and hence no indication for star-formation, beyond the radius where the gas density falls below a value very close to the critical value. From Kennicutt et al. (1989)

to have star formation the gas surface density has to exceed a critical value :

$$\sigma > \sigma_c = \frac{\kappa a_0}{\pi G} \tag{6.108}$$

It is observed that this condition is indeed met in the central part of disk galaxies, where star formation occurs.

7. Feedback processes in star formation

We are bits of stellar matter that got cold by accident, bits of a star gone wrong.

— Sir Arthur Eddington

In the local Universe, only 4% of baryons have been converted into stars, while theory of gas cooling and gravitational collapse would have expected that, over the entire life of the Universe (∼13.7 Gyr) about 80% of the baryons should have been converted into stars. Hence some mechanisms must have been responsible for suppressing the formation of stars and/or removing gas from galaxies, hence preventing it from star forming stars. Such mechanisms are globally identified as *negative feedback* processes. We will see that these processes are partially associated with star formation itself and partially associated with accretion onto supermassive black hole.

7.1 Stellar Winds

Mass loss from massive and evolved stars occurs in the form of stellar wind (the solar wind is a nearby example).
The properties of the stellar wind depend on the star :

- Post-main sequence stars nearing the ends of their lives often eject large quantities of mass in massive ($\dot{M} > 10^{-3}$M$_\odot$yr-1) but relatively slow ($v <$10 km s^{-1}) winds. Post-main sequence stars include red giants and super giants, and asymptotic giant branch stars. The winds are likely to be driven by radiation pressure on dust condensing in the upper atmosphere of the stars.
- Massive O and B stars have stellar winds with lower mass loss rates ($\dot{M} < 10^{-6} M_\odot yr^{-1}$) but very high velocity ($v > 1000 - 2000 kms^{-1}$). Such winds are driven by radiation pressure on the resonance absorption lines in heavy

elements such as carbon and nitrogen. These high-energy stellar winds provide significant feedback and drive shocks into the ISM - they have a two-shock structure. The freely-expanding stellar wind hits an inner termination shock where it's kinetic energy is thermalized, producing 10^6 K, X-Ray emitting plasma. The hot, high-pressure, shocked wind expands, driving a shock into the surrounding interstellar gas. If the surrounding stellar gas is dense enough (number densities n >0.1 cm^{-3} typically), the swept up gas radiatively cools far faster than the hot interior, forming a thin, relatively dense shell around the hot, shocked wind.

Although stellar winds from main sequence stars do not strongly influence the evolution of the stars, during the later, post-main sequence phase, mass lost by stellar winds can decide the fate of the star. Many intermediate mass stars become white dwarfs at the ends of their lives rather than exploding as supernovae only because they lost enough mass in their winds. It can be shown that about 20% of the energy of the stellar wind is converted into kinetic energy of the surrounding ISM.

7.2 Additional feedback effect from O-B massive stars

Massive stars on the main sequence have additional feedback effects, as a consequence of their high luminosity, which dominate the total luminosity of the newly formed stars (as a consequence of $L \propto M^4$) on the Main Sequence, and of their high temperature, producing large amount of ionising photons. More specifically, additional feedback effects of O-B stars :

- Photoionisation, and hence heating, of the ISM.
- Strong radiation pressure onto the dust in the ISM clouds can drive powerful winds, which eject gas out of the galaxy.

7.3 Supernovae

Stellar winds are therefore efficient at putting energy into the ISM, however they represent a relatively minor feedback process in galaxies which are actively producing very massive stars, i.e. those with masses greater than about $8M_\odot$ as these end their lives as supernovae which have a very dramatic effect on the galaxy. The total energy input to the ISM from each supernova explosion is of order 10^{43}-10^{44}J. Moreover the supernovae feedback also processed stellar material and therefore are responsible for the metallicity enrichment of the ISM as we have already discussed. Typically one solar mass of material is ejected back into the ISM.

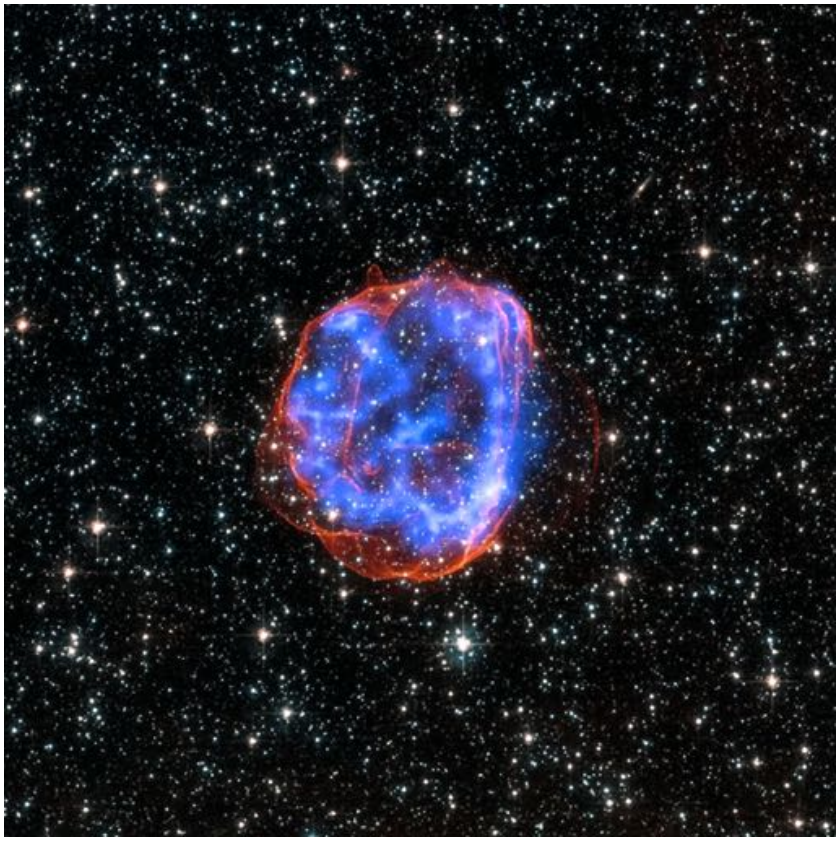


Figure 7.1 : X-ray & Optical Images of SNR E0519-69.0 a supernova remnant in the LMC. Here, multimillion degree gas is seen in X-rays from Chandra (blue). The outer edge of the explosion (red) and stars in the field of view are seen in visible light from Hubble. Source : NASA/CXC/Rutgers/J.Hughes (X-Ray) ; NASA/STScI (Optical)

Figure 7.2 : Images of the starburst galaxy M82. Optical light from stars (yellow-green/Hubble Space Telescope) shows the disk of a modest-sized, apparently normal galaxy. The Spitzer Space Telescope infrared image (red) shows that cool gas and dust are also being ejected. Chandra's X-ray image (blue) reveals gas that has been heated to millions of degrees by the violent outflow. The eruption can be traced back to the central regions of the galaxy where stars are forming at a furious rate, some 10 times faster than in the Milky Way Galaxy. Source : NASA

7.4 Starburst and ultraluminous galaxies

Starburst and ultraluminous galaxies are the most extreme examples of star forming systems. The definition of both of these classes is somewhat vague, but a working definition is that for a starburst the measured rate of star formation is so large that the available gaseous fuel would be used in a period of time very much less than a Hubble time, say about 100 Myr. Ultraluminous galaxies have far-infrared luminosities greater than 10^{12} $L_\odot$ and infrared star formation rates of order 50-1000 $M_\odot yr^{-1}$.

The archetypal starburst galaxy is M82. The star-formation is concentrated in a nuclear region - a nuclear starburst. The associated supernovae drive a galactic-scale outflow of hot gas which is clearly visible in the X-ray image shown on Figure 7.2.

7.5 Global stellar feedback

Stellar winds, photoionisation and supernova explosion have a cumulative effect of suppressing star formation, i.e. have a negative feedback effect onto star formation. More specifically :

- heating by shocks and photoionisation makes the collapse and fragmentation of molecular clouds more difficult (because efficient cooling requires radiation from molecules and dust which are often destroyed in these environments)
- Galactic winds (mostly produced by the cumulative effect of SN explosions and radiation from young O-B stars) expel gas out of the galaxy, hence remove fuel available for star-formation.

The latter mechanism is most important in low mass galaxies for which the gravitational potential well is not deep enough to retain the gas.

7.6 Feedback from Supermassive Accreting Black Holes

Supermassive Black Holes are found in the nuclei of most galaxies, and their masses are seen to be proportional to the mass of the stellar spheroid (bulge or whole galaxy in the case of an elliptical galaxy) in a ratio of $M_{BH} \sim 10^{-3} M_{sph}$.
The accretion of black holes is a much more efficient process of converting mass into radiation, than the nuclear fusion in the stellar interiors. Indeed the accreting matter, as a consequence of angular momentum conservation, forms an accretion disc. Within the accretion disc, viscosity converts gravitational energy into thermal energy, making the disc very hot (temperatures in excess of 10^5K), hence the thermal energy is radiated away (Black Body radiation, mostly in the UV).
The luminosity from black hole accretion can be very high, resulting in very luminous nuclei, called Active Galactic Nuclei (AGN), which can photoionise gas clouds over large galactic regions. QSOs (or Quasars) are the most extreme version of AGNs, in which luminosity associated with the accreting supermassive black hole is so high ($> 10^{12} L_\odot$) to outshine the rest of the galaxies (hence their images appear star-like, hence their name "Quasi Stellar-Objects", or QSOs).

7.6.1 The Eddington Limit

Accretion can occur only if the luminosity of the accretion disc is lower than the so-called Eddington limit. The latter is given by the condition that the (inwards) gravitational force must be larger than the (outwards) radiation pressure. Radiation pressure is dominated by Thomson scattering of photons on electrons, while gravitational force is dominated by protons (photons and electrons are coupled

through Columbian drag forces), therefore the condition for accretion is :

$$\frac{1}{4\pi r^2}\frac{L}{c}\sigma_e < \frac{Gm_p M_{BH}}{r^2} \tag{7.1}$$

therefore :

$$L < L_{edd} = \frac{4\pi c G m_p}{\sigma_e} M_{BH} \tag{7.2}$$

replacing the numerical values :

$$L_{edd} \approx 3 \times 10^4 \left(\frac{M_{BH}}{M_\odot}\right) L_\odot \tag{7.3}$$

This can be inverted to find the minimum black hole mass that can be associated with an accretion luminosity :

$$M_{BH} > 3 \times 10^{-5} \left(\frac{L}{L_\odot}\right) M_\odot \tag{7.4}$$

For QSOs :

$$L_{QSO} \sim 10^{12} - 10^{14} L_\odot \tag{7.5}$$

$$M_{BH} > 3 \times 10^7 - 3 \times 10^9 M_\odot \tag{7.6}$$

7.6.2 Mass to energy efficiency

In accretion discs the loss of gravitational energy allows matter in the disc to move towards inner orbits. Consider an element of mass dm, in the accretion disc, moving from an orbit with radius $r + dr$ to an orbit with radius r. From the virial theorem half of the variation of gravitational potential energy must be radiated away:

$$\mathrm{d}E_{rad} = -\frac{1}{2}\frac{GMdm}{r + \mathrm{d}r} - \left(-\frac{1}{2}\frac{GM\mathrm{d}m}{r}\right) \tag{7.7}$$

where M is the mass of the central object (in this case, the supermassive black hole). Hence the luminosity is :

$$\mathrm{d}L = \frac{\mathrm{d}E_{rad}}{\mathrm{d}t} = \frac{1}{2}GM\frac{\mathrm{d}M}{\mathrm{d}t}\left(\frac{1}{r} - \frac{1}{r+dr}\right) = \frac{1}{2}GM\dot{M}\frac{\mathrm{d}r}{r^2} \tag{7.8}$$

by integrating :

$$L = \int_{R_{out}}^{R_{in}} \mathrm{d}L = \frac{1}{2}GM\dot{M}\left(\frac{1}{R_{in}} - \frac{1}{R_{out}}\right) \approx \frac{1}{2}\frac{GM\dot{M}}{R_{in}} \tag{7.9}$$

The efficiency of conversion of matter into energy is given by :

$$L = \epsilon \dot{M} c^2 \tag{7.10}$$

therefore

$$\epsilon = \frac{GM}{2c^2 R_{in}} \tag{7.11}$$

From general relativity, it can be shown that, for a non-rotating black hole, the innermost stable orbit is :

$$R_{in} = 3R_{Sch} = 6\frac{GM}{c^2} \tag{7.12}$$

where $R_{Sch} = 2GM/c^2$ is the so-called Schwartzschild radius. Therefore :

$$\epsilon = 1/12 \approx 0.1 \tag{7.13}$$

which is much higher than efficiency achieved by the nuclear fusion of hydrogen in the interior of stars (ϵ_{nuc} ~0.007).

7.6.3 AGN Feedback

As a consequence of the high mass-to-energy conversion efficiency, a black hole of mass M_{BH}, in the process of accreting matter, must have radiated an amount of energy given by :

$$E_{BH} = 0.1 M_{BH} c^2 \tag{7.14}$$

The gravitational binding energy of a bulge, or of a spheroidal galaxy, of mass M_{gal} is given by :

$$E_{gal} \sim M_{gal}\sigma^2 \tag{7.15}$$

where σ is the stellar velocity dispersion. Since M_{BH} ~$10^{-3} M_{sph}$, it follows that :

$$\frac{E_{BH}}{E_{gal}} \approx 10^{-4} \left(\frac{c}{\sigma}\right)^2 \tag{7.16}$$

Most galaxies have $\sigma < 400$km s^{-1}, hence :

$$\frac{E_{BH}}{E_{gal}} > 80 \tag{7.17}$$

therefore the energy produced by the growth of the black hole exceeds the binding energy of the galaxy by a large factor. So in principle black hole accretion can seriously "harm" its host galaxy. If even a small fraction of the produced energy is transferred to the gas in the galaxy, then an accreting supermassive black hole (i.e. an AGN) can have profund effects on the evolution of its host galaxy.

Understanding AGN feedback onto the host galaxy is currently subject of extensive observational and theoretical studies. Here we only summarise that black hole accretion can provide powerful feedback onto the host galaxy and surrounding medium in two forms :

- when it is radiating at "Quasar"-like luminosities ($L > 10^{12} L_{\odot}$, this generally happens when a BH with mass $>10^{7} M_{\odot}$ accretes close to the Eddington rate) then radiation pressure can produce a powerful nuclear wind, which shocks and heats the ISM in the host galaxy, and generates a powerful outflow that expels gas out of the galaxy. Direct radiation pressure from the AGN onto the dusty clouds also contribute to the outflow. AGN-driven massive outflows have recently been observed in quasar host galaxies. Both observations and theoretical calculations show that powerful quasars can potentially completely clean a galaxy out of their ISM, hence completely quenching star formation. This is often referred to as "quasar-mode" feedback.

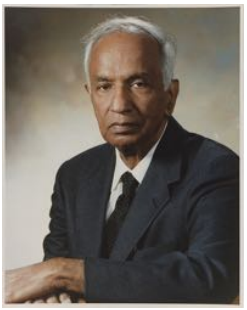
- when the accreting black holes produce powerful radio jets (this happens generally when the BH is accreting at significantly sub-Eddington rates) then the energy injected into the intergalactic and intracluster medium keeps such medium hot, preventing it to cool and feed the central galaxy with new gas. This feedback effect onto the intracluster medium is observed as cavities generated into the X-ray emitting medium, associated with the radio-lobes created by the relativistic radio jet. This is often referred to as "radio-mode" feedback.

The two processes are thought to act in a sequence. Once the elliptical massive galaxy has been cleaned of its gas by the quasar mode, then the radio-mode prevents the massive galaxy from accreting new gas and therefore prevents additional star formation. The galaxy then evolves passively into an old elliptical.

7.6.4 The relative role of different feedback processes

We will see more in detail in the next chapter that the stellar mass function of galaxies differs significantly from the mass function of Dark Haloes in which galaxies are hosted, the former being much flatter at lower masses and much steeper at high masses. If baryons in all halos were forming stars with the same efficiency, then the two mass functions should be identical in shape (scaled by the M/L ratio). The currently favoured scenario is that negative feedback processes are responsible for such deviations. More specifically, stellar feedback is responsible for reducing the efficiency of star formation predominantly in low mass galaxies, while AGN feedback is responsible for quenching star formation predominantly in high mass galaxies.

8. Galaxies interaction and triggering star formation



The black holes of nature are the most perfect macroscopic objects there are in the universe: the only elements in their construction are our concepts of space and time.

— Subrahmanyan Chandrasekhar

In this chapter, we will discuss how interaction between galaxies can trigger the star formation.

8.1 Collisions between stellar system

8.1.1 Stellar collisional cross section and relaxation time

We start by estimating the collision cross section for strong gravitation interactions making the following assumptions :

- We identify a pair-wise collision as one in which the star is significantly deflected
- this will occur if :

$$\frac{1}{2}mv^2 \sim \frac{Gm^2}{r} \tag{8.1}$$

  which gives the scattering radius :

$$r_s \sim \frac{2Gm}{v^2} \tag{8.2}$$

- The effective cross section is therefore defined as :

$$\pi r_s^2 \sim \pi \left(\frac{2Gm}{v^2}\right)^2 \tag{8.3}$$

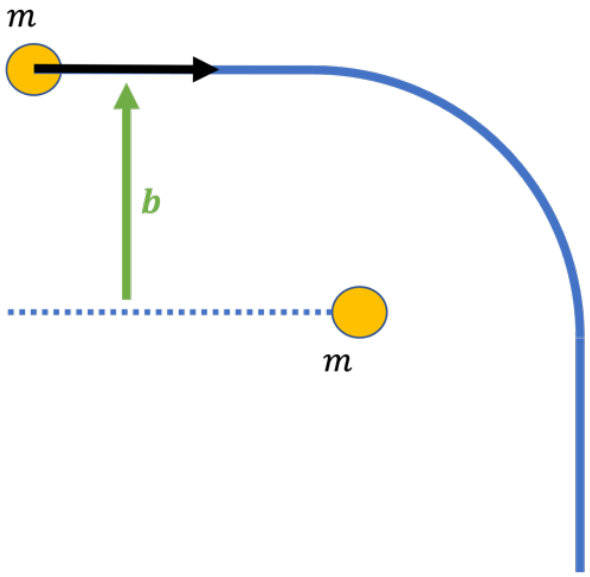


Figure 8.1 : Stellar scattering via gravitational interaction

The mean free path for strong collision is $\lambda = \frac{1}{n\sigma}$ and the collision time is $t_s = \frac{\lambda}{v}$, or :

$$t_s = \frac{v^3}{4\pi G^2 m^2 n} = 4 \times 10^{12}\,\mathrm{yr} \left(\frac{v}{10\mathrm{km/s}}\right)^3 \left(\frac{m}{M_\odot}\right)^{-2} \left(\frac{n}{1\mathrm{pc}^{-3}}\right)^{-1} \tag{8.4}$$

We can then conclude that even strong gravitational collisions do not occur.

However, we still need to consider interactions which give rise to a small interaction. For example, consider a star which is almost undeflected : the net force over the interaction is perpendicular to the direction of travel and for an impact parameter[1] b, we have :

$$F_\perp = \frac{GmMb}{(b^2 + v^2t^2)^{3/2}} = M\frac{dv_\perp}{dt} \tag{8.5}$$

integrating over the interaction gives :

$$\Delta v_\perp = \frac{1}{M}\int_{-\infty}^{+\infty} F_\perp dt = \frac{2Gm}{bv} \tag{8.6}$$

[1]The impact parameter is the perpendicular distance to the closest approach if the projectile were undeflected

The momentum is conserved by the fact that the second star (here assumed stationary) must have a $\Delta v_\perp$ in the opposite sense. As the star undergoes many small deflections, we may assume they are in random directions giving $< \Delta v_\perp >= 0$.

We now integrate over all encounters.
The number of collisions with $b \to b + \mathrm{d}b$ is $vt \times 2\pi b \mathrm{d}b \times n$:

$$< \Delta v_\perp^2 >= \int_{b_{min}}^{b_{max}} \left(\frac{2Gm}{bv}\right)^2 nvt 2\pi b \,\mathrm{d}b = \frac{8\pi G^2 m^2 nt}{v} \ln\left(\frac{b_{max}}{b_{min}}\right) \tag{8.7}$$

and we write :

$$\ln \Lambda = \ln\left(\frac{b_{max}}{b_{min}}\right) \tag{8.8}$$

The relaxation time is the time taken for $< \Delta v_\perp^2 >= v^2$, hence :

$$\boxed{t_r = \frac{v^3}{8\pi G^2 m^2 n \ln \Lambda} = \frac{t_s}{2 \ln \Lambda}} \tag{8.9}$$

8.1.2 Collisions in stellar systems

We can quantify the degree to which stars interact using the collisional relaxation time (eq.8.9). We need to estimate $\ln \Lambda = \ln(b_{max}/b_{min})$. For an isolated system of N stars and size R we can estimate :

- $b_{max} \sim R$
- $b_{min} \sim r_s$ for consistency

We can also apply the virial theorem to this system :

$$2 \times \frac{1}{2} N m v^2 = \frac{G(Nm)^2}{R} \tag{8.10}$$

and obtain $v^2 R = GNm$.
Then :

$$\Lambda = \frac{b_{max}}{b_{min}} = R \times \frac{v^2}{2Gm} = \frac{1}{2} N \tag{8.11}$$

The typical time for a star to cross the system is defined as the crossing time :

$$t_c \approx \frac{R}{v} \tag{8.12}$$

Since $n = \frac{N}{\frac{4}{3}\pi R^3}$ we find :

$$\frac{t_r}{t_c} = \frac{v}{R} \frac{v^3 \frac{4}{3}\pi R^3}{8\pi G^2 m^2 N \ln \Lambda} = \frac{N}{6 \ln(N/2)} \tag{8.13}$$

For a galaxy, with $N \sim 10^{11}$ stars : $t_c \sim 10^7 - 10^8$ yr and $t_r \sim 10^8 t_c$ much longer than the Hubble time. Similarly, for a globular cluster with $N \sim 10^6$, we find that $t_r \sim 10^{10}$ yr. Therefore most of the stellar systems are collisionless.

8.1.3 Dynamical friction

By definition, the dynamical friction (also called the Chandrasekhar friction), is loss of momentum and kinetic energy of moving bodies through gravitational interactions with surrounding matter in space. It was first discussed in detail by Subrahmanyan Chandrasekhar in 1943.
Recall the result we had for the change in velocity perpendicular to nearly undeflected path of a particle of mass M as it passes a mass of m :

$$\Delta v_\perp = \frac{2Gm}{bv} \tag{8.14}$$

By conservation of momentum, the particles of mass m should also suffer a change in perpendicular velocity of :

$$\Delta V_\perp = \frac{2GM}{bv} \tag{8.15}$$

The total change in kinetic energy of the system due to these changes in perpendicular velocity is :

$$\Delta E_{k,\perp} = \frac{M}{2}\left(\frac{2Gm}{bv}\right)^2 + \frac{m}{2}\left(\frac{2GM}{bv}\right)^2 = \frac{2G^2 mM(M+m)}{b^2 v^2} \tag{8.16}$$

This energy can only come from the forward motion of M, i.e. $\Delta E_{k,\perp} + \frac{M}{2}\Delta v_{||}^2 = 0$, but :

$$\frac{1}{2}\Delta(v_{||}^2) = v_{||}\Delta v_{||} \approx v\Delta v_{||} \tag{8.17}$$

hence :

$$-\Delta v_{||} \approx \frac{\Delta E_{k,\perp}}{M v_{||}} = \frac{2G^2 m(M+m)}{b^2 v^3} \tag{8.18}$$

Finally we again integrate over all impact parameters to give :

$$-\frac{dv}{dt} = \int_{b_{min}}^{b_{max}} nv\frac{2G^2 m(M+m)}{b^2 v^3} 2\pi b\, db = \frac{4\pi G^2 (M+m)\rho}{v^2}\ln\Lambda \tag{8.19}$$

We now apply this model to the interaction of two galaxies. We identify M as the mass of an entire galaxy and $\rho = nm$ as the density within the interacting galaxy. We again define a relaxation time :

$$t_r = \frac{v}{|\dot{v}|} = \frac{v^3}{4\pi G^2 M\rho \ln\Lambda} \tag{8.20}$$

This will be the timescale on which dynamical friction acts to dissipate the bulk kinetic energy of the interacting systems and allow the galaxies to merge.

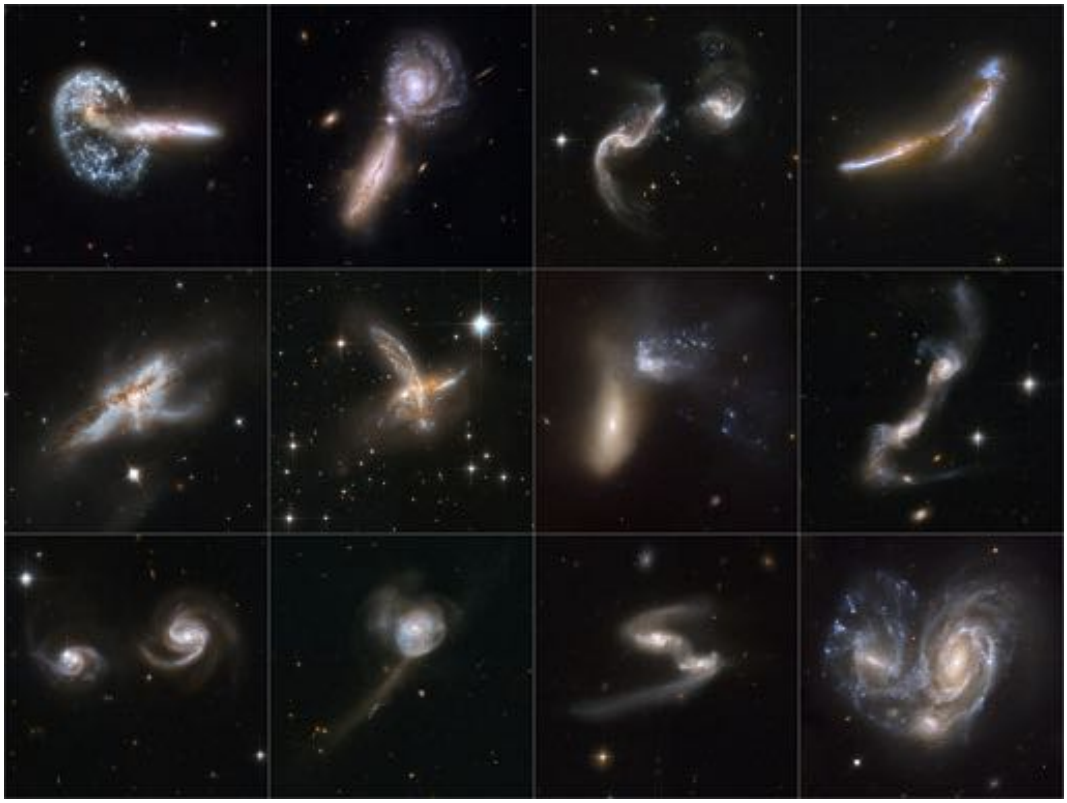


Figure 8.2 : Example of interacting galaxies as seen by the *Hubble* Space Telescope. Source : NASA, ESA, the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration, and A. Evans (University of Virginia, Charlottesville/NRAO/Stony Brook University)

8.2 Interacting galaxies

In the first part of this course, we have always considered galaxies as independent isolated systems. But as we all know, galaxies show a remarkable range of different structures. We begin this section by reviewing some of these.
Many galaxies, including our own, have lower-mass companions with who they are interacting gravitationally. In many cases, the effects of these interactions are relatively small. One famous example is M51 for which its companion is possibly the cause of the extremely well defined spiral arms (Figure 8.3). Simulations show that even modest tidal perturbations can trigger the formation of tidal densities waves as we have discussed earlier.

Figure 8.4 is an image from the *Hubble* Space Telescope showing two interacting spiral galaxies (LEDA 62867 and NGC 6786) and again with well-defined spiral

Figure 8.3 : M51, also known as the Whirlpool Galaxy, is interacting with a small companion which is probably responsible for the well defined arms. Source : NASA, ESA, S. Beckwith (STScI) and the Hubble Heritage Team (STScI/AURA)

Figure 8.4 : HST color image of LEDA 62867 and NGC 6786. Simulations show that although at a very early stage, this system will lead to a merger between the two galaxies in few billion years. Source : NASA / ESA / HST

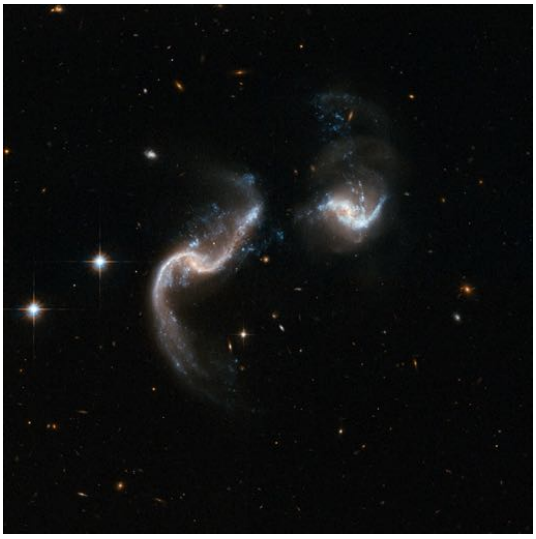


Figure 8.5 : HST color image of Arp 256.Source : NASA / ESA / HST

arms. A supernovae has been observed in the large spiral in 2004. Although at a very early stage, it is possible that this system will lead to a merger between the two galaxies. Another interesting system to look at is Arp 256 (Figure 8.5), a stunning system with two spiral galaxies in an early stage of merging. The *Hubble* image shows two galaxies with strongly disrupted morphologies and an astonishing number of blue knots of star formation which have been triggered by the interaction. Note also the characteristic tidal tails associated with the galaxy to the left. Tidal tails are a direct indication of a strong gravitational interaction.

In advanced stages (Figure 8.6) intense regions of star formation appear as long threadlike structures located between the main galaxy cores. The system almost qualifies as an ultra-luminous system, but has not yet reached the late stage of coalescence that is the norm for the most ultra-luminous system. IC 883 (Figure 8.7) has a very disturbed, complex, central region with two tidal tails of approximately the same length emerging at nearly right angles : one diagonally to the top right of the frame and the other to the bottom right. The twin tidal tails suggest that IC 883 is the remnant of the merger of two gas-rich galaxies. The collision appears to have triggered a burst of star formation, indicated by a number of bright star clusters in the central region.
Figure 8.8 shows the early merger Arp 270 (where the galaxy disks are still very distinct). Diffuse X-ray emitting gas with $T < 10^6$K is seen associated with two galaxy disks and several point sources of varying hardness (temperatures) are seen

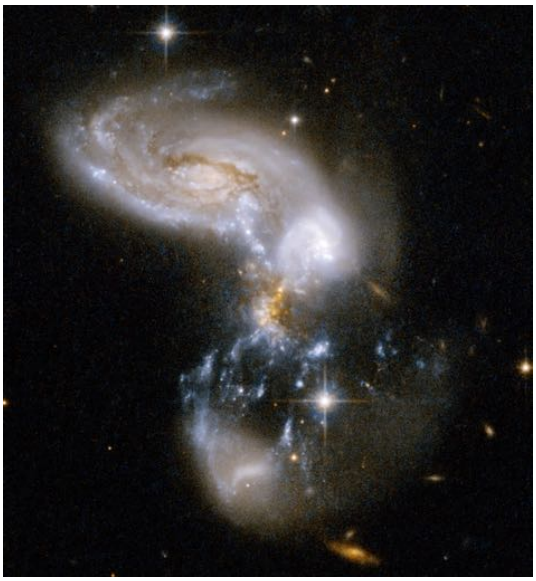


Figure 8.6 : HST color image of II ZW 96.Source : NASA / ESA / HST

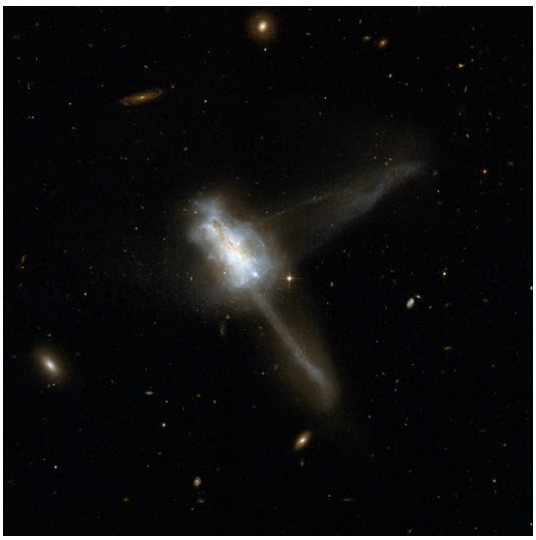


Figure 8.7 : HST color image of ICC 883 (also known as Arp 193) .Source : NASA / ESA / HST

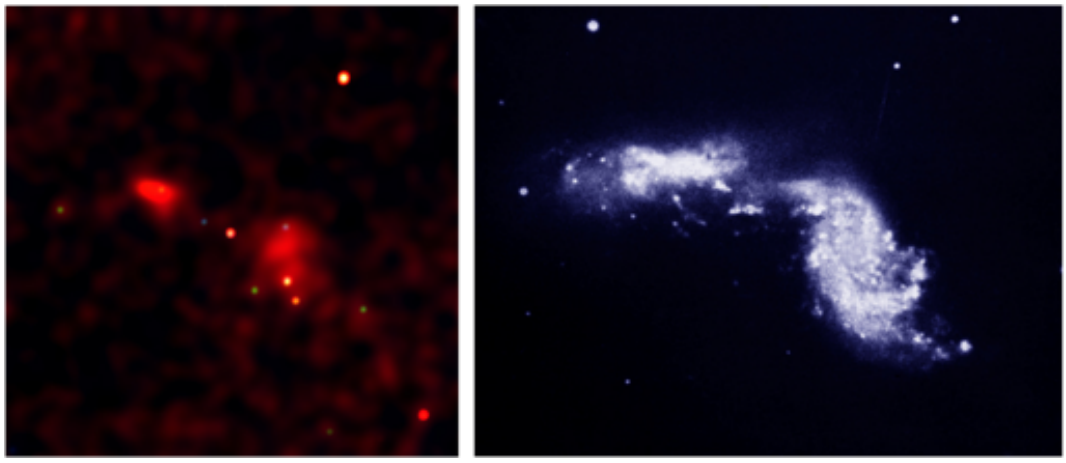


Figure 8.8 : Two views of the galaxies Arp 270 : (left) X-Ray image from the *Chandra* Space Observatory colour coded such that red representents soft X-ray emission (0.2-0.9 keV), green intermediate (0.9-2.5 keV) and blue hard X-Ray (2.5-10 keV). (Right) optical image. Source : NASA/U. Birmingham/A.Read

scattered within the galaxies. Of particular note are the hard X-ray sources seen where the disks collide, probably resulting from strong star-formation.

Despite the very extensive range of observed structures, the physics of these processes seem to be relatively simple, dominated by pure gravitation interactions. We can successfully model galaxies in this context as being composed just of collisionless massive particles - either stars and/or dark matter. An example of a simulation of a pair of galaxies which undergo merger is shown on Figure 8.9. We can apply our analysis of dynamical friction developed earlier. The expression we had was :

$$t_r = \frac{v}{|\dot{v}|} = \frac{v^3}{4\pi G^2 M \rho \ln \Lambda} \tag{8.21}$$

To get a first estimate, we approximate $\ln \lambda \sim 1$ and take similar values to those used in the simulation :

- $v \sim 200$ km/s
- $M \sim 10^{10} \mathrm{M}_\odot$
- central density : $\rho \sim 10^8 \mathrm{M}_\odot$ kpc^{-3}

Our dynamical friction relaxation time then gives as timescales 8×10^8yrs. In the simulation, the time taken from the point where the discs first overlap to completion of merger is about 1 billion years, in remarkably good agreement with our

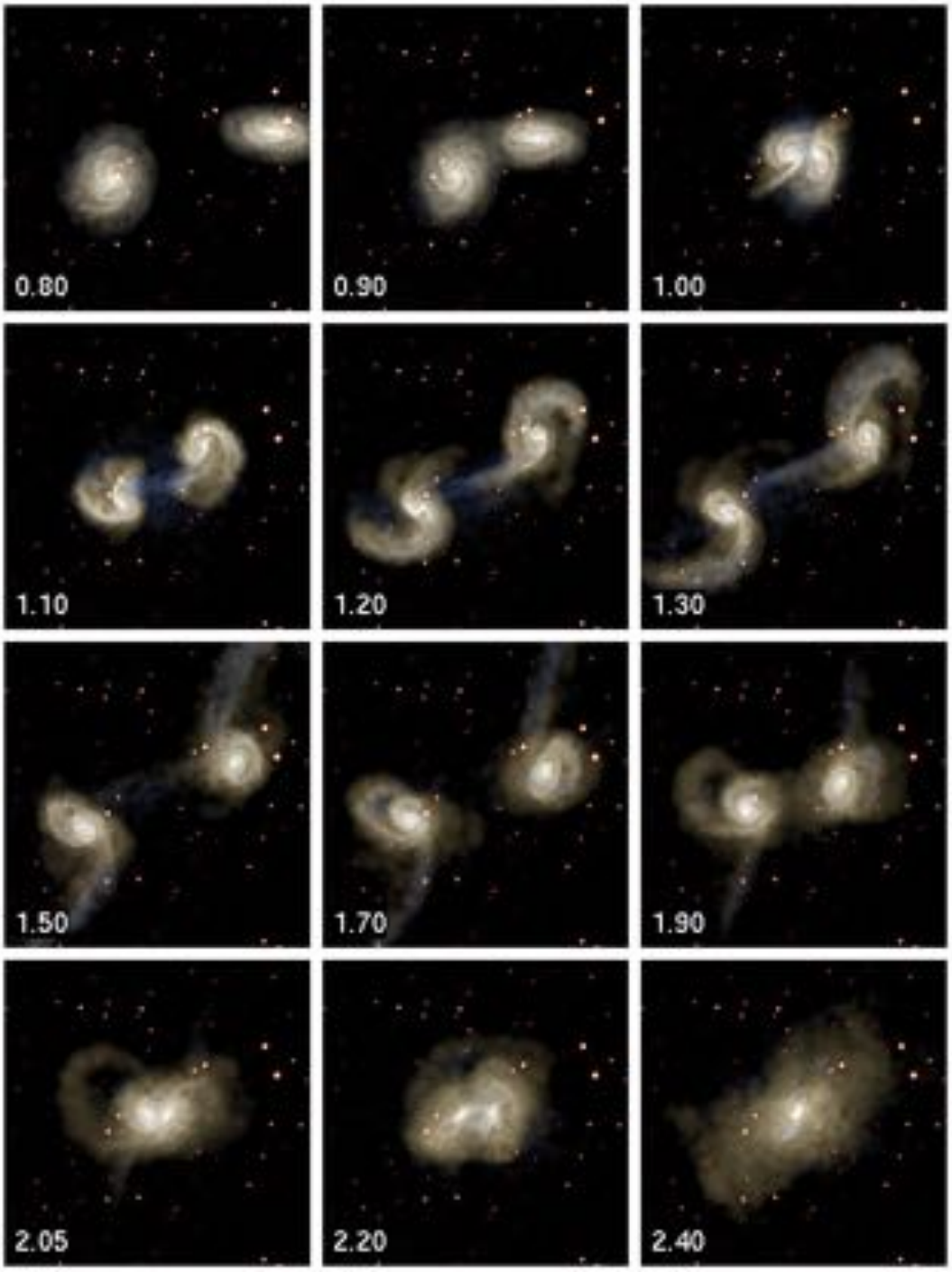


Figure 8.9 : Simulation of two merging galaxies with similar masses. The time is indicated in billion years. Source : Springel et al. (1999)

simple estimate.

Triggering of star formation during galaxy interactions is a complex process. To illustrate some of the physics. Figure 8.11 show the results of a simulation which includes a simple model for star formation which is determined by cloud build-up as well as following the gas in the galaxy. What we see is that as a result of strong perturbating the system with the passage of a satellite galaxy :

- Spiral density waves are induced with a very clear two-armed spiral in the gas
- Shocks and cloud-collisions dissipate the kinetic energy of the cloud population and gas accumulates on the inner Lindblad resonance.
- The high gas density leads to a burst of star formation
- After a further period of time, the gas dissipates more kinetic energy and accumulates in the nucleus of the galaxy giving rise to a second burst of star formation
- Models expect that a fraction of the gas accumulated in the nuclear region can accrete onto the nuclear supermassive black hole, i.e. trigger quasar activity (or more generally an AGN).

8.2.1 ULIRGs as merging systems

These expectations and predictions of theoretical models and numerical simulations are clearly observed in ULIRGs. The ultra-luminous galaxies (ULIRGs) invariably show strong signs of interactions and mergers. The sample of images of distant ULIRGs shows a range of disturbed morphologies. Tidal interactions and merging after the circular motion of clouds causing cloud-cloud collisions resulting into loss of angular momentum, which allows clouds to flow towards the center of the galaxy to feed star formation in the central region and, often, black hole accretion (i.e. AGN/quasar activity). Indeed, in ULIRGs we observe that star formation is generally concentrate in the central few hundred parsecs (1kpc at most) and that they also often host a luminous AGN (typically a quasar, which is however often deeply obscured by dust).

8.2.2 Nuclear fuelling through stellar bars

An additional mechanism that can drive gas towards the center, hence triggering a nuclear starburst or accretion onto a supermassive black hole, is the formation of a stellar bar (often triggered by even mild galaxy interaction) :

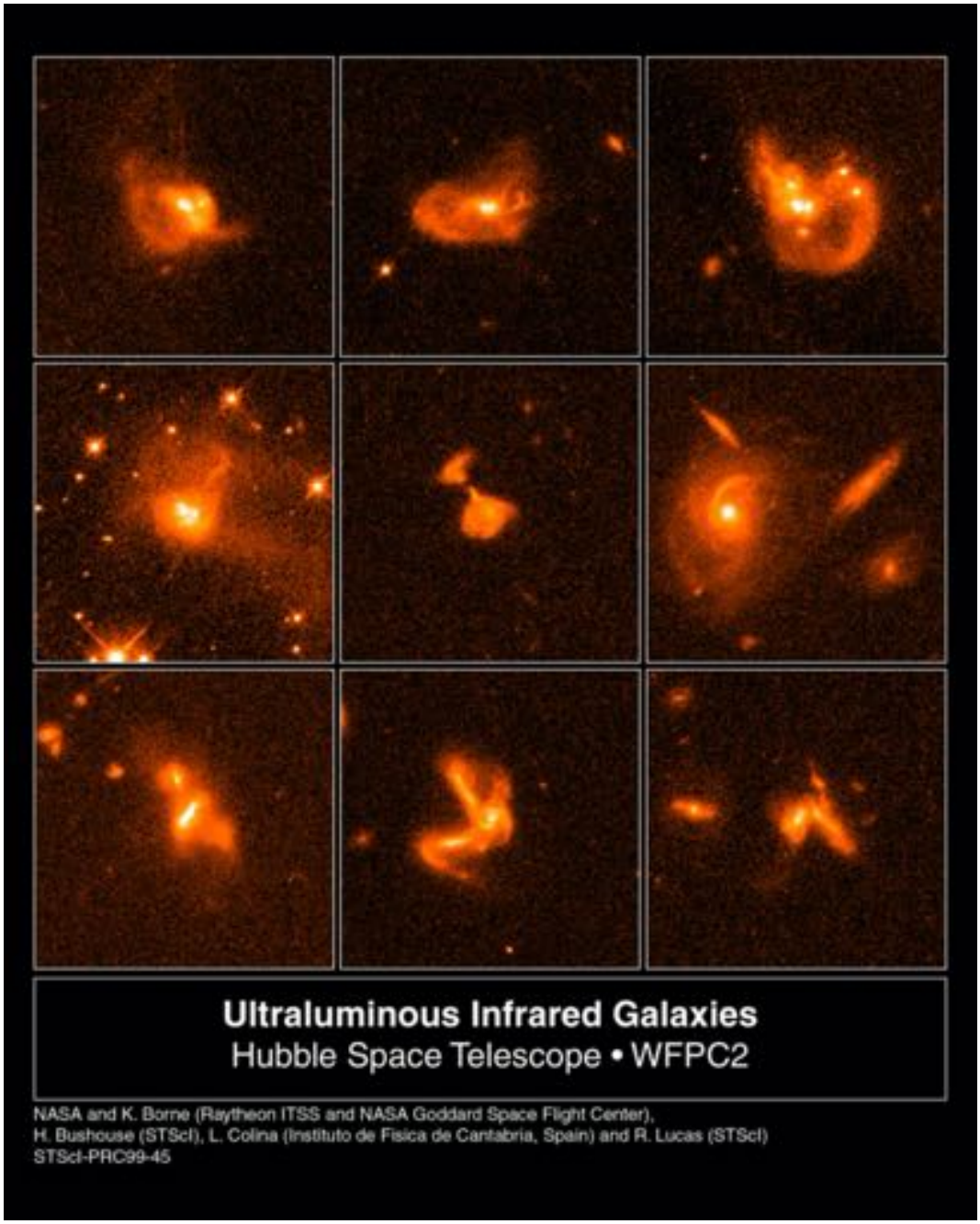


Figure 8.10 : Exemple of Ultra-Luminous Infrared Galaxies observed with the *Hubble* Space Telescope. Source : NASA/ ESA/ HST

- the bar is a very strong non-linear perturbation which can exist within the inner Lindblad resonance
- Gas on circular orbits encountering the bar at supersonic velocities is shocked
- kinetic energy is dissipated and the gas accumulates in the bar
- Orbits in the bar are highly elongated bringing gas and stars to the nucleus

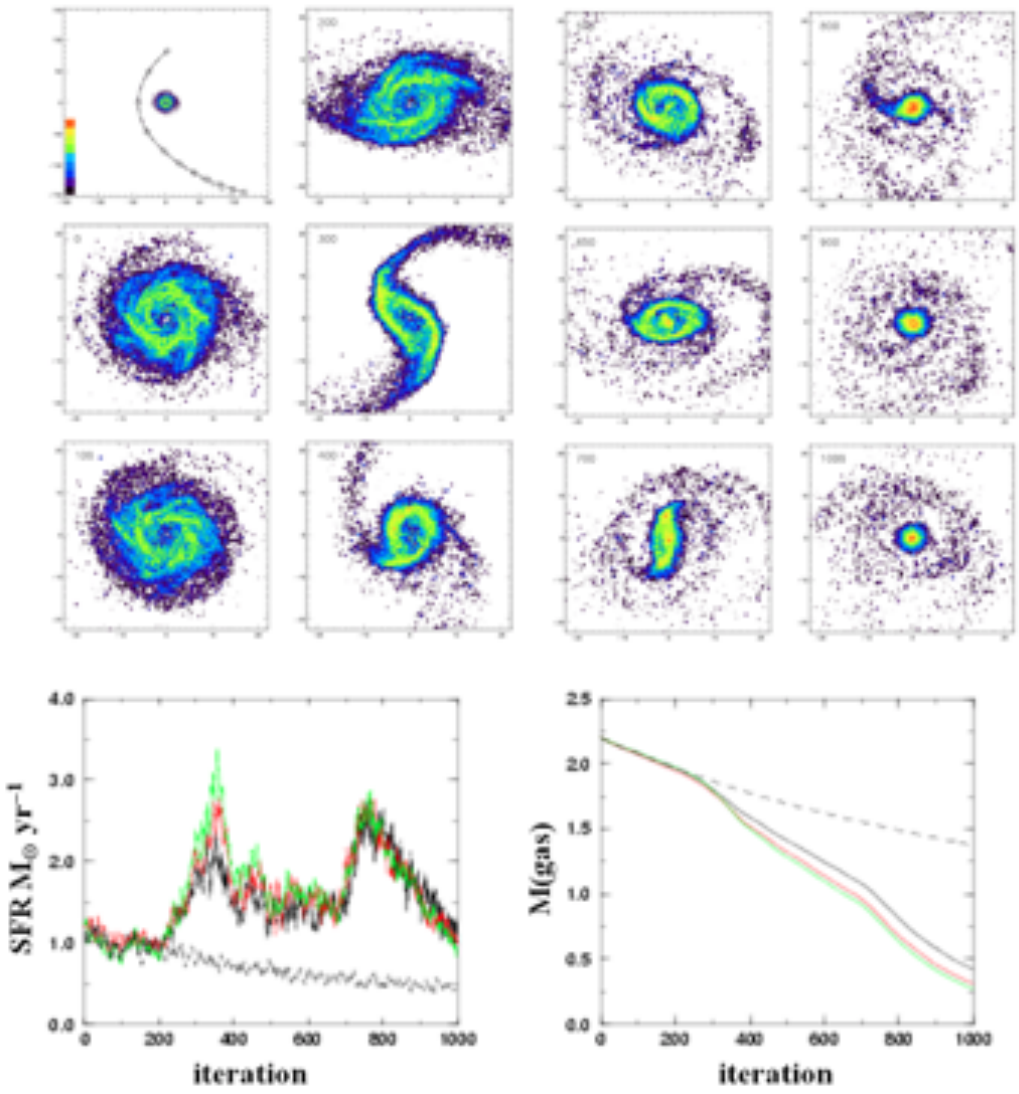


Figure 8.11 : Simulation of a disc galaxy including gas and a simple model for star formation. The images show the gas density. The system is perturbed by the passage of a satellite galaxy. Two bursts of star formation are triggered in this system.

9. Gravitational instabilities in the cosmological context

In a spiral galaxy, the ratio of dark-to-light matter is about a factor of 10. That's probably a good number for the ratio of our ignorance-to-knowledge. We're out of kindergarten, but only in about third grade.

— Vera Rubin

In this chapter we will describe the cosmological evolution of structure formation by considering the growth of perturbations in the early Universe. We will based our discussion on your last term course on Relativistic Astrophysics and Cosmology.

9.1 Starting point

We assume an expanding universe governed by the cosmological field equations as follow :

$$\frac{\ddot{R}}{R} + \frac{4\pi G\rho}{3}(1+\epsilon) - \frac{\Lambda}{3} = 0 \tag{9.1}$$

$$\left(\frac{\dot{R}}{R}\right)^2 - \frac{8\pi G\rho}{3} - \frac{\Lambda}{3} = -\frac{kc^2}{R^2} \tag{9.2}$$

We assume the so-called ΛCDM cosmology with :

- A flat geometry with $\Omega_\Lambda + \Omega_m = 1$
- $\Omega_\Lambda \sim 0.7$, $\Omega_m \sim 0.3$ and $\Omega_b \sim 0.044$
- The seeds of structure are quantum fluctuations which are amplified by inflation. This gives the Harrison-Zel'dovich spectrum in which the power in the fluctuations is approximately independent of linear scale.

We recall hereafter the definition of some cosmological parameters :

$$H = \frac{\dot{R}}{R} \tag{9.3}$$

$$\Omega_\Lambda = \frac{\Lambda}{3H^3} \tag{9.4}$$

$$\Omega_m = \frac{\rho}{3H^2/8\pi G} \tag{9.5}$$

$$1 - \Omega_m - \Omega_\Lambda = -\frac{kc^2}{R^2H^2} \tag{9.6}$$

9.2 Jeans instability in an expanding Universe

In the following section, we will start with the equations of hydrodynamics and will eventually do a perturbation analysis. Our approach will be to look at deviations from the smooth expansion of the universe in a co-moving frame. In this frame, we will have a background of the cosmological expansion, but the deviations we expect to be small and therefore can be approximated, not only by a perturbation analysis, but also by locally using non-relativistic equations for the fluid and treating gravitational perturbations as sufficiently small that a Newtonian approximation is justified.

9.2.1 Fluid equation in co-moving coordinates

The relationship between proper distance and comoving coordinates is :

$$r = R(t)\chi \tag{9.7}$$

Differentiating with respect to times gives the velocity :

$$u = \dot{R(t)}\chi + R(t)\dot{\chi} = \dot{R(t)}\chi + v \tag{9.8}$$

where we identify $\dot{R(t)}\chi$ as arising from the Hubble expansion and $v = R(t)\dot{\chi}$ is the peculiar velocity, i.e. the velocity which is superimposed on the Hubble flow. Given our approximation, the equations describing the fluid have their familiar form:

$$\text{The equation of continuity: } \left(\frac{\partial \rho}{\partial t}\right)_r + \nabla_r \cdot (\rho u) = 0 \tag{9.9}$$

$$\text{The Euler equation: } \left(\frac{\partial u}{\partial t}\right)_r + u \cdot \nabla_r u = -\frac{1}{\rho}\nabla_r p - \nabla\Phi \tag{9.10}$$

$$\text{The Poisson's equation : } \nabla^2\Phi = 4\pi G\rho \tag{9.11}$$

In these expressions, the differentials are with respect to proper distance. We now transform these equations to a co-moving frame. To do so, we note the following :
• Transforming the gradient to co-moving coordinates is given by:

$$\nabla_r \to \frac{1}{R}\nabla_\chi \tag{9.12}$$

• For the time derivative, this becomes :

$$\left(\frac{\partial}{\partial t}\right)_r \to \left(\frac{\partial}{\partial t}\right)_\chi + \left(\frac{\partial \chi}{\partial t}\right)_r \cdot \nabla_\chi = \left(\frac{\partial}{\partial t}\right)_\chi - \frac{\dot{R}}{R}\chi \cdot \nabla_\chi \tag{9.13}$$

where we have used :

$$\left(\frac{\partial \chi}{\partial t}\right)_r = \left(\frac{\partial r/R}{\partial t}\right)_r = -\frac{\dot{R}}{R^2}r = -\frac{\dot{R}}{R}\chi \tag{9.14}$$

We start by rewriting the equation of continuity (from now all partials are at constant χ) :

$$\frac{\partial \rho}{\partial t} - \frac{\dot{R}}{R}\chi \cdot \nabla_\chi \rho + \frac{1}{R}\nabla_\chi \cdot (\rho(\dot{R}\chi + v)) = \frac{\partial \rho}{\partial t} - \frac{\dot{R}}{R}\chi \cdot \nabla_\chi \rho \tag{9.15}$$

$$+ \frac{\dot{R}}{R}\chi \cdot \nabla_\chi \rho + \frac{\dot{R}}{R}\rho \nabla_\chi \cdot \chi + \frac{1}{R}\nabla_\chi \cdot (\rho v) \tag{9.16}$$

$$= \frac{\partial \rho}{\partial t} + 3\frac{\dot{R}}{R}\rho + \frac{1}{R}\nabla_\chi \cdot (\rho v) \tag{9.17}$$

or

$$\boxed{\frac{\partial \rho}{\partial t} + \frac{1}{R}\nabla_\chi \cdot (\rho v) = -3\rho\frac{\dot{R}}{R}} \tag{9.18}$$

The new term is just related to the change in density due to the overall cosmic expansion

We now transform the Euler equation, following the same approach, and we obtain:

$$\frac{\partial v}{\partial t} + \frac{1}{R}v \cdot \nabla_\chi v + \frac{\dot{R}}{R}v = -\frac{1}{R\rho}\nabla_\chi P - \frac{1}{R}\nabla_\chi \Phi - \ddot{R}\chi \tag{9.19}$$

The final term can be written as :

$$\ddot{R}\chi = \frac{1}{2}\ddot{R}\nabla_\chi \chi^2 = \frac{1}{R}\nabla_\chi \left(\frac{1}{2}R\ddot{R}\chi^2\right) \tag{9.20}$$

The reason for this manipulation is that we can define a new potential function $\phi_g = \Phi + \frac{1}{2}R\ddot{R}\chi^2$, and the Euler equation then becomes :

$$\boxed{\frac{\partial v}{\partial t} + \frac{1}{R}v \cdot \nabla_\chi v + \frac{\dot{R}}{R}v = -\frac{1}{R\rho}\nabla_\chi P - \frac{1}{R}\nabla_\chi \phi_g} \tag{9.21}$$

Before rewriting the Poisson's equation, we return to the cosmological field equations for the homogeneous universe :

$$\frac{\ddot{R}}{R} + \frac{4\pi G\bar{\rho}}{3}(1+\epsilon) - \frac{\Lambda}{3} = 0 \tag{9.22}$$

$$\left(\frac{\dot{R}}{R}\right)^2 - \frac{8\pi G\bar{\rho}}{3} - \frac{\Lambda}{3} = -\frac{kc^2}{R^2} \tag{9.23}$$

where we have explicitly included the mean density of the universe $\bar{\rho}$.
In the limit of a non-relativistic fluid and at early enough epochs so that we may ignore the Λ term, then the previous equations can be written as :

$$\frac{\ddot{R}}{R} + \frac{4\pi G\bar{\rho}}{3} = 0 \tag{9.24}$$

Finally we consider Poisson's equation :

$$\frac{1}{R^2}\nabla^2_\chi \phi_g = \frac{1}{R^2}\nabla^2_\chi\left(\Phi + \frac{1}{2}R\ddot{R}\chi^2\right) \tag{9.25}$$

in spherical polar coordinates :

$$\nabla^2_\chi \frac{1}{2}\chi^2 = \frac{1}{2\chi^2}\frac{\partial}{\partial\chi}\chi^2\frac{\partial\chi^2}{\partial\chi} = 3 \tag{9.26}$$

Substituting for $\nabla^2\Phi$ from the Poisson's equation and using the field equation to eliminate $\ddot{R}$ we get :

$$\boxed{\frac{1}{R^2}\nabla^2_\chi \phi_g = 4\pi G\rho - 4\pi G\bar{\rho} = 4\pi G\bar{\rho}\Delta} \tag{9.27}$$

where we have for the density :

$$\rho = \bar{\rho}(1+\Delta) \tag{9.28}$$

where Δ is the density contrast.

9.2.2 Governing equations for the overdensity

We can make further simplifications by keeping only terms which are first order in small quantities. We have already introduced v, the small peculiar velocity, and for the pressure, assuming an equation of state in which P is only a function of ρ, we expand about its mean value $P(\bar{\rho})$:

$$P \approx P(\bar{\rho}) + \frac{\mathrm{d}P}{\mathrm{d}\rho}\bar{\rho}\Delta = P(\bar{\rho}) + c_s^2\bar{\rho}\Delta \tag{9.29}$$

From the cosmological principle $\nabla P(\bar{\rho}) = 0$ and we assume the adiabatic sound speed c_s^2 to be a constant. To first order, in small quantities we have :

$$\frac{\partial \Delta}{\partial t} + \frac{1}{R}\nabla_\chi \cdot v = 0 \tag{9.30}$$

$$\frac{\partial v}{\partial t} + \frac{\dot{R}}{R}v = -\frac{c_s^2}{R}\nabla_\chi \Delta - \frac{1}{R}\nabla_\chi \phi_g \tag{9.31}$$

$$\frac{1}{R^2}\nabla_\chi^2 \phi_g = 4\pi G\bar{\rho}\Delta \tag{9.32}$$

To complete the analysis, we take the time derivative of the continuity equation, the divergence of the Euler equation and eliminate the dependence of the peculiar velocity. Finally we seek solutions of the form :

$$\Delta = \Delta(t)\exp(ik_c \cdot \chi) \tag{9.33}$$

This gives the following equation for the time dependant overdensity for wave number $k = k_c/R$:

$$\boxed{\frac{\mathrm{d}^2\Delta}{\mathrm{d}t^2} + 2\left(\frac{\dot{R}}{R}\right)\frac{\mathrm{d}\Delta}{\mathrm{d}t} = (4\pi G\rho - k^2c_s^2)\Delta} \tag{9.34}$$

9.2.3 The growth of instabilities

The growth of density perturbations is governed in the linear regime by eq.9.34. Let's contrast the growth of instabilities in a non-expanding universe.

We recover the analysis we performed for the stability of an interstellar cloud as follows :

- recover the equivalent to a static universe by setting : $\dot{R} = 0$
- seek solutions of the form : $\Delta = \Delta_0 \exp i(kc \cdot \chi - \omega t)$

We then get the dispersion relation we had before :

$$\omega^2 = c_s^2 k^2 - 4\pi G\rho \tag{9.35}$$

This gives exponentially growing modes for $c_s^2 k^2 < 4\pi G\rho$, or :

$$\lambda > \lambda_J = c_s \left(\frac{\pi}{G\rho}\right)^{1/2} \tag{9.36}$$

when $\lambda \gg \lambda_J$ the modes grow like $\exp(t/\tau)$ where $\tau \sim (4\pi G\rho)^{1/2}$.

In the expanding Universe, we consider the simplest model of a flat universe the Einstein-de-Sitter universe which has the following properties :

$$\Omega_0 = 1 \tag{9.37}$$

$$\frac{R}{R_0} = \left(\frac{t}{t_0}\right)^{2/3} = \left(\frac{3}{2} H_0 t\right)^{2/3} \tag{9.38}$$

$$H^2 = \frac{8\pi G\rho}{3} \tag{9.39}$$

$$4\pi G\bar{\rho} = \frac{2}{3t^2} \tag{9.40}$$

We also consider the case where the gravitational attraction is much stronger than the pressure force :

$$4\pi G\bar{\rho} \gg c_s^2 k^2 \tag{9.41}$$

The equation for Δ reduced to :

$$\frac{\mathrm{d}^2\Delta}{\mathrm{d}t^2} + \frac{4}{3t}\frac{\mathrm{d}\Delta}{\mathrm{d}t} - \frac{2}{3t^2}\Delta = 0 \tag{9.42}$$

It is easy to verify that for solutions of the form $\Delta \propto t^n$, the growing modes have :

$$\boxed{\Delta \propto t^{2/3} \propto R \propto (1+z)^{-1}} \tag{9.43}$$

The key element of this result is that the perturbations only grow algebraically, and not exponentially with time. This basic result is similar for other cosmologies as well.

We can modify this analysis to account for the early radiation dominated phase of the Universe :

- the expression $4\pi G$ in the equation for Δ is replaced by $32\pi G/3$
- furthermore in the radiation dominated phase of the universe $R \propto t^{1/2}$

we find that :

$$\Delta \propto t \propto R^2 \propto (1+z)^{-2} \tag{9.44}$$

9.3 The need for dark matter

The results of the previous section lead to a major problem. We start by restating the problem then provide a detailed analysis of how dark matter resolves the problems.

In the early, radiation dominated, universe, when matter and radiation are strongly coupled : $\rho \propto 1/R^3 \propto T^3$, giving :

$$\Delta = \frac{\delta\rho}{\rho} \approx 3\frac{\delta T}{T} \tag{9.45}$$

From Planck observations of the CMB, we have an excellent measurement of $\frac{\delta T}{T} \sim 10^{-5}$, implying $\Delta \approx 3 \times 10^{-5}$ at the epoch of recombination ($z \approx 1500$). After recombination, the matter perturbations grow via gravity approximately according to the results of the previous section implying :

$$\Delta(t = t_0) \sim 0.05 \tag{9.46}$$

which is not the case !

9.4 Perturbations with dark matter

Let us now consider the presence of baryonic and dark matter represented by density contrasts Δ_B and Δ_D respectively. We now need to solve the coupled equations :

$$\frac{\mathrm{d}^2\Delta_B}{\mathrm{d}t^2} + 2\left(\frac{\dot{R}}{R}\right)\frac{\mathrm{d}\Delta_B}{\mathrm{d}t} = A\bar{\rho_B}\Delta_B + A\bar{\rho_D}\Delta_D \tag{9.47}$$

$$\frac{\mathrm{d}^2\Delta_D}{\mathrm{d}t^2} + 2\left(\frac{\dot{R}}{R}\right)\frac{d\Delta_D}{dt} = A\bar{\rho_B}\Delta_B + A\bar{\rho_D}\Delta_D \tag{9.48}$$

$$\tag{9.49}$$

which are valid for both matter dominated universe and for the radiation dominated case depending on the choice of the constant A.

Consider first the epoch immediately after recombination so that $A = 4\pi G$:

- To illustrate the result we consider the case where $\Omega_D \approx 1$ and $\Omega_B \ll 1$
- the second equation reduces to the case we have already considered :

$$\bar{\rho_b}\Delta_B \ll \bar{\rho_D}\Delta_D \tag{9.50}$$

and

$$R = R_0 \left(\frac{3}{2} H_0 t\right)^{2/3} \tag{9.51}$$

- We write the solution as $\Delta_B = BR$ where B is a constant.

The equation for Δ_B is then :

$$\frac{d^2\Delta_B}{dt^2} + 2\left(\frac{\dot{R}}{R}\right)\frac{d\Delta_B}{dt} = 4\pi G\bar{\rho_D}\Delta_D \tag{9.52}$$

changing independent variable to R, we find (after a little algebra) :

$$R^{3/2}\frac{d}{dR}\left(R^{-1/2}\frac{d\Delta_B}{dR}\right) + 2\frac{d\Delta_B}{dR} = \frac{3}{2}B \tag{9.53}$$

which has the solution $\Delta_B = B(R - R_0)$, or :

$$\Delta_B = \Delta_D\left(1 - \frac{z}{z_0}\right) \tag{9.54}$$

which implies that the amplitude of the baryonic perturbation quickly grows to that of the dark matter no matter how small the baryonic perturbations is at $z = z_0$, e.g. at recombination.

9.5 Evolution of perturbations

We can now begin to understand the process of structure formation :

- In the very early Universe, any initial perturbations on scales larger than the Jeans length grow as R^2 since we are radiation dominated. At this stage the amplitude of the perturbations in baryonic matter, dark matter and radiation are equal since they are coupled.
- If the dark matter is made of heavy particules, as we assume in the Cold Dark Matter (CDM) cosmologies, then the cold dark matter decouples from the radiation at an early epoch.
- At a redshift of $z_{eq} \approx 4 \times 10^4 \Omega h^2$ the universe switches from radiation to matter dominated although the baryonic gas remains tightly coupled to the radiation
- Perturbations in the CDM grow at a rate of $\propto R$, however perturbations in the radiation and baryonic gas are damped :

  - photon pressure and diffusion of photons are both important and indeed the fluctuations in the photon+baryon fluid oscillate rather than grow
  - this explains the low values of $\Delta T/T$ which are observed

- After the epoch of recombination, the baryonic matter decouples from the radiation and quickly perturbations in the baryonic gas start to follow those in the CDM as we have shown.

To finish this part of our discussion, we calculate the Jeans mass just after the recombination ; this will give an indication of the typical masses of the first structure to form.
The Jeans length is given by :

$$\lambda_J = c_s \left(\frac{\pi}{G\rho}\right)^{1/2} \tag{9.55}$$

and the sound speed is $c_s^2 = 5k_BT/3m_H$.
The Jeans mass is the mass within this perturbations :

$$M_J = \frac{4\pi}{3}\left(\frac{\lambda_J}{2}\right)^3 \rho_m \tag{9.56}$$

and $\rho_m = (1+z_{eq})^3\rho_c\Omega_m$.
Therefore, the numerical application of the previous equation gives $M_J \sim 3 \times 10^5 M_\odot$.

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

Find the joy and satisfaction in what you do, and don't be too swayed by what other people think of your choices.

— Hiranya Peiris

In this chapter we continue our discussion of cosmological structure formation. Firstly we discuss the primordial perturbation spectrum and how this evolves. In order to discuss galaxy formation we need to consider how the small perturbations we discussed in the last chapter grow - this requires us to follow their evolution into the non-linear regime. We will then discuss the hierarchical clustering model for structure formation. All of this will be done considering just cold dark matter. To form real galaxies, we then discuss how the baryonic gas evolves and how the system form stars.

10.1 The power spectrum of the fluctuations

10.1.1 The two-point correlation function

The power-spectrum[1] of the fluctuations is a key parameter for structure formation: it determines the mass-spectrum of the initial perturbations and also the initial spatial distribution of structure.

We start by defining the *two-point correlation function*, which describes the excess of probability of finding a density enhancement (i.e. a galaxy) at distance r from a density enhancement, randomly selected in a uniform, random distribution. The

[1] By definition, the power spectrum is commonly defined as the Fourier transform of the auto-correlation function

number of galaxies in the volume element dV at distance r from any galaxy is :

$$\mathrm{d}N(r) = N_0[1 + \xi(r)]\mathrm{d}V \tag{10.1}$$

where N_0 is the average background number density of galaxies. $\xi(r)$, the two-point correlation function, can also be written in terms of the probability of finding pairs of galaxies separated by distance r :

$$dN_{pair} = N_0^2[1 + \xi(r)]\mathrm{d}V_1\,\mathrm{d}V_2 \tag{10.2}$$

The two-point correlation function can be directly related to the density contrast $\Delta = \delta\rho/\rho$. We write $\rho = \rho_0[1 + \Delta(x)]$ and so the pairwise numbers of galaxies separated by r is :

$$\mathrm{d}N_{pair}(r) = \rho(x)\mathrm{d}V_1\,\rho(x+r)\mathrm{d}V_2 \quad = \rho_0^2[1 + \Delta(x)][1 + \Delta(x+r)]\,\mathrm{d}V_1\,\mathrm{d}V_2 \tag{10.3}$$

Taking averages over a large number of volume elements, the mean value of Δ is zero by definition, and therefore the two-point correlation function becomes :

$$\mathrm{d}N_{pair}(r) = \rho_0^2[1 + < \Delta(x)\Delta(x+r) >]\,\mathrm{d}V_1\,\mathrm{d}V_2 \tag{10.4}$$

hence :

$$\xi(r) =< \Delta(x)\Delta(x+r) > \tag{10.5}$$

10.1.2 The power-spectrum of the fluctuations

We start by defining the Fourier transform for $\Delta(r)$:

$$\Delta(\vec{r}) = \frac{V}{(2\pi)^3}\int \Delta_k \exp(-i\vec{k}\cdot\vec{r})\,\mathrm{d}^3\vec{k} \tag{10.6}$$

$$\text{with } \Delta_k = \frac{1}{V}\int \Delta(r)\exp(i\vec{k}\cdot\vec{r})\,\mathrm{d}^3\vec{x} \tag{10.7}$$

Parseval's theorem [2] gives :

$$\frac{1}{V}\int \Delta^2(\vec{r})d^3x = \frac{V}{(2\pi)^3}\int |\Delta_k|^2\,\mathrm{d}^3\vec{k} \tag{10.8}$$

where $|\Delta_k|^2$, also noted $P(k)$ is the power spectrum of the fluctuations.
The left hand side of eq.10.8 is the mean squared density contrast, hence :

$$< \Delta^2 >= \frac{V}{(2\pi)^3}\int |\Delta_k|^2 d^3k = \frac{V}{(2\pi)^3}\int P(k)\,\mathrm{d}^3\vec{k} \tag{10.9}$$

[2] In mathematics, Parseval's theorem usually refers to the result that the Fourier transform is

Since the two-point correlation function is spherically symmetric, the element of k-space can be written $\mathrm{d}^3\vec{k} = 4\pi k^2\,\mathrm{d}k$, then:

$$< \Delta^2 >= \frac{V}{2\pi^2}\int |\Delta_k|^2 k^2\,\mathrm{d}k = \frac{V}{2\pi^2}\int P(k)k^2\,\mathrm{d}k \tag{10.10}$$

We can also write $\Delta(x)$ as a fourier series :

$$\Delta(\vec{x}) = \Sigma_k \Delta_k \exp(-i\vec{k}\cdot\vec{x}) \tag{10.11}$$

But $\xi(\vec{r}) =< \Delta(\vec{x})\Delta(\vec{x}+\vec{r}) >$, hence :

$$\xi(\vec{r}) = \Big\langle \sum_k \sum_{k'} \Delta_k \Delta^*_{k'} e^{-i(\vec{k}-\vec{k'})\cdot\vec{x}} e^{i\vec{k'}\cdot\vec{r}} \Big\rangle \tag{10.12}$$

Given the orthogonality of the Fourier basis, the cross terms vanish except for those for which $k = k'$, and therefore :

$$\xi(r) = \sum |\Delta_k|^2 \exp(i\vec{k}\cdot\vec{r}) \tag{10.13}$$

or in terms of Fourier integral :

$$\xi(r) = \frac{V}{(2\pi)^3}\int |\Delta_k|^2 e^{i\vec{k}\cdot\vec{r}}\mathrm{d}^3\vec{k} \tag{10.14}$$

Since $\xi(r)$ is real, we are only interested in the integral of the real part of $e^{ik\cdot r}$, i.e $\cos(k\cdot r) = \cos(kr\cos\theta)$. Moreover, because of the spherical symmetry of the two-point correlation function, we integrate over the angular part of the volume element $\frac{1}{2}\sin\theta\,\mathrm{d}\theta$:

$$\xi(r) = \frac{V}{2\pi^2}\int |\Delta_k|^2 \frac{\sin kr}{kr} k^2\,\mathrm{d}k = \frac{V}{2\pi^2}\int P(k)\frac{\sin kr}{kr}k^2\,\mathrm{d}k \tag{10.15}$$

Similarly, the inverse transform gives for the power spectrum :

$$P(k) = \frac{1}{V}\int_0^\infty \xi(r)\frac{\sin kr}{kr}4\pi r^2\,\mathrm{d}r \tag{10.16}$$

10.1.3 The initial Power-Spectrum

The observations of the CMB discussed last term suggest a power-law form for the perturbation spectrum with no preferred scale :

$$P(k) = |\Delta_k|^2 \propto k^n \tag{10.17}$$

then $\xi(r)$ has the following form :

$$\xi(r) \propto \int \frac{\sin kr}{kr} k^{n+2} dk \tag{10.18}$$

- $\frac{\sin kr}{kr} \sim 1$ for $kr \ll 1$ and decreases rapidly when $kr \gg 1$, hence we can integrate k from 0 to $k_{max} \approx 1/r$ to estimate the dependence of the amplitude of the correlation function on the scale r :

$$\xi(r) \propto r^{-(n+3)} \tag{10.19}$$

- the mass in the perturbation is $M \sim \rho r^3$, hence :

$$\xi(M) \propto M^{-(n+3)/3} \tag{10.20}$$

- The density contrast for a mass scale M, $\Delta(M)$, is :

$$\Delta(M) =< \Delta^2 >^{1/2} \propto M^{-(n+3)/6} \tag{10.21}$$

10.1.4 The Harisson-Zel'dovich Power Spectrum

The Harisson-Zel'dovich power spectrum has $n = 1$ hence :

$$\Delta(M) \propto M^{-2/3} \qquad \text{and} \qquad \xi \propto r^{-4} \propto M^{-4/3} \tag{10.22}$$

10.1.5 Evolution of the Power Spectrum and transfer functions

The power spectrum is modified from its initial form as the universe evolves. This is an extremely complicated process since it involves the physical interaction of three fluids in an expanding universe: baryonic matter, non-baryonic matter and photons. We can only give a flavour of this subject and show some results of detailed calculations.

The transfer function $T(k)$ describes how the shape of the initial power spectrum $\Delta_k(z)$ in the dark matter is modified by different physical processes :

$$\Delta_k(z = 0) = T(k) f(z) \Delta_k(z) \tag{10.23}$$

where :

- $\Delta_k(z = 0)$ is the power spectrum at the present epoch
- $f(z) \propto R$ is the linear growth factor between the scale factor at redshift z and the present epoch.

In the following, we will outline one process which is important for the adiabatic

unitary; loosely, that the sum (or integral) of the square of a function is equal to the sum (or integral) of the square of its transform.

cold-dark matter we have been considering.

Consider an initial power-law power spectrum such as:

$$P(k) = |\Delta_k|^2 \propto k^n \tag{10.24}$$

A critical point in the evolution of the perturbations is when their size is equal to the horizon size. For a perturbation of size r this happens when $r \approx ct$ - we say that the perturbation has come trough, or entered, the horizon.

- before the perturbation entered the horizon during the radiation-dominated era, their density contrasts grew as $\Delta_k \propto R^2$ on all scales
- if the perturbations came through the horizon during the radiation-dominated phase, the dark matter perturbations were gravitationally coupled to the radiation-dominated plasma, and their amplitudes were then stabilised.
- therefore as soon as the perturbations came through the horizon the perturbations ceased to grow until the epoch of equality.
- after that time, all perturbations grew as $\Delta_k \propto R$

Between crossing their particle horizons at scale factor R_H and the epoch of equality R_{eq}, the amplitudes of the perturbations were damped by a factor $(R_H/R_{eq})^2$ relative to the unmodified spectrum :

$$\Delta_k \propto k^{n/2} \left(\frac{R_H}{R_{eq}}\right)^2 \tag{10.25}$$

Since $k \propto R_H^{-1}$ it follows that the transfer function $T(z)$ has the asymptotic forms :

$$T_k = 1 \qquad \text{for} \qquad M \geq M_{eq}, k \leq k_{eq} \tag{10.26}$$

$$T_k \propto k^{-2} \qquad \text{for} \qquad M \leq M_{eq}, k \geq k_{eq} \tag{10.27}$$

Thus, for small masses, the 'processed' power spectrum $P(k) \propto T_k^2$ is flatter than the input spectrum of perturbations by a power k^{-4} :

$$P(k) = |\Delta_k|^2 \propto k^{n-4} \tag{10.28}$$

and

$$\xi(r) \propto r^{-(n-1)} \text{ or } \xi(M) \propto M^{-(n-1)/3} \tag{10.29}$$

- For the Harisson-Zel'dovich spectrum with $n = 1$, the processed two-point correlation function is flat at small wavelengths.
- Detailed modelling of the transfer function shows that in addition to this large-scale damping we expect acoustic oscillations as discussed last term, and observed in the power spectrum of the CMB.

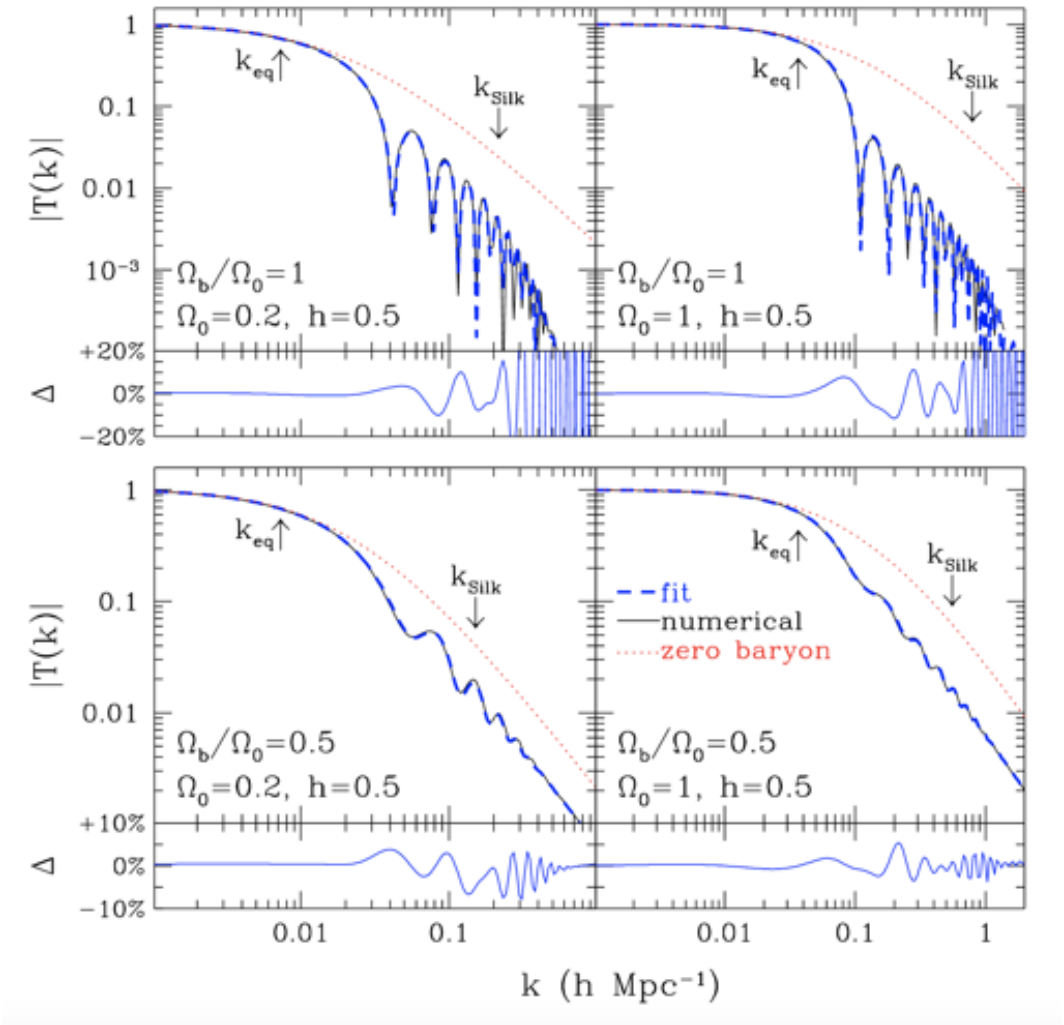


Figure 10.1 : Four examples of the transfer functions for models of structure formation with baryons only (top) and with mixed cold and baryonic matter. From Einsenstein and Hu 1998

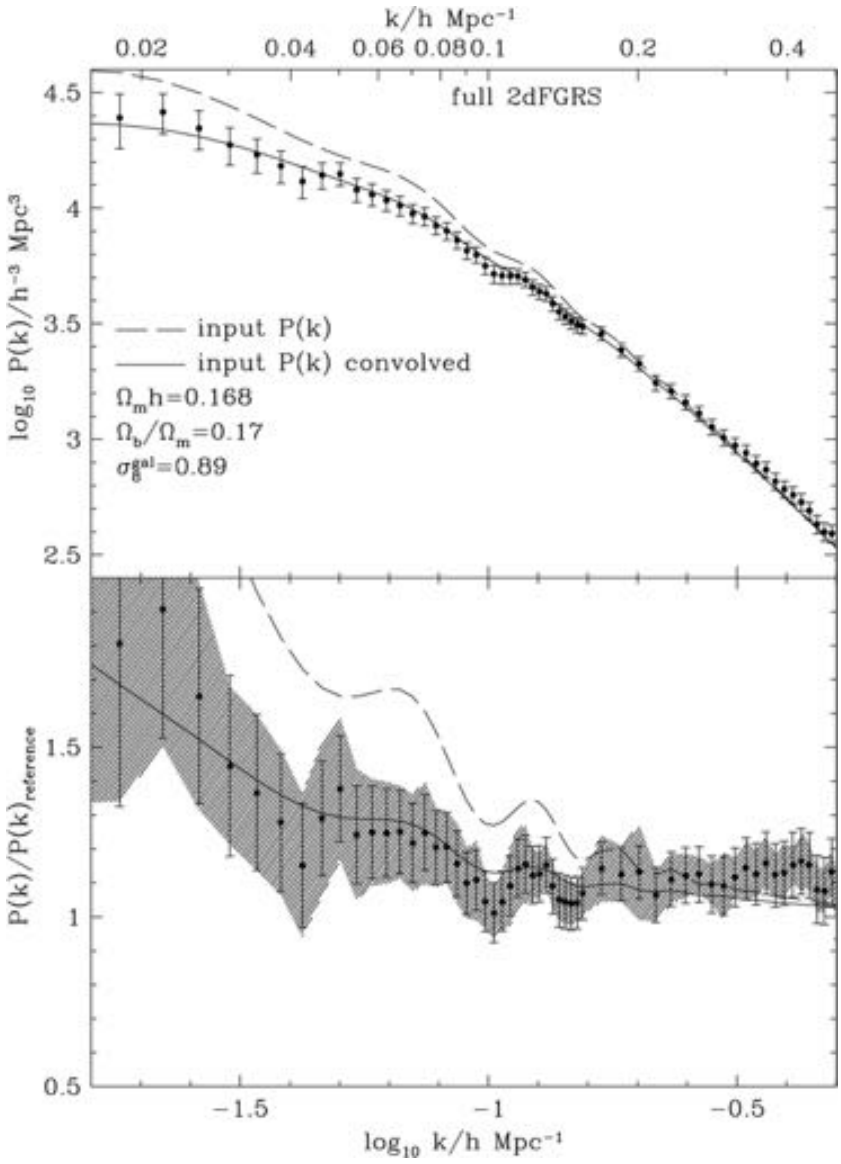


Figure 10.2 : The power spectrum of the three-dimensional distribution of galaxies in the 2dF Galaxy Redshift Survey. The points with error bars are the best estimates of the observed power spectrum once the biases and corrections for incompleteness are taken into account. In the lower panel, the data from the upper panel have been divided by a reference cold dark matter model with $\Omega_D = 0.2$, $\Omega_\Lambda = 0$ and $\Omega_B = 0$ which has a smooth power-spectrum. The grey dashed-line is a best fitting model before convolution with the window function for the survey. The solid line shows the best fit once the model is convolves with the window function.

10.2 Non-linear collapse of a spherical overdensity

To model the evolution of perturbations into the non-linear regime, we consider a single spherical overdensity. Although this is clearly an unrealistic model, it has the advantage of having an analytical solution. The reason for this is that just as in the classical Gauss's theorem, in general relativity for a perfectly spherical geometry the mass inside a sphere is not influenced by the material outside the sphere. We can therefore consider a sphere of gas as if it is a separate universe. This model was introduced in the Relativistic Astrophysics and Cosmology course. Now we find a detailed solution for this case.

Our assumptions are as follows :

- We imagine just one spherical overdense region of radius R_s and density ρ_s
- For the background cosmology, we consider a spatially flat universe and for algebraic simplicity we model this as an Einstein-de-Sitter universe. We write for the mean density of the Universe $\bar{\rho}$ with scale factor R.
- To fix our ideas, we take recombination as a reference epoch. At this epoch, we take the mean density in the universe to be given by $\rho_0 = \bar{\rho}(t_0)$ and the scale factor for the Universe is R_0. At this epoch, the mass and radius of the overdensity are M and R_{s_0} and its density is :

$$\rho_{s_0} = \frac{3M}{4\pi R_{s_0}^3} \tag{10.30}$$

- To simplify the algebra further, we choose a co-moving coordinates such that the outer radius of our spherical overdensity has $\chi = 1$, then the scale factor is just the radius our overdensity would have if it expended with the Hubble flow in our flat universe. At $t = t_0$ therefore $R_{s_0} = R_0$.

The evolution of both the universe as a whole and the overdensity are described by the field equations :

$$\frac{\ddot{R}}{R} + \frac{4\pi G\rho}{3}(1+\epsilon) - \frac{\Lambda}{3} = 0 \tag{10.31}$$

$$\left(\frac{\dot{R}}{R}\right)^2 - \frac{8\pi G\rho}{3} - \frac{\Lambda}{3} = -\frac{kc^2}{R^2} \tag{10.32}$$

10.2.1 Background Universe

For our Einstein-de-Sitter model, $k = 0$, $\Lambda = 0$, $\epsilon = 0$, and we have the standard solution :

$$\Omega_m = 1 \tag{10.33}$$

$$\frac{R}{R_0} = \left(\frac{t}{t_0}\right)^{2/3} = \left(\frac{3}{2}H_0 t\right)^{2/3} \tag{10.34}$$

$$H^2 = \frac{8\pi G\bar{\rho}}{3} \tag{10.35}$$

$$4\pi G\bar{\rho} = \frac{2}{3t^2} \tag{10.36}$$

10.2.2 Evolution of the overdensity

The overdensity means that for the region of the Universe within the radius R_s the universe must be closed (since we have an overdensity) and hence $k = 1$.
The equation for the radius of the overdensity follows from eq.10.32 :

$$\left(\frac{\dot{R}_s}{R_s}\right)^2 - \frac{8\pi G}{3}\frac{\rho_{s_0}R_{s_0}^3}{R_s^3} = -\frac{kc^2}{R_s^2} \tag{10.37}$$

At $t = t_0$ the radius of the overdensity is R_{s_0}, the mean cosmological density is ρ_0 and the overdensity is $\Omega_{s_0}\rho_0$. To solve this equation, we have to follow the following steps :

- we change independent variable to the conformal time : $\eta = \int_0^t \frac{cdt'}{R(t')}$
- we introduce a dimensionless radius : $a = R_s/R_{s_0}$
- we define the constant $a_m = 8\pi G\rho_{s_0}R_{s_0}^2/3c^2$
- we also define :

$$\Omega_s = \frac{8\pi G\rho_s}{3H_s^2} \qquad \Omega_{s_0} = \frac{8\pi G\rho_{s_0}}{3H_{s_0}^2} \tag{10.38}$$

  If we further assume that the perturbation is still approximately following the Hubble flow at t_0 then $H_{s_0} \approx H_0$ and $\rho_{s_0} \approx \Omega_{s_0}\rho_0$ where ρ_0 is the mean density in our flat Einstein-de-Sitter universe at this epoch which we have taken to be recombination.

With these definitions, we get the following differential equation :

$$\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}\eta}\right)^2 = \frac{a_m}{a} - 1 \tag{10.39}$$

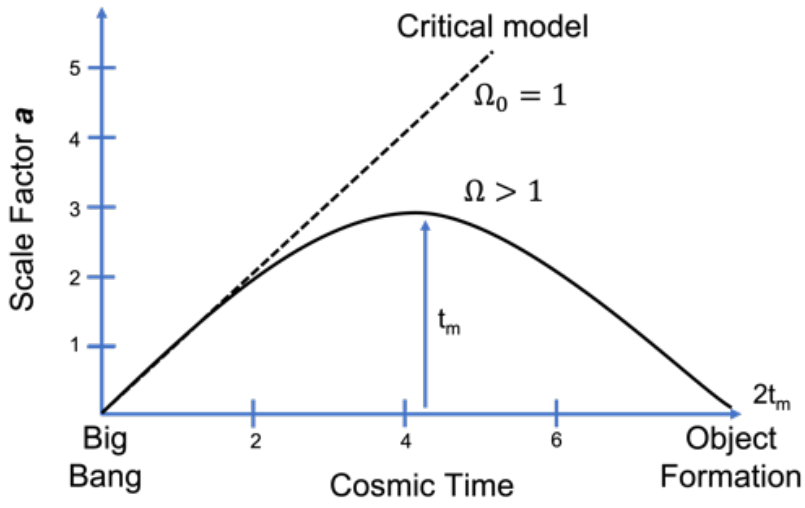


Figure 10.3 : Non-linear collapse of a spherical overdensity

The solution with appropriate boundary conditions for our Big-Bang cosmology is :

$$a = \frac{R_s}{R_{s_0}} = \frac{a_m}{2}(1 - \cos\eta) \tag{10.40}$$

$$t = \frac{R_{s_0}}{c}\frac{a_m}{2}(\eta - \sin\eta) \tag{10.41}$$

From the general solution to the dynamical equations, the overdensity will reach a maximum radius, R_m at a time t_m where :

$$R_m = a_m R_{s_0} = \frac{\Omega_{s_0}}{\Omega_{s_0} - 1} R_{s_0} \tag{10.42}$$

$$t_m = \frac{\pi}{2}\frac{R_{s_0}}{c} a_m = \frac{\pi \Omega_{s_0}}{2H_{s_0}(\Omega_{s_0} - 1)^{3/2}} \tag{10.43}$$

where we have used $R_{s_0} = \frac{c}{H_{s_0}(\Omega_{s_0} - 1)^{1/2}}$.

- The overdensity will then collapse to its final state at a time $t = 2t_m$
- We call the point where the overdensity reaches its maximum radius the **turn around**
- The density within the collapsing overdense region at turn around is :

$$\rho_s(t_m) = \rho_{s_0}(R_{s_0}/R_m)^3 \approx \rho_0 \Omega_{s_0}(R_{s_0}/R_m)^3 = \rho_0 \Omega_{s_0}\left(\frac{\Omega_{s_0} - 1}{\Omega_{s_0}}\right)^3 \tag{10.44}$$

whereas the mean density in the Universe at this time will be :

$$\bar{\rho}(t_m) = \rho_0 (R_{s_0}/R(t_m))^3 \tag{10.45}$$

where $R(t_m)$is the radius a sphere of radius R_{s_0} at time t_0 would have in the smooth cosmological expansion at the time t_m.

For an Einstein-de-Sitter cosmology :

$$\frac{R(t_m)}{R_{s_0}} = \left(\frac{3}{2} H_0 t_m\right)^{2/3} \tag{10.46}$$

hence

$$\frac{\rho_s(t_m)}{\rho(t_m)} \approx \frac{(\Omega_{s_0} - 1)^3 \Omega_{s_0}^2}{(3H_0 t_m/2)^2} \tag{10.47}$$

but

$$H_0 t_m \approx H_{s_0} t_m = \frac{\pi \Omega_{s_0}}{2(\Omega_{s_0} - 1)^{3/2}} \tag{10.48}$$

Therefore

$$\frac{\rho_s(t_m)}{\bar{\rho}(t_m)} \approx \left(\frac{3\pi\Omega_{s_0}}{4\pi(\Omega_{s_0} - 1)^{3/2}}\right)^2 \times \frac{(\Omega_{s_0} - 1)^3}{\Omega_{s_0}^2} = \left(\frac{3\pi}{4}\right)^2 \tag{10.49}$$

Therefore independent of the initial overdensity, we find at turn around :

$$\frac{\rho_s(t_m)}{\bar{\rho}(t_m)} \sim 5.6 \tag{10.50}$$

The collapsed object forms at a time $2t_m$. The redshifts of turn around, z_m, and the formation redshift of the collapsed object, z_f, are related in the Einstein-de-Sitter universe by :

$$\frac{1 + z_m}{1 + z_f} = \frac{R(2t_m)}{R(t_m)} = 2^{2/3} \approx 1.59 \tag{10.51}$$

The collapse is halted by the internal pressure and the end state will be determined by the virial equilibrium : object is virialised :

$$2K_v + \Phi_v = 0 \tag{10.52}$$

where K_v and Φ_v are respectively the internal energy (thermal + turbulent) and gravitational potential energy of the final virialised object. If the collapsing object has little kinetic energy at turn around (i.e. its peculiar velocity is small), then conservation of energy gives :

$$-\frac{GM^2}{R_m} = K_v - \frac{GM^2}{R_v} = -\frac{1}{2}\frac{GM^2}{R_v} \tag{10.53}$$

or $R_v = \frac{1}{2}R_m$ and the density will be eight times the density at turn around. The object has now fully decoupled from the Hubble flow.

The density of the universe at the formation epoch is given by :

$$\bar{\rho}(z_f) = \left(\frac{1+z_f}{1+z_m}\right)^3 \bar{\rho}(z_m) \tag{10.54}$$

Hence we can write for the virialised density :

$$\rho_v \sim 5.6 \times 8 \times \bar{\rho}(z_m) = (1.59)^3 \times 5.6 \times 8 \times \bar{\rho}(z_f) \tag{10.55}$$

hence :

$$\boxed{\rho_v \sim 200\rho(z_f) \sim 200\rho(z=0)(1+z_f)^3} \tag{10.56}$$

That is to say that the final de-coupled virialised object has a density of 200 times that of the universe at the epoch it forms.

10.3 Application to the Milky Way

As a first application of this result, we can estimate the formation epoch of different objects. Recalling that this result applies to all matter and is therefore dominated by dark matter we can estimate the following :

- The Milky Way :

$$M_{DM} \approx 3 \times 10^{11} M_\odot \tag{10.57}$$

$$R_{DM} \approx 50 \text{ kpc} \tag{10.58}$$

- calculating ρ_{DM} and comparing to $\bar{\rho_0}$ gives z_f ~2.5. Larger, less dense objects must form later, e.g. for a typical cluster on this analysis z_f ~1. Of course the problem is much more complicated than this as we now discuss.

10.4 Hierarchical structure formation

The typical Jeans mass we found is crucial in understanding the process of galaxy formation. The structures which initially form in a CDM cosmology are much smaller than the scale of a typical galaxy which we find now. Indeed more detailed calculations show that since the dark matter decouples at a very early epoch, perturbations on a wide range of scales are present after recombination.

The picture we have is then as follows :

- The initial structures, over-densities, which form are comparatively small by galaxy standards
- These collapse under their own self-gravity to form dark matter halos
- These dark matter halos are of course subject to gravitational interactions and can merge under their mutual gravitational interaction to form larger structures. The timescale for merger will be of order a few crossing times.

This process has been followed via numerical simulations such as the Illustris simulations which we have already referred to. However can we gain some analytical insight into these processes ?

10.4.1 The Press-Schechter Mass Function

The analysis begin with the assumption that the primordial density perturbations were Gaussian fluctuations. Thus, the phases of the waves which make up the density distribution were random and the probability distribution of the amplitudes of the perturbations could be described by a Gaussian function :

$$p(\Delta) = \frac{1}{\sqrt{2\pi}\sigma(M)} \exp\Big[- \frac{\Delta^2}{2\sigma^2(M)}\Big] \tag{10.59}$$

where $\Delta = \delta\rho/\rho$ is the density contrast associated with the perturbations of mass M and :

$$< \Delta^2 >= \Big\langle \left(\frac{\delta\rho}{\rho}\right)^2 \Big\rangle = \sigma^2(M) \tag{10.60}$$

Assume that :

- Perturbations grow according to the linear theory until they reach a critical density contrast Δ_c, they evolved rapidly into bound objects with mass M
- The perturbations had a power-law power-spectrum $P(k) = k^n$
- For the Einstein-de-Sitter model $\Omega_0 = 1$, $\Omega_\Lambda = 1$, so that the perturbations grow as $\Delta \propto R \propto t^{2/3}$

For fluctuations of a given mass M, the fraction $F(M)$ of those which became bound at a particular epoch were those with amplitudes greater than Δ_c :

$$F(M) = \frac{1}{\sqrt{2\pi}\sigma(M)} \int_{\Delta_c}^{\infty} \exp\left[-\frac{\Delta^2}{2\sigma^2(M)}\right] \mathrm{d}\Delta = \frac{1}{2}[1 - \Phi(t_c)] \tag{10.61}$$

where $t_c = \Delta_c/\sqrt{2}\sigma$, and $\Phi(x)$ is the probability integral :

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, \mathrm{d}t \tag{10.62}$$

We can express the mean-squared density contrast in terms of power-spectrum of fluctuations :

$$\sigma^2(M) = \left\langle \left(\frac{\delta\rho}{\rho}\right)^2 \right\rangle =< \Delta^2 >= AM^{-(3+n)/3} \tag{10.63}$$

where A is a constant. We can also express t_c in terms of the mass distribution :

$$t_c = \frac{\Delta_c}{\sqrt{2}\sigma(M)} = \frac{\Delta_c}{\sqrt{2}A^{1/2}} M^{(3+n)/6} = \left(\frac{M}{N^\star}\right)^{(3+n)/6} \tag{10.64}$$

where $M^\star = (2A/\Delta_c^2)^{3/(3+n)}$.
Since the amplitude of the perturbation $\Delta(M)$ grew as $\Delta(M) \propto R \propto t^{2/3}$, it follows that $\sigma^2(M) = \Delta^2(M) \propto t^{4/3}$, therefore $A \propto t^{4/3}$. Hence :

$$M^\star \propto A^{3/(3+n)} \propto t^{4/(3+n)} \tag{10.65}$$

and

$$M^\star = M_0^\star \left(\frac{t}{t_0}\right)^{4/(3+n)} \tag{10.66}$$

where $M_0^\star$ is the value of $M^\star$ at the epoch t_0.
The fraction of perturbation with masses in the range M to $M + dM$ is :

$$\mathrm{d}F = -\frac{\partial F}{\partial M}\mathrm{d}M \tag{10.67}$$

N.B.:The minus sign appearing because F is a decreasing function of increasing M

In the linear regime, the mass of the perturbation is $M = \bar{\varrho}V$, where $\bar{\varrho}$ is the mean density of the universe. Once the perturbation became non-linear, collapse ensued and ultimately a bound object of mass M was formed. The space density per unit mass of perturbations in the mass range M and $M + dM$ that will become bound is:

$$N(M) = \frac{\mathrm{d}n(M)}{\mathrm{d}M} = -\frac{\bar{\rho}}{M}\frac{\partial F}{\partial M} \tag{10.68}$$

Since :

$$\frac{d\Phi}{dx} = \frac{2}{\sqrt{\pi}} e^{-x^2} \tag{10.69}$$

we find :

$$\boxed{N(M) = \frac{1}{2\sqrt{\pi}}\left(1 + \frac{n}{3}\right)\frac{\bar{\rho}}{M^2}\left(\frac{M}{M^\star}\right)^{(3+n)/6} \exp\left[-\left(\frac{M}{M^\star}\right)^{(3+n)/3}\right]} \tag{10.70}$$

This formalism results in only half the total mass density being condensed into bound objects because of the fact that, according to this simple analysis, only the positive density fluctuations developed into bound systems. The underlying cause of this factor of two discrepancy is the fact that the above analysis is based upon the linear theory of the growth of the perturbations. Once the perturbations developed to large amplitude, mass was accreted from the vicinity of the perturbation and N-body simulations show that most of the mass was indeed condensed into discrete structures.
The Press-Schechter mass function can be written as :

$$N(M) = \frac{\bar{\rho}}{\sqrt{\pi}} \frac{\gamma}{M^2} \left(\frac{M}{M^\star} \right)^{\gamma/2} \exp \left[- \left(\frac{M}{M^\star} \right)^{\gamma} \right] \tag{10.71}$$

where $\gamma = 1 + (n/3)$ and $M^\star = M^\star(t_0)(t/t_0)^{4/3\gamma}$

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

In the beginning there were only probabilities. The universe could only come into existence if someone observed it. It does not matter that the observers turned up several billion years later. The universe exists because we are aware of it.

— Martin J. Rees

In this final chapter, we consider the formation of real galaxies and in particular the evolution of the baryonic gas. While gravitation stills plays a central role, gas heating and cooling now become our focus. In particular, we want to understand the way in which gas is processed into stars and hence also planetary systems and how this is linked to the overall cosmological evolution of the Universe.

11.1 The first objects

Following recombination, baryonic gas is neutral. As the Universe expands, a number of things occur :

- The CMB cools proportional to $1/R$ hence is temperature is given by :

$$2.7 \times (1+z)\mathrm{K} \tag{11.1}$$

- The baryonic gas not associated with self-gravitating objects cools adiabatically and faster than this proportional to $1/R^2$ since its adiabatic index is $5/3$
- The process of hierarchical structure formation is occurring all the time with the merger of dark matter halos to form ever larger self-gravitating objects.

At first however, there is no star formation or formation of black holes and AGN: the Universe is dark apart from the CMB radiations, this is **the Dark Ages**. At some point the density of baryonic gas in halos becomes sufficiently large that the first stars and/or AGN form. This is a critical point in the history of the Universe: it marks the end of the dark ages, and the start of the formation of the structure we now observe. The first objects produce UV and/or X-Ray emission which starts to reionise the neutral hydrogen outside of the densest regions. This epoch when the first objects form is therefore called the epoch of reionisation (see chapter 1). Importantly, it is observable via the 21 cm transition of hydrogen redshifted at low frequencies corresponding to this epoch.

One piece of physics we need is the effect on the spin temperature of hydrogen when we have a small flux of UV radiation. At high densities, collisions keep the spin temperature equal to the kinetic temperature of the gas. As the density falls, collisions become less important, and interaction with the photon field determines the spin temperature. Initially this is the CMB, but once the first objects start to form, a small optical/UV flux is present and this acts to excite neutral hydrogen to an excited electronic energy state and as the hydrogen returns to the ground state the probability of returning to either of the split levels of the ground state is determined by the kinetic temperature of the gas. This is the so-called *Wythousen-Field effect* and it results in a resonant coupling between the spin temperature and the kinetic gas temperature again.
Recall that the spin temperature is directly related to the populations of the ground state of hydrogen:

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} \exp\left(-\frac{h\nu_{21}}{k_B T}\right) \tag{11.2}$$

The evolution of the spin temperature (Figure 11.1) can be understood as follows :

- After recombination gas and the CMB cool adiabatically, the kinetic temperature is therefore : $T_K < T_{CMB}$
- Initially, the density is large enough that collisions ensure : $T_S \sim T_K$
- As density falls, CMB determines $T_S \rightarrow T_{CMB}$
- Overdensity in the dark matter collapse forming the first objects
- As first objects form, UV resonant scattering couples : T_S to T_K
- Gas temperature increases and so does T_S due to heating from stars and AGN

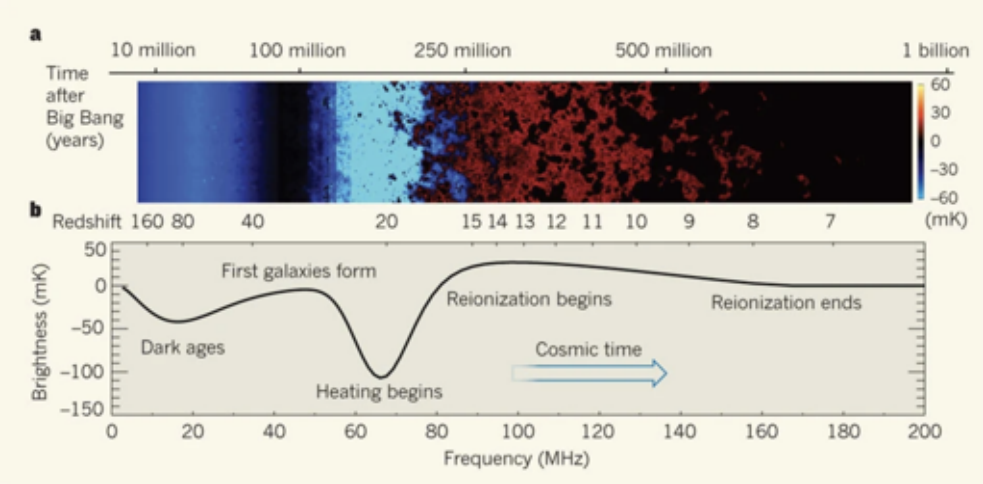


Figure 11.1 : Evolution of the Spin Temperature. From Pritchard & Loeb 2010

- UV and X-Ray flux ionise the gas and $T_S \to 0$

What we observe is the hydrogen seen against the CMB : we are in the Rayleigh-Jeans limit and can therefore think in terms of the equivalent temperature of the radiation. In fact, the observed brightness of the 21 cm line is given by :

$$T_b = 27 x_{HI}(1+\delta_b)\left(\frac{T_s - T_{CMB}}{T_S}\right)\left(\frac{1+z}{10}\right)^{1/2}\left(\frac{\partial_r v_r}{(1+z)H(z)}\right)^{-1} \text{mK} \qquad (11.3)$$

where x_{HI} is the ionisation fraction, δ_B is the baryon overdensity and the last term is the peculiar velocity relative to the Hubble flow.

Of course the formation of the first object does not all occur at once, and is not homogeneous and we expect an evolving structure of reionisation with Cosmic epoch which we will be able to observe with the next generation of radio telescopes currently being constructed.

11.2 Baryonic gas in dark matter halos

The raw material for star formation is the baryonic gas. We have seen how the overdensity in the baryonic gas catches up that in the dark matter distribution :

$$\Delta_B = \Delta_D\left(1 - \frac{z}{z_0}\right) \qquad (11.4)$$

This leads to a simple picture of how we expect the baryonic gas to behave :

- Baryonic gas should fall into pre-existing or growing dark-matter halos
- Typically infall velocities will be of order the free-fall speed ($\sim GM/R$), this greatly exceeds the sound speed in the baryonic gas
- The gas therefore passes through a shock - a structure formation shock - and is heated

11.2.1 Hydrostatic equilibrium

To proceed, we develop a model for the gas and dark matter :

- The matter forms a virially supported structure of radius :

$$R_v = \frac{1}{2} R_m \tag{11.5}$$

  where R_m is the radius at turnaround (see previous chapter).

- The mean matter density in the halo is given by :

$$\rho_v = \Delta_f \bar{\rho}(z_f) = \Delta_f \rho_0 (1+z_f)^3 \tag{11.6}$$

  where ρ_0 is the total matter density in the Universe at the current epoch and $\Delta_f \sim$200 as we showed in the previous chapter.

- Assume further that the matter has a density structure given by the singular isothermal sphere model - the "pressure support" for the dark matter we take to be due to random motions of the non-interacting dark matter particles. For the baryons, we have normal pressure support.
- The baryonic gas is in hydrostatic equilibrium in this potential well

The halo is characterised by its total mass - the virial mass M_v :

- dark matter mass : $M_D = \frac{\Omega_D}{\Omega_M} M_v$
- baryonic mass : $M_B = \frac{\Omega_B}{\Omega_M} M_v$

The virial radius of the halo is given by :

$$R_v = \left(\frac{3M_v}{4\pi\rho_v} \right)^{1/3} = \left(\frac{3M_v}{4\pi\Delta_f\rho_0} \right)^{1/3} (1+z_f)^{-1} \tag{11.7}$$

The equation of hydrostatic equilibrium for the baryonic gas is :

$$\frac{1}{\rho_g} \frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{GM(r)}{r^2} \tag{11.8}$$

Recall that for the singular isothermal sphere :

$$\rho = \frac{a^2}{2\pi G r^2} \tag{11.9}$$

$$M = \frac{2a^2 r_0}{G} \tag{11.10}$$

where the sphere was of mass M, radius r_0 and a is the sound speed.
We can therefore write :

$$\rho(r) = \frac{M_v}{4\pi R_v r^2} = \frac{\rho_v R_v^2}{3r^2} \tag{11.11}$$

and hence :

$$M(r) = M_v \frac{r}{R_v} \tag{11.12}$$

The equation of state of the baryonic gas is $p = \rho_g kT/\mu$ where μ is the mean mass per perticule (i.e. =0.69 for ionised primordial gas). Substituting into eq.11.8, we find that the virial temperature for the gas is given by :

$$kT_v = \frac{GM_v\mu}{2R_v} \propto M_v^{2/3}(1+z) \tag{11.13}$$

$$\text{or} \qquad T_v \approx 3.6 \times 10^5 \left(\frac{M_v}{10^{12} M_\odot}\right)^{2/3} (1+z) \qquad \text{K} \tag{11.14}$$

11.2.2 Halo growth

From our discussions in the previous chapter, we saw that the hierarchical structure formation scenario suggests that halos continue to grow after their initial formation epoch by continually accreting dark-matter and gas via mergers of smaller structures. Further we argued that the Press-Schecheter formalism gave a good description of this process. Our picture therefore is one in which the halos are continuing to evolve throughout their lifetime.

11.2.3 Gas cooling

This gas will cool. In chapter 2, we discussed this in some detail. At the virial temperatures indicated by the above analysis and at lower temperature, the cooling will be dominated by line emission. We are also in the low-density limit therefore the cooling rate will be given by :

$$\epsilon = n^2 \Lambda(T) \approx n^2 \sum_i \frac{g_{u,i}}{g_{l,i}} C_{ul,i} h\nu_i . \exp\left(-h\nu_i/k_B T\right) \tag{11.15}$$

- For a line to be excited and contribute significantly to the sum : $kT > h\nu_i$

- We approximate the exponential terms to approximately unity, therefore we do the sum only over excited lines
- The collisional de-excitation rates (C_{ul}) for the lines which are important for cooling at these temperatures typically have a $T^{-1/2}$ temperature dependence:

$$\epsilon \approx n^2 \Lambda_0 \left(\frac{T}{T_0}\right)^{-1/2} \tag{11.16}$$

The cooling time for the gas depends on radius and is given by :

$$t_c(r) = \frac{\frac{3}{2}\rho_g(r) k_B T_\nu/\mu}{n^2 \Lambda(T)} = \frac{3\mu k_B}{2\Lambda_0 T_0^{1/2}} \frac{T_\nu^{3/2}}{\rho_g} = A\frac{T_\nu^{3/2}}{\rho_g} \tag{11.17}$$

We find the characteristic cooling time for the halo by averaging over all the particles in the halo :

$$\bar{t_c} = \frac{1}{M_v}\int_0^{R_v} t_c(r)\rho_g 4\pi r^2 \,\mathrm{d}r = \frac{A}{M_v}\int_0^{R_v} \frac{T_v^{3/2}}{\rho_g}\rho_g 4\pi r^2 \,\mathrm{d}r \tag{11.18}$$

$$= \frac{A}{M_v} T_v^{3/2} \frac{4}{3}\pi R_v^3 = A\frac{T_v^{3/2}}{\rho_v} \tag{11.19}$$

$$\propto \frac{\left(M_v^{3/2}(1+z_f)\right)^{3/2}}{(1+z_f)^3} \propto M_v (1+z)^{-3/2} \tag{11.20}$$

Inserting numerical values :

$$\frac{\bar{t_c}}{yr} \sim 6.4 \times 10^{10} \left(\frac{M_v}{10^{12} M_\odot}\right)(1+z_f)^{-3/2} \tag{11.21}$$

Cooling is therefore more efficient in low-mass halos and those which form at high-redshift.
If we equate the cooling time to the age of the universe at a given epoch, we find a condition for all the baryonic gas to cool :

- To illustrate the result we take an Einstein-de-Sitter universe for which the age of the Universe at redshift z is given by :

$$t_u = 1.2 \times 10^{10}(1+z_f)^{-3/2} \text{ yr} \tag{11.22}$$

- if t_c is much less than t_u we expect all the gas to cool in this simple model - this predicts a many small-mass halos where all the gas has been converted to stars.

- Note also they have the same redshift dependence therefore at all epochs the low-mass objects cool more efficiently

Since the gas density depends on radius, within each halo there is a radius inside of which the cooling time is less than the Hubble time : this is called the cooling radius. Outside of this radius, we do not expect gas cooling to be efficient. The cooling radius moves out with time.

11.2.4 Star Formation

Star formation happens in the cool gas which accumulates at the center of the potential well. There will be continuous input to the reservoir of cold gas as the warm gas at the virial temperature in the halo cools at a rate :

$$\dot{M}_{c,acc} \approx \int 4\pi r^2 \frac{f_g^2 \rho_g^2 \Lambda(T)}{\frac{3}{2}\mu k_B T_\nu} \, \mathrm{d}r \tag{11.23}$$

where $f_g(r)$ is the fraction of gas remaining in the hot phase at any time at a given radius.
As a consequence of angular momentum conservation, the accreting gas settles into a disc. As we have seen in chapter 5, the gas in the disc becomes gravitationally unstable, and therefore forms stars, when its surface density is:

$$\sigma > \frac{\kappa a_0}{\pi G} \tag{11.24}$$

Once the gas surface density is above this threshold, star formation will occur following the Schmidt-Kennicutt relation.

11.3 The galaxy Luminosity Function and the galaxy populations

We have seen in chapter 6 that the Luminosity Function follow a Schechter function :

$$\frac{\mathrm{d}n}{\mathrm{d}L} = \Phi(L) = \frac{\Phi^\star}{L^\star} \exp(-L/L^\star) \left(\frac{L}{L^\star}\right)^\alpha \tag{11.25}$$

Fits to the data give α ranging from -1.1 to -1.8 (depending on the redshift). In chapter 9, however, we derived the form for the mass spectrum of halos via the Press-Schechter formalism :

$$N(M) = \frac{1}{2\sqrt{\pi}} \left(1 + \frac{n}{3}\right) \frac{\bar{\rho}}{M^2} \left(\frac{M}{M^\star}\right)^{(3+n)/6} \exp\left[-\left(\frac{M}{M^\star}\right)^{(3+n)/3}\right] \tag{11.26}$$

where the index n is the index of the power-spectrum of the fluctuations in the early universe.

We can see straight away that there are inconsistencies : If we assume a constant mass-to-light ratio then to match the observed data we require $n = 3$. However, the Harrison-Zel'dovich spectrum has $n = 1$ and the reprocessed spectrum $n = -3$. In either case, we predict a much steeper spectrum than observed.

In fact the problem is much worse than this. We have seen that cooling is very efficient in the low mass halos. In fact for halos not much less massive than that of the Milky Way, the cooling time is less than the Hubble time. Therefore we predict :

- Low mass halos should have processed all of the gas into stars and have done so at early epochs. This would steepen the luminosity function even further.
- High-mass halos should still be gas rich and still actively forming stars in all cases since the cooling is long.

In fact what we observe is :

- Elliptical galaxies in the local universe with little gas and evolved stellar populations
- many small irregular galaxies which are very gas rich and still forming stars - we do observe dwarf ellipticals but these do not dominate
- we observe many high-redshift galaxies with extremely large star formation rates - these ULIRGs have inferred masses which mean they have to be the precursors of the giant elliptical galaxies in the local Universe

The model we have outlined has problems. But there are also significant successes :

- The shape of the luminosity function
- the formation redshifts for objects of different mass
- the cooling argument suggests a change in behaviour in mass of order the mass of the Milky Way - this is approximately where the break in the luminosity function occurs. The star formation in higher-mass objects is suppressed due to longer cooling times
- Very massive halos form very late and will have little cooling - the model matches the observed clusters extremely well where we have very hot massive systems which can take longer to be a single halo.

11.4 Feedback

The key to solving these problems appears to be the inclusion of feedback. We have already seen how feedback provides significant energy input (mostly from supernovae) into the ISM. Including supernova feedback helps in two ways :

- The energy input heats the gas in the disc suppressing star formation
- The supernovae drive winds which eject cold gas from the disc and put it back into the halo - if these are sufficiently energetic the gas is moved to large radii. If gas is expelled beyond the cooling radius the feedback can halt star formation

A second process is also effective in suppressing star formation in the lowest mass halos and this is reionisation of the gas from photons forming UV background produced by star formation and AGN activity.
More sophisticated models require numerical solution to follow the hot and cold gas. The model in outline is as follows and is called a semi-analytic model :

- The process of halo growth is not taken from analytical models, but from numerical N-body models. In fact the basic buildup of the halos follows well the predictions of Press-Schechter, but the numerical approach includes the sudden increases in virial mass and gas which occurs as halos merge in the hierarchical structure formation scenario
- Halo structure is modelled much as we have described
- Cooling and star formation are modelled as we have described
- Supernova feedback is included by simply heating and ejecting the cold gas into the halo. The supernova rate is determined from the IMF as the fraction of stars with masses above $8M_{\odot}$ and the efficiency is treated as a free parameter
- Reionisation feedback is also included
- As mentioned in Chapter 7, recently feeedback from AGN has also been included as an additional heating source - analytic models of jet-producing AGN are incorporated into the model. The AGN are triggered because of black hole growth at the centres of galaxies. This is currently modelled quite simply. The effects are on the most massive galaxies for which the available energy (proportional to $\epsilon M_{BH}c^2$) is largest.

Another test is to compare to the measured star formation rate density (i.e. average

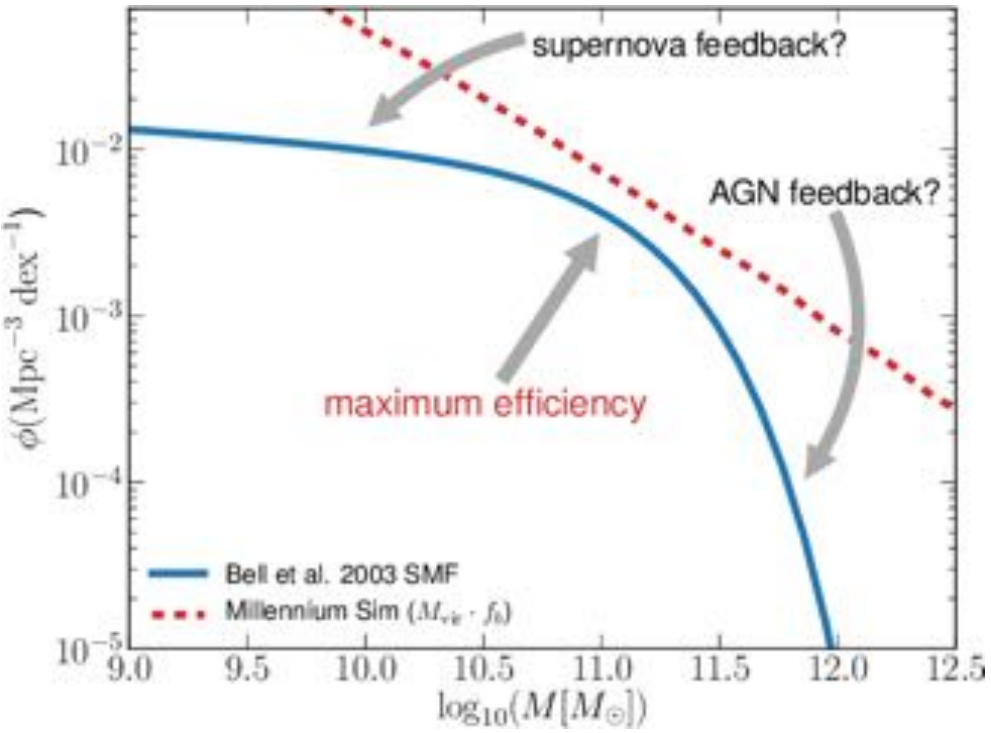


Figure 11.2 : Modelling the galaxy mass function with feedback included. The dashed red line shows the expected shape of the mass function from the dark-halos mass function (assuming constant mass-to-light ratio).From Mutch et al. (2013)

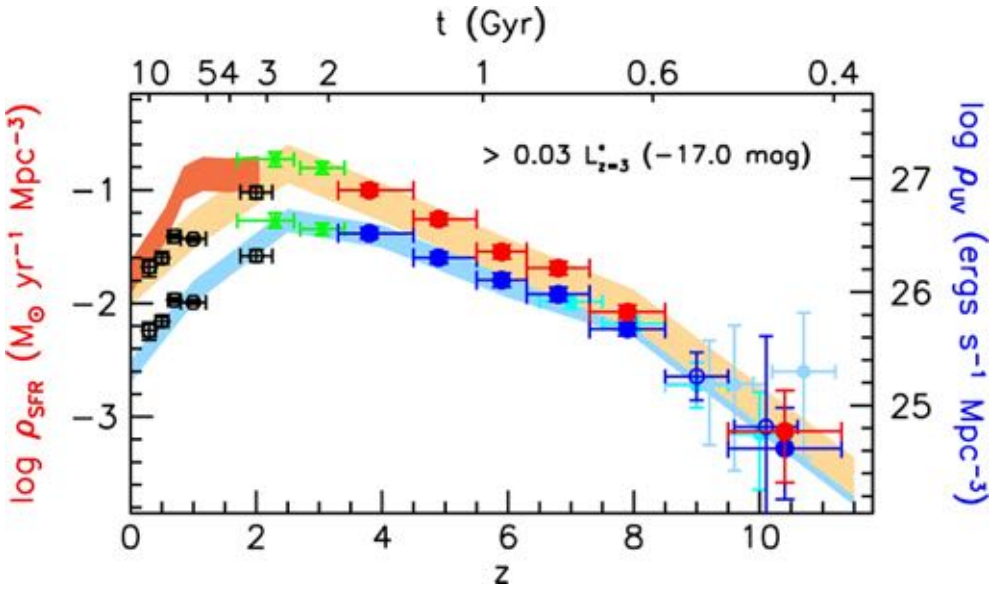


Figure 11.3 : Star Formation Rate density over cosmic time. From Bouwens et al. (2015)

star-formation rate per init volume in the Universe) as a function of redshift. We don't have time to discuss how the experiment is done, but the results are shown in Figure 11.3. The general trend is that the average star formation density increases by about a factor of 10 or 20 between $z = 0$ and $z = 2$. Currently the most accepted scenario is that this increase of the star formation rate density in the universe is primarily due to a much larger amount of gas in high-z galaxies, as a consequence of gas cooling from the halo and from the intergalactic medium ; as a consequence of the Schmidt-Kennicutt relation, the larger amount of gas results into a higher star-formation rate in each galaxy. Including the effect of AGN feedback suppresses star formation still further especially at low-redshift in the massive halos which otherwise now have cooling time comparable to the Hubble flow. Figure 11.4 shows that the inlcusion of the various form of feedback is succesfull in reproducing the mass function of galaxies over a large range of redshift.

11.5 Comments and Conclusion

In this course, we have discussed our current understanding on structure formation in the Universe, mostly based on advances made over the past 10-20 years. The models we have developed have several successes (e.g. the galaxy mass function),

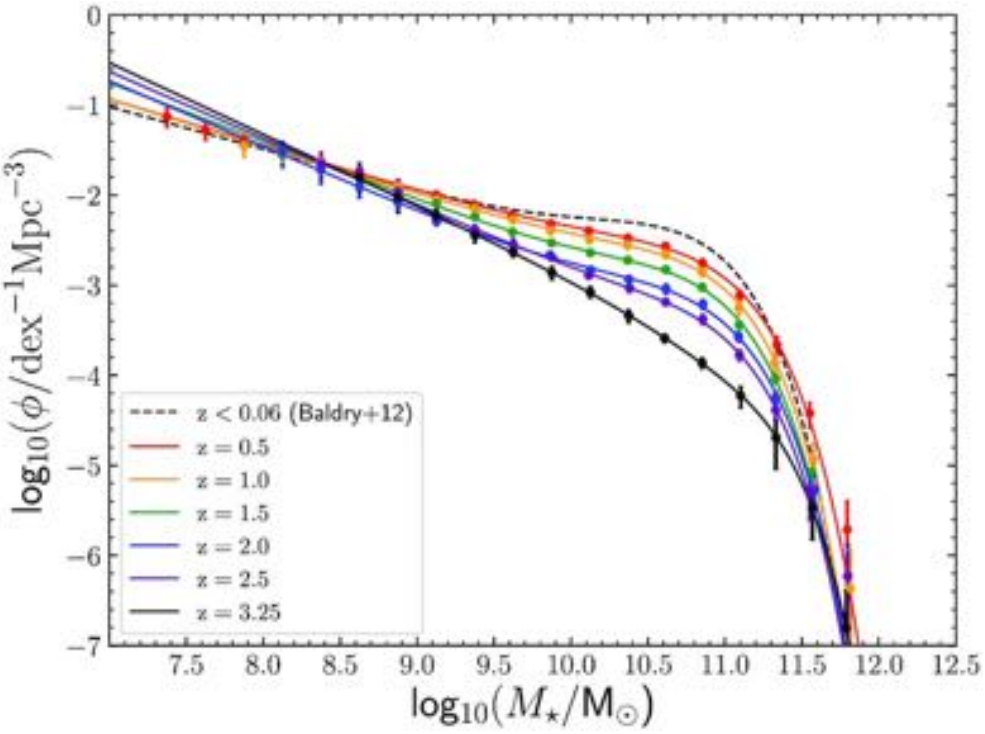


Figure 11.4 : Evolution of the observed Galaxy Stellar Mass Function from $z \sim 0$ to $z \sim$3.25 in *Hubble* surveys. From McLeod et al. (2021)

and leads to the emergence of a consistent picture. However several concerns arise with these models :

- Each time we face a problem raised by observations, we have to include more complex physics to the models to overcome it.
- Our current models can not explain the high-star formation rate objects seen at high-redshift (e.g. $>1000\ M_{\odot}/yr$)
- All these models have free parameters that are still poorly understood (e.g. the escape fraction of galaxies at high-redshift)

However, the future is very exiting with the arrival of several new observatories whose main goals will be to address all the previous points. In December 2021, the Ariane V rocket launched the *James Webb* Space Telescope, a 10 billion dollars project, which will start observing in June this year. Moreover, within the next decade, the European Extremely Large Telescope (ELT), the Square Kilometre Array (SKA) and many other facilities will provide new, detailed information on the properties of galaxies, and will drastically improve our understanding of the mechanisms and physical processes responsible for galaxy formation, evolution and transformation.

Appendices

A Examples 1

A.1

Show that integration of the transfer equation for a constant source function S_ν in a region extanding from $s = 0$ to $s = L$ gives :

$$I_\nu - I_\nu(0) = (S_\nu - I_\nu(0))(1 - e^{-\tau_\nu}) \tag{27}$$

A.2

The ground state of neutral hydrogen is split into two states by the effects of nuclear spin. It is usefull to describe the relative populations of the two states by introducing an *excitation* temperature defined by :

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} e^{-h\nu/k_B T_{ex}} \tag{28}$$

where $h\nu$=$E_2 - E_1$.
A cloud of hydrogen with an excitation temperature T_{ex} is observed against the cosmic microwave background, assuming the excitation temperature of the cloud is the same at all cosmological epochs, describe how observations of the cloud depend on the redshift at which it is observed ?

A.3

A single spherical dust grain of radius a is illuminated by a planar wave of flux F. What temperature will the grain reach in steady state if assuming it is a perfect thermal absorber and emitter ? *You may assume the only heating source for the grain is the radiation field.* A star of luminosity L_ν is surrounded by a cloud of dust in which all the grains are spherical with a radius a and number density of grains n_g. The cloud has an inner radius r_i which is much larger than radius of the star.

1. Assuming perfect thermal absorption and emission for the individual grains, and that the cloud is optically thin, find an expression for the grain temperature at a radius r from the star. *Hint : in steady state the total radiative energy flow through any radius must be constant.*

2. Now assume the more realistic case in which the efficiency of absorption by dust grains in the infrared depends on the frequency as $Q_{abs} \propto \nu^\beta$ (typically β ~1.5-2.0). A good approximation in this case is that a grain only absorbs at optical/UV wavelenghts and emits only in the infrared - explain why this approximation is reasonable. Find how the dust temperatures scales

with distance from the star in this case [assuming that the absorption cross section of dust grains at optical/UV wavelengths can approximated with their geometrical cross section].

A.4

A cloud of pure hydrogen of number density n_H is illuminated by an O-type star which radiates a total of $S_\star$ ionising photons per second. The ionisation cross section σ_i can be assumed to be independent of frequency and α_B is the total recombination coefficient. Typical values are : $\alpha_B(T) = 2 \times 10^{-16} T_e^{-3/4}$ m^{3}s^{-1} ; σ_i=6.8×10$^{-22}m^2$; $n_H = 10^8$m^{-3} and $S_\star = 10^{49}$s^{-1}.

1. A simple model for the effect of the star on the surrounding gas is to assume that within a sphere of radius r_1 (Strömgren sphere) the gas is fully ionised by the available ionising photons from the star producing an HII region or nebula. Outside this sphere the gas is neutral. By considering the equations for ionisation balance within this sphere show that :

$$r_1 = \left(\frac{3}{4\pi} \frac{S_\star}{n_H^2 \alpha_B} \right)^{1/3} \tag{29}$$

2. Estimate the radius of the Strömgren Sphere

3. A more sophisticated model aims to determine the ionisation fraction $X = n_e/n_H$. Show that at any point in the nebula the ionisation fraction satisfies :

$$\frac{X^2}{1-X} = F_S \frac{1}{n_H} \frac{\sigma_i}{\alpha_B} \tag{30}$$

where F_S is the flux of ionising photons. By approximating $F_S \sim S_\star/4\pi r^2$, show that the approximation $X \sim 1$ is justified for the case considered here.

A.5

Show that for a spherically symmetric self-gravitating cloud of gas with a polytropic equation of state of the form $p = K\rho^\Gamma$ that the temperature is related to the gravitational potential Φ_g by :

$$k_B T = \frac{1-\Gamma}{\Gamma} \mu \Phi_g \tag{31}$$

A.6

Starting from the dispersion relation, define the Jeans length and hence determine an expression for the Jeans mass. How does this expression relate to

Bonnor-Ebert mass and the limiting mass determined from considering the limiting stability of an isothermal cloud ?

A.7

Estimate the Jeans mass for the following : a giant molecular cloud complex ; a dark cloud ; a dense core ; the warm neutral medium ; gas towards the centre of a rich cluster.

A.8

Consider the virial equilibrium of an isothermal cloud of gas of mass M_0, radius r_0 and temperature T_0.

1. Assuming that the density within the cloud is constant, show that the gravitational potential energy is given by :

$$\Omega = -\frac{3}{5}\frac{GM_0^2}{r_0} \tag{32}$$

2. Neglecting external pressure, show that a necessary condition for collapse of the cloud is :

$$\frac{r_0}{a_T} > \sqrt{5} t_{ff} \tag{33}$$

where $a_T = \sqrt{kT/\mu}$ is the isothermal sound speed and $t_{ff} \approx \sqrt{r_0^3/GM_0}$ is the free fall time. Interpret this result.

A.9

Show that the gravitational potential energy of the singular isothermal sphere of mass M_0 and radius r_0 is given by :

$$\Omega = -\frac{GM_0^2}{r_0} \tag{34}$$

Verify that the cloud is in virial equilibrium. In lectures, we defined a dimensionless mass $m = p_0^{1/2} G^{3/2} M_0 / a_T^4$ where p_0 is the external pressure and a_T is the isothermal sound speed. Evaluate this dimensionless quantity for the singular isothermal sphere. Use these results to comment on the stability of the singular isothermal sphere.

A.10

A cloud of gas in hydrostatic equilibrium has a density distribution given by

the singular isothermal sphere :

$$\rho(r) = \frac{a_T^2}{2\pi G r^2} \tag{35}$$

where a_T is the isothermal sound speed. The cloud is perturbed and undergoes inside out collapse in which the free-fall radius R_{ff} propagates into the cloud at a velocity a_T. Find an expression for the accretion rate, $\dot{M}$ onto the growing protostar. How does the density vary within the free fall radius ?

A.11

Given the observed brightness distribution of a late-type spiral galaxy what form would the rotation curve be expected to take in the absence of dark matter ?

B Examples 2

B.1

A molecular cloud has formed stars with an Initial Mass Function $\xi(m) = dN/dm = \xi_0 m^{-\theta}$ in the mass range $m_l < m < m_u$ (with $m_l << m_u$), where dN is the number of stars formed in the mass interval m and $m+dm$, and ξ_0 is a constant. Derive the expression for the total mass of stars formed and for the total luminosity emitted by stars, assuming they are all on the Main sequence and that $L \propto M^4$. Discuss the result in the case of a Salpeter IMF, i.e. θ=2.35

B.2

For the potential $\Phi_g = -A/r^\alpha$, find expressions for the frequency and radius of circular orbits as a function of the angular momentum per unit mass l. Find also an expression for the epicyclic frequency. Show that closed orbits are possible in a non-rotating frame when $\alpha = 1$ and $\alpha = -2$. Comment on these cases.

B.3

In the outer regions of two-armed spiral galaxy, we measure a flat rotation curve with rotational speed of 190km/s. The spiral arms are seen to end at a radius of 8.2kpc. Estimate the pattern speed. Useful astronomical units for the pattern speed are km s^{-1} kpc^{-1}

B.4

Estimate the Collisional relaxation time (in years) for : (a) stars in a globular cluster ; (b) stars in a galaxy ; (c) galaxies in a cluster of galaxies.

B.5

It has been observationally found that the star formation rate (SFR) in (late type) galaxies is proportional to their stellar mass : $SFR = K \times M_\star$, where K is a constant (at least within small redshift intervals). Assuming that the initial stellar mass is M_0, that the initial gas mass is negligible, and that galaxies are subject to a constant inflow rate of gas Φ_0, find how the stellar mass and gas fraction evolve as a function of time. The result shows that at late evolutionary times this model fails to describe the evolution of individual galaxies ; discuss why that is the case and how the model can be improved to obtain a more realistic description of galaxy evolution.

B.6

A development of the closed-box model for the evolution of gas and metallicity in a galaxy is to assume that there is infall of gas. One model assumes that the infall rate is linked to the star formation rate and we take the infall rate to be equal to $\alpha\nu\Psi$ where (1-α) is the fraction of gas returned to the ISM by supernovae, and ν is a constant. Show that the metallicity now has the following dependence on the amount of gas g, the yield P and the total initial gas mass M_0:

$$Z = \frac{P}{\nu}\Big(1 - \Big(\frac{g}{M_0}\Big)^{\nu/(1-\nu)}\Big) \tag{36}$$

B.7

A supermassive black hole has been found at $z = 7.5$ (age of the Universe 800 Myr) with a mass of $M = 8 \times 10^8 M_\odot$. Assume that the black hole seed out of which this black hole has evolved formed at $z = 20$ (age of the Universe ∼200 Myr) and assume the extreme case that, since its formation, the black hole has been continuously accreting at the Eddington rate, with a radiative efficiency ϵ=0.1. Find therefore the minimum mass of the black hole's progenitor and discuss the implications of your finding.

B.8

Show that the relaxation time scale for dynamical friction to act is :

$$t_r = \frac{V^3}{4\pi G^2 M \rho \ln \Lambda} \tag{37}$$

Hence show that the number of times two colliding galaxies pass through one another before merging is proportional to V^4

B.9

Compare the integrated feedback from supernovae to that from the winds of supermassive O and B stars. Assume all stars with masses greater than $3M_\odot$ have 1000 km s^{-1} winds and loose a total of $1M_\odot$, that all stars with masses greater than $8M_\odot$ end their lives as a supernova, and the IMF is a Salpeter IMF with a lower mass limit of 0.1 $M_\odot$.

B.10

(a) Show that in co-moving coordinates in the weak-field, low-velocity Euler's equation can be written as :

$$\frac{\partial v}{\partial t} + \frac{1}{R} v \cdot \nabla_\xi v + \frac{\dot{R}}{R} v = -\frac{1}{R\rho} \nabla_\xi P - \frac{1}{R} \nabla_\xi \Phi_g \tag{38}$$

for an appropriate choice of the gravitational potential.

(b) Show further that thus gravitational potential satisfies the Poisson equation for the overdensity.

B.11

Why is dark matter crucial to understand how density fluctuations grow to form galaxy-scale structures ? Starting from the coupled equations for the growth of over densities in the dark matter and baryonic matter, show that when dark-matter dominates over baryonic matter :

$$\Delta_B = \Delta_D \Big(1 - \frac{z}{z_0}\Big) \tag{39}$$

B.12

Estimate the epoch for which a cluster of galaxies may form using the non-linear collapse model developed in the lectures

B.13

If the peculiar velocity of halos remains constant with epoch, estimate how the merger rate of halos varies with cosmic epoch if all of the halos are of equal mass and are in place at some early epoch. Without detailed calculation, discuss how the cosmic star formation rate might be expected to vary with cosmic epoch.