Chapter 1 | *Introduction*

1.1 Background

The main thrust of this thesis is to obtain an understanding of the physics that is important to the formation, evolution and dynamics of galaxies by using large-scale computer simulations. In order to understand galaxy formation it is necessary that we first obtain a basic knowledge of the cosmological framework in which we are working, both in order to understand the physical conditions in which galaxies are thought to form and so that we can assess the impact of galaxies on the surrounding universe. In this chapter we will therefore outline the basic features of the hot Big Bang (BB) model and the physical mechanisms by which the observable structure in the universe today came to be formed. We will see that although it is possible to summarise the properties of the universe at very early times using only a few equations and constants, the evolution of these initial conditions exhibits an inexorable march from simplicity to complexity, until the universe we see today is an immeasurably intricate and complex place.

We begin by sketching some of the notable turning points in the history of the study of the universe, before reviewing the current state of the Hot BB theory. Finally we address in some detail the features of galaxies that are most relevant to this thesis. Notably, the properties of the interstellar medium in galaxies, and the process of star formation.

1.1.1 The Development of the Hot Big Bang Model

The universe is the sum of all matter that exists and the space in which all events occur or could occur. Cosmology, taken as a whole, is the study of the origin and evolution of the universe. The study of cosmology has a long history involving science, philosophy, and religion and is amongst the oldest of humanity's pursuits. We are concerned with *physical cosmology*, the application of the principles of physics to the universe at large. The first application of modern physics to cosmology occurred with Isaac Newton's 1687 publication of *Principia Mathematica* and his use of the theory of gravity to

explain the motions of celestial bodies. One fundamental difference between Newton's cosmology and those preceding it was the application of the Copernican principle (no 'special' observers should be proposed and so bodies both on and off the earth should be subject to the same physical laws). Newton believed that the Universe must be both infinite and static. If the distribution of matter did not extend forever, he realised, then it would inevitably collapse inwards due to its own self-gravity. One major objection to this model of the universe was popularised by Heinrich Wilhelm Olbers in 1824 (although first proposed by Kepler in 1610) and is now known as *Olbers' paradox*

Olbers argued that if we live in an infinite, transparent universe filled with stars, then in any direction one looks in the night sky, one's line of sight will fall on the surface of a star. This implies that the sky should have a uniform brightness, equal in luminosity to the surface of a star. This fact is in obvious disagreement with the fact that the night sky is dark. This argument was so strong that it puzzled scientists for over 75 years. Surprisingly, the first essentially correct solution to this paradox came from the author and poet Edgar Allan Poe in his 1847 poem 'Eureka':

> *"Were the succession of stars endless, then the background of the sky would present us an uniform luminosity, like that displayed by the Galaxy – since there could be absolutely no point, in all that background, at which would not exist a star. The only mode, therefore, in which, under such a state of affairs, we could comprehend the voids which our telescopes find in innumerable directions, would be by supposing the distance of the invisible background so immense that no ray from it has yet been able to reach us at all. That this may be so, who shall venture to deny?"*

Poe argues that since light has a finite speed and the universe is not infinitely old, the light from the most distant sources has not yet reached us. This solution was put forward independently on a firmer scientific basis by Kelvin (1901). These first arguments against the idea of an infinite, static universe marked the beginnings of an investigation into to the properties and dynamics of a constantly evolving universe, governed by the laws of physics.

Modern scientific cosmology is usually considered to have started with Albert Einstein's publication of the theory of general relativity (Einstein (1916)), in which he wrote down the field equations that describe the relationship between space-time and matter-energy. The most fundamental assumption of cosmology from the days of Einstein through to the 1980's was that the universe is both *homogeneous* and *isotropic* on large

scales, this assumption became such a necessary part of virtually all cosmological models that it is known as the cosmological principle (however, see section 1.1.2, for a discussion of inflationary theory, which may provide a physical explanation for the observed isotropy and homogeneity of the universe)

In 1935 A. G. Walker and H. P. Robertson independently found a metric that determines the space-time interval between two events in a dust filled universe that obeys the cosmological principle. These models were first studied in detail by A. Friedmann in 1922 (and independently by G. Lemaitre in 1927) who solved Einstein's field equations for an isotropic, homogeneous universe to obtain what are now known as the Friedmann equations

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{kc^2}{a^2} \tag{1.1}$$

$$\frac{\ddot{a}}{a} = \frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}\,. \tag{1.2}$$

In these equations the matter in the universe is described as a fluid with density ρ and pressure p. The universe has curvature k and expansion factor $a = a(t)$ is normalised such that a at the present epoch is unity. c represents the speed of light and Λ represents a constant arbitrarily added by Einstein to his original formulation of General relativity in order to allow that the Universe is static. H is the Hubble constant that characterises the rate of expansion of the Universe

The Friedmann equations reveal the astonishing fact that there is a connection between the density of the universe and its geometry. Assuming that the cosmological constant is zero then for any given expansion rate there is a corresponding critical density, ρ_c, that will yield $k = 0$, given by

$$\rho_c(t) = \frac{3H(t)^2}{8\pi G}\,. \tag{1.3}$$

The ultimate behaviour of the universe depends on its mean density. If $\rho < \rho_c$, then the universe is unbound and a can continue to increase to infinity. If $\rho > \rho_c$ then the universe is bound and a inevitably collapses to 0. The ratio of the average density to the critical density is an often used quantity so we define

$$\Omega_X = \frac{\rho_X}{\rho_c}\,, \tag{1.4}$$

where ρ_X represents the density of a particular component of the universe (e.g. baryons (b), cold dark matter (c), radiation (r) or dark energy (Λ)). The total density parameter is represented by Ω_0.

Despite this theoretical work on the evolution of an isotropic, homogeneous universe, observational results about the distribution of matter in the universe were still few and far between. In 1916, when Einstein published the theory of general relativity the phrase 'the large-scale distribution of matter' was generally taken to mean the distribution of stars in the Milky Way galaxy (Peebles (1981)). The nature and distribution of 'spiral nebulae' was uncertain, as was whether or not they were associated with the Milky Way or if they were 'island universes' completely decoupled from our galaxy (Kant (1798)) . It was not until the pioneering work of Hubble (1929b) who, following the discovery of a relation between the periodicity of a pulsating Cepheid variable star and its magnitude, was able to use Cepheid variables as sources of known luminosity and discovered that the distance to the nearby galaxy M31 was approximately 275kpc, and therefore external to the Milky Way, which was known to be less than 100kpc across. This marked the first time that any object had been conclusively proven to lie outside of our galaxy.

It was in the same year that Hubble proved conclusively that at least one spiral nebula was of extragalactic origin that he also demonstrated a linear relationship between the redshift of the emission lines in the spectra of other galaxies (Hubble (1929a)), defined as

$$z = \frac{\lambda_{obs} - \lambda_{lab}}{\lambda_{lab}} = \frac{\Delta\lambda}{\lambda}\,, \tag{1.5}$$

where z is the redshift, λ_{obs} is the wavelength at which a feature is observed and λ_{lab} is the wavelength the same feature has in the laboratory. It is natural to interpret redshifts as being due to the Doppler shift of the light from the galaxy due to the motion of the galaxies away from us. Hubble's law is written

$$v = Hd\,, \tag{1.6}$$

where v represents the recessional velocity of other galaxies and d represents their distance from us. H, the constant of proportionality, is the Hubble constant as first introduced in equation 1.1. This discovery indicated that the universe was expanding, providing evidence for the usefulness of the Friedman solutions.

The Hubble law also fits with the concept of homogeneity, it is easy to show that if the Hubble law holds for some observer, A, moving with the Hubble flow it must also hold for all other observers moving with the same flow, this is consistent with the Copernican principle. To demonstrate this consider figure 1.1. Each of the three points represents a different observer. If the observer at position $\mathbf{r}_a$ observes the Hubble law to be true, that

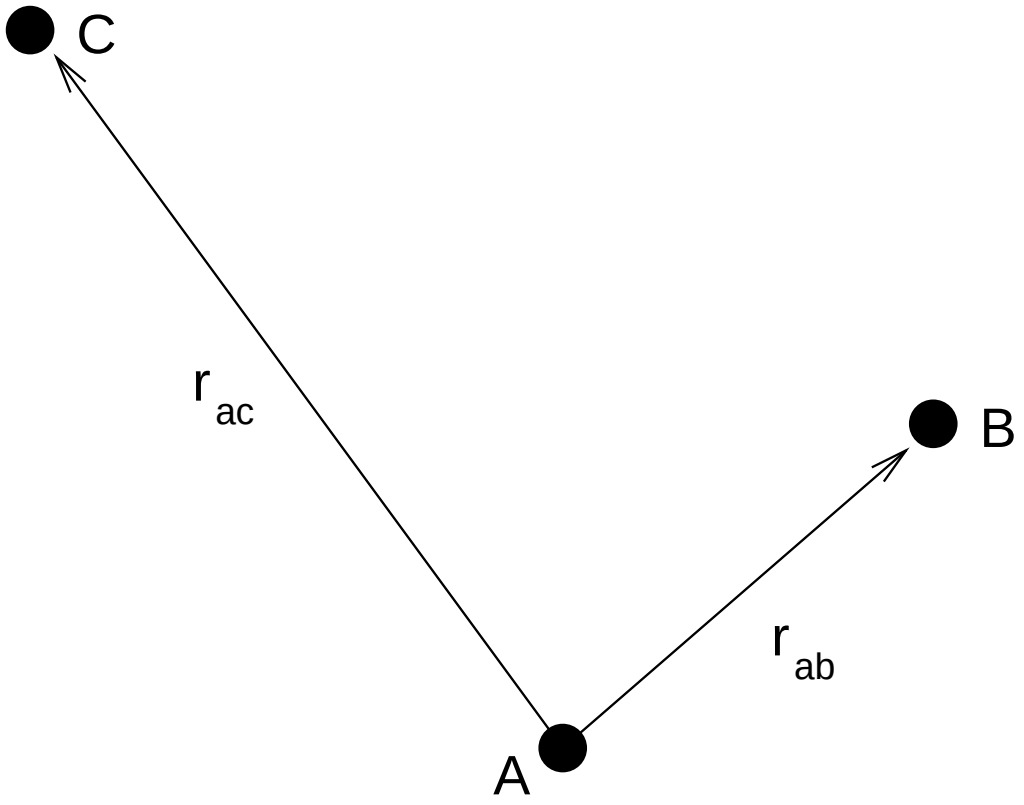

Figure 1.1: Demonstration of the Hubble law. Considering three observers, A, B and C we can show that if the Hubble law holds for any one of the observers then it must necessarily hold for the others.

is

$$\mathbf{v}_{ab} = H\mathbf{r}_{ab}\,, \tag{1.7}$$

and

$$\mathbf{v}_{ac} = H\mathbf{r}_{ac} \tag{1.8}$$

where $\mathbf{v}_{ab}$ is the relative velocity between observer A and observer B. By a simple piece of vector addition we can see that the Hubble law also holds between points b and c,

$$\mathbf{v}_{bc} = H\mathbf{r}_{ab} - H\mathbf{r}_{ac} = H\mathbf{r}_{bc}\,. \tag{1.9}$$

The Hubble law is therefore consistent with the concept of homogeneity.

The ideas of homogeneity and isotropy, when coupled with Einstein's theory of general relativity provide a coherent picture of the universe in which at early times the Universe was small, dense and hot and has expanded and cooled through to the present day. In the following sections we will flesh out the BB paradigm by working chronologically from very early times (10^{-32}s after the BB) to the present day, highlighting some of

the physical mechanisms that operated to form the rich variety of structure we observe today.

1.1.2 The Early Evolution of the Universe

In the following two sections we will detail some of the major features of the evolution of the Universe in a chronological order. The cutoff between the 'early' evolution of the universe and the 'late' evolution is somewhat arbitrary, and has been chosen to correspond to a time shortly after the time at which the cosmic microwave background was formed. This roughly represents the boundary between the parts of the evolution of the universe we can treat analytically (early times) and those that we are forced to treat with numerical tools (late times). In our discussion of the early evolution of the Universe we will consider three different physical processes, each of which represents a fundamental change in the makeup of the Universe. These processes are inflation, nucleosynthesis and decoupling.

Inflation

Despite the great successes of the BB theory in explaining the relative abundances of chemical elements and the existence of the CMB (see the rest of this section), in the early 90's BB theory was plagued by a number of serious problems. In this section we will discuss three of these problems and introduce a period of *inflation* as a possible solution.

The Flatness Problem: It has been known for a long time that the present day mean density of the Universe – parameterised by Ω_0 – is at least to order of magnitude, unity. However, the $\Omega_0 = 1$ universe is unstable. In its very early history, an Ω_0 only very slightly above 1 would have resulted in a very rapid big crunch, while with an Ω_0 only very slightly below 1, the universe would have expanded so fast that stars and galaxies could not have formed. The fact that approximately 13.7 billion years after its formation, the universe still has a value of Ω_0 so close to unity indicates that Ω_0 must have been within one part in 10^{22} (Peacock (1999)) of unity at early times. We can demonstrate this by substituting an expression for Ω_0 into the Friedman equations to obtain

$$\Omega(t)_0 - 1 = \frac{kc^2}{\dot{a}^2}\,. \tag{1.10}$$

Dividing this into the result at the present epoch we obtain

$$\Omega(t)_0 - 1 = \frac{\dot{a_0}^2}{\dot{a}^2}\,. \tag{1.11}$$

Putting in numbers here (Assuming $a \propto t^{1/2}$, i.e. the Universe is dominated by radiation) at this epoch we obtain that $\dot{a}_0^2/\dot{a}^2 < 10^{-48}$. In other words $|\Omega_0 - 1|$ is extremely finely tuned to zero. The problem is that a simple BB theory cannot explain how an Ω_0 so close to unity could arise, the fine tuning of this parameter is indicative of hidden physics.

The Horizon Problem: This problem relates to the uniformity of the CMB (Partridge and Wilkinson (1967); Misner (1968); section 1.1.2). In a BB with only matter and radiation, two widely separated regions (e.g. the regions probed by two patches of the CMB separated by 180 degrees on the sky) will never have been in causal contact (that is, at no point in the history of the Universe will it have been possible for a photon to travel from one region to the other). If two regions cannot have interacted at any point in the past then they should not have been able to smooth out any irregularities in the CMB. This is in contradiction with the observation that the CMB is isotropic and homogeneous to better than one part in 100000 (Smoot et al. (1992))

The Magnetic Monopole Problem: Grand unified theories (GUT) predict the existence of topological defects and massive relic particles, notably magnetic monopoles ('t Hooft (1974)). These particles are expected to form at around the GUT scale ($t \sim 10^{-35}$s). If no more than one monopole forms per horizon scale then the mean density of the Universe would be so high that it would recollapse within $\sim 10^5$ years (Narlikar and Padmanabhan (1991)). Inflation was suggested as a solution to these problems by Guth (1981). According to Guth the BB picture is essentially correct, but with the addition of an *inflationary phase* during which the Universe expands exponentially (as in the classical de Sitter (1917) model) for a short period of time.

The theory of inflation posits that the universe underwent a period of exponential growth ($a \propto e^{Ht}$) within the first 10^{-32}s. This exponential expansion was theorised to occur due to the presence of a quantum scalar field in the early universe (the *inflaton*). In the original models of Guth this scalar field, ϕ, has a potential, $V(\phi, T)$, that depends upon temperature such that at $T \gg T_c$ (where T_c is some critical temperature), the field has one minimum at $\phi = 0$. As the temperature of the Universe decreases the shape of the potential changes until at $T = T_c$ the potential has two minima, and at $T \ll T_c$ the minimum potential is actually at some finite value of ϕ. When this happens, the system is in a state of 'false vacuum', and over the course of time, thermal fluctuations and quantum tunnelling will move the scalar field from $\phi = 0$ to the new, lower potential state, For the time during which the universe is in the 'false vacuum' state, the potential difference between the true vacuum and the false vacuum is the dominant energy

density in the universe and it behaves like a de Sitter universe.

A thorough calculation reveals that Guth's original model does not work (Guth (1983b)), since if the false vacuum is stable for long enough to inflate the universe, the phase transition will proceed in a patchy manner and lead to an inhomogeneous universe incompatible with observations.

Many improvements on the standard model of inflation have been suggested, from changes in the form of the potential (e.g. Linde (1982a,b)), to chaotic inflation whereby the initial value of the potential is slightly different from zero and it slowly 'rolls' to the zero potential (Linde (1983)).

A period of inflation can solve each of the three problems presented so far. The horizon problem is solved because regions outside of each others horizon at the time of CMB emission ($z \sim 1000$) are in contact during the inflationary, de Sitter period and so can thermalize. A period of inflation also has the effect of smoothing out any regions of curvature, forcing that $\Omega_0 \approx 1$.

The temperatures at which magnetic monopoles and other exotic relics would be expected to be produced are greater than $T \sim 10^{27}$K ($t \sim 10^{-34}$s), inflation occurs after this time and is thought to double the Universe in size roughly 60 times (Peacock (1999)) from a size of $\sim 10^{-25}$m to around $\sim 10^{-7}$m, diluting the density of relic particles so that we would never expect to observe them.

Probably the most compelling feature of inflation theory is that it provides a natural mechanism for generating tiny fluctuations in the initial density field due to quantum thermal fluctuations present in any de Sitter universe. The form of these fluctuations was first calculated by four groups working independently (Hawking (1982); Bardeen et al. (1983); Starobinsky (1982); Guth (1983a)), who predicted that the spectrum of perturbations laid down by inflation would be that of a *nearly scale-invariant Gaussian random field* (see section 2.4). The properties of Gaussian random fields are discussed in great detail by Bardeen et al. (1986); a Gaussian random field is defined as any random field for which each value is drawn independently from a Gaussian probability density function. The properties of a Gaussian random field are specified entirely by its power spectrum and in the case of cosmological initial conditions the field is thought to have the form

$$P(k) = Ak^{n_s} \,. \tag{1.12}$$

This prediction has been borne out by the WMAP Cosmic Microwave Background experiment (Spergel et al. (2003, 2006)), which by probing the statistical distribution of density

perturbations at the time of decoupling found that the spectral index, n_s, which is 1 for a scale invariant field, and is predicted to be between 0.92 and 0.98 by simple models of inflation (Steinhardt (2004)), was found to be 0.95. Additionally galaxy surveys (e.g. York et al. (2000)) show galaxy clustering properties consistent with a spectrum of perturbations of this form. These experiments are important confirmations of the theory of inflation and show that the form of the initial density perturbations in the Universe is very close to that predicted by inflation. Other methods of fluctuation generation, for example cosmic strings (Albrecht and Stebbins (1992)) and global textures (Park et al. (1991)) do not make such predictions for the form of the initial density perturbations.

It is worth noting, however, that currently inflation theory is understood principally in terms of its predictions for the initial conditions of the hot, early universe. Despite having passed many observational tests, the physics in inflation theory is still largely *ad hoc* modelling and as such there are open questions about the theory, particularly with regards to the hypothesised inflaton field, which does not currently fit into the standard model of particle physics

Big Bang Nucleosynthesis

The mechanism of stellar nucleosysnthesis (the fusing of primordial hydrogen into heavier elements) was first discussed by Salpeter (1952) and Burbidge et al. (1957), who found that if stellar nucleosynthesis were the only mechanism by which elements heavier than hydrogen could be formed then we vastly under-predict the abundance of Helium, which makes up over 20% of the observable Universe.

The hot, dense universe predicted by the BB model at early times provides an ideal mechanism for generation of the primordial mix of heavy elements we observe. This process is reviewed thoroughly by Boesgaard and Steigman (1985) and sketched only in the barest detail here.

At a temperature of just below 10^{12}K, the universe contained a mixture of photons, electron-positron pairs, and neutrinos. There was also a smaller number of protons and neutrons that were constantly transformed into each other[1]. As long as the rate of the reactions that convert protons and neutrons between each other remains high enough, the neutron-proton ratio will be described by the Boltzmann law. At 10^{12}K this ratio is 0.985. As the universe expands the reaction rates drop rapidly and the neutron-proton ratio becomes 'frozen' at the value it has when $T \approx 10^{10}$K, which was $n_n/n_p = 0.223$.

[1] the following are the important reactions $n \rightleftharpoons p^+ + e^- + \bar{\nu}_e$; $n + e^+ \rightleftharpoons p^+ + \bar{\nu}_e$; $p^+ + e^- \rightleftharpoons n + \nu_e$

At this point the universe is still too hot for heavier nuclei to form and the free neutrons continue to decay via the process

$$n \rightleftharpoons p^+ + e^- + \bar{\nu}_e \,. \tag{1.13}$$

This reaction has a half-life of 617s and continues to dominate until the temperature of the universe drops as far as 10^9K. This took approximately 229s, and the neutron-proton ratio becomes 0.213. Below 10^9K protons and neutrons readily combine to form deuterium and then helium via the following reaction chains

$$^2\mathrm{H} + {}^2\mathrm{H} \rightleftharpoons {}^3\mathrm{H} + {}^1\mathrm{H}\,, \tag{1.14}$$

$$^3\mathrm{H} + {}^2\mathrm{H} \rightleftharpoons {}^4\mathrm{He} + n\,, \tag{1.15}$$

and

$$^2\mathrm{H} + {}^2\mathrm{H} \rightleftharpoons {}^3\mathrm{He} + n\,, \tag{1.16}$$

$$^3\mathrm{He} + {}^2\mathrm{H} \rightleftharpoons {}^4\mathrm{He} + {}^1\mathrm{H}\,. \tag{1.17}$$

The buildup of large amounts of heavier elements is blocked by the fact that there are no stable elements with atomic numbers 5 or 8 at these temperatures and so only trace amounts of heavier elements (e.g. $^7\mathrm{Li}$ and $^7\mathrm{Be}$) are formed (Dicus and Teplitz (1980)). Detailed calculations of the abundances of the heavier elements were carried out by Yang et al. (1984)

Detailed calculations of the predicted primordial abundances of the elements generated during this period provide a very strong test of BB theory. Although complicated by stellar nucleosynthesis the fraction of $^4\mathrm{He}$ observed in the universe agree incredibly well with the BB predictions (see e.g. Izotov et al. (1994)). The abundances of other elements (e.g. Li) are not in such good agreement with the predictions of BBN, but it is unclear whether the origin of this discrepancy is observational, stellar, nuclear or more fundamental (Coc and Vangioni (2005)). The fact that the helium abundance in the universe never drops below 23%, no matter where it is observed is an important confirmation of the theory.

The Cosmic Microwave Background

At the present epoch, we observe a homogeneous background of electromagnetic radiation, well described by a black body with a temperature of 2.73K, which gives it a peak wavelength in the microwave section of the electromagnetic spectrum. The existence of

this cosmic microwave background (CMB) was first predicted qualitatively by Alpher et al. (1948), and observed by Penzias and Wilson (1965).

After nucleosynthesis took place and the primordial mixture of hydrogen, helium and traces of other elements were formed, the universe was made up of a hot plasma of photons, electrons and baryons. Due to the adiabatic expansion of the universe the temperature of this plasma was decreasing. When the temperature of the universe fell to approximately 3000K (redshift of ~ 1100), it became favourable for protons and electrons to combine into neutral hydrogen in a process called *recombination* (Peebles (1968)).

Before this occurrence the photons are kept in thermal equilibrium with the baryons due to the large cross section for the interaction between photons and electrons (Thomson scattering). When recombination takes places this cross section drops to virtually zero and the matter and radiation fields begin to evolve independently (*decoupling* of the photons and baryons). At the moment of decoupling the spectrum of the photons is well described by a black body at a temperature of $\sim 3000K$, by the present day the radiation has been redshifted by a factor a_{dec}/a_0 and has a temperature of approximately 2.73K.

Due to the initial density perturbations put in place by a period of inflation, the CMB is not perfectly smooth, there are small fluctuations in its temperature, these fluctuations are of order one part in 100,000 (Smoot et al. (1992)). These fluctuations arise due to a large variety of physical mechanisms, some of which will be summarised here. *Primary* fluctuations are those that arise due to physical effect before or at the epoch of decoupling. *Secondary* anisotropies arise due to scattering along the line of sight. It is important to note that an accurate calculation of the effect of these perturbations requires a full numerical treatment of the Boltzmann equation (see e.g. Seljak and Zaldarriaga (1996)), so we will discuss them only qualitatively. Primary anisotropies include

- Sachs-Wolfe perturbations. Photons from high density regions at decoupling have to climb out of potential wells and so are redshifted
- Adiabatic perturbations. In high density regions, both the radiation and matter will be compressed and have a higher temperature

These effects are important for perturbations for which only gravity has been important up until the time of recombination, that is: those with wavelengths larger than the horizon size at decoupling. On smaller scales a variety of additional physical processes act on the perturbations.

Radiation pressure tends to erase anisotropies, whereas the gravitational force acts to

make baryons collapse in overdense regions. These two competing effects create acoustic oscillations, which give the CMB its characteristic peak structure. The peaks correspond, roughly, to resonances occurring during recombination when a particular mode is at its peak amplitude. The location of the first acoustic peak in the CMB spectrum depends sensitively upon the total matter density in the universe and thus its location is of crucial importance. Many experiments including Toco (Miller et al. (1999)), Boomerang (Mauskopf et al. (2000)) and Maxima-1 (Hanany et al. (2000)) were designed primarily to determine its location. In the ΛCDM model, the peak shows up right where it would be expected if the curvature of the universe was zero (i.e. $k = 0$).

Secondary anisotropies, which arise due to the interaction between the CMB photons and the intervening medium include:

- The integrated Sachs-Wolfe effect. Photons travel through evolving potential wells and are blueshifted as they fall in to the potential, then redshifted again as they climb back out. If the potential deepens as the photons are travelling through it there will be a net redshift of photons

- The Sunyaev-Zel'dovich effect (see e.g. Sunyaev and Zeldovich (1980)). Photons that interact with high energy electrons (for example those present in the hot diffuse gas of a galaxy cluster), may undergo inverse Compton scattering, in which some of the energy of the electrons is transferred to the photons, leading to an apparent upward shift in CMB temperatures in the directions of galaxy clusters. Some galaxy clusters have been observed via this effect (Birkinshaw et al. (1984))

Other secondary effects including gravitational lensing of the CMB photons can introduce non-gaussianity into the CMB spectrum, this must be carefully controlled for.

See Hu and Dodelson (2002) for a comprehensive recent review of the different mechanisms by which anisotropies are embedded in the CMB.

1.1.3 The Late Stage Evolution of the Universe

After the cosmic background radiation was decoupled from the baryonic component of the universe its subsequent evolution is governed primarily by the effects of gravity, which drives the collapse of objects on all scales, from the largest superclusters down to individual stars. In this section we investigate the process by which galaxies are formed.

Large Scale Structure Formation

Gravity causes the density perturbations laid down by inflation to amplify themselves; overdense regions tend to become more overdense and underdense regions tend to become more underdense. Over time this leads to the collapse of virialised 'haloes' of matter. These haloes continue to interact with each other gravitationally; merging hierarchically to form ever larger structures.

The formation of virialised haloes is a non-linear problem and as such we must resort to numerical tools to investigate it in detail. In section 2.2 we introduce the most widely used numerical tools, and in section 2.4 discuss the generation of the initial conditions that go into one of these codes.

The dynamics of gas in these virialised haloes is a complex and dynamic process. In this section we sketch an outline of how this dense gas is thought to form the wide variety of galaxies we see in the present day universe. In section 1.2 we will discuss the properties of the interstellar medium in present day galaxies, and in chapter 3 will investigate how this system can be simulated.

According to many recent galaxy formation models (Somerville and Primack (1999); Cole et al. (2000); Bower et al. (2006)) diffuse gas that falls into a halo is shock heated to the virial temperature of that halo. When the virial temperature of the halo (White and Rees (1978)) is high enough the gas can cool radiatively (Rees and Ostriker (1977)) and may collapse into the halo. Haloes gain angular momentum from tidal torques operating during their formation (see e.g. Mo et al. (1998)) and so the gas is accreted into the form of a disk. Dense, cold gas in the disk of a galaxy will collapse into stars (see section 1.3.1 for an overview). Various feedback processes from stars regulate the star formation rate in the galaxy (Dekel and Silk (1986)).

The dynamics and evolution of the gas in galaxies is very complex. Stars are thought to form in molecular clouds in a complex interstellar medium (ISM) in which magnetic fields (Safier et al. (1997)), cosmic rays, turbulence (Krumholz and McKee (2005)), relativistic jets (Klamer et al. (2004)), molecules, dust, and radiative transfer may all play some role. Additionally mergers between galaxies can move the gas in the galaxy far from equilibrium.

Reionization

The formation of the first stars and galaxies marks a transition between the smooth universe we have examined at high redshift and the clumpy universe we see at low redshift. It is currently thought that this occurs at redshifts ~ 50 (Reed et al. (2005b); Gao et al. (2005)).

After recombination (see 1.1.2) the hydrogen in the universe is neutral and thus absorbs photons blueward of the Lyman α line very strongly (Gunn and Peterson (1965))[2]. Neutral hydrogen therefore has the effect of efficiently absorbing all UV photons. However, modelling of the transmitted light from observed quasars implies that at least as far back as $z \leq 5.8$ the neutral fraction of hydrogen was less that 10^{-6} (Fan et al. (2000)). We therefore require some mechanism by which the neutral hydrogen that exists from $z \sim 1000$ down to $z \sim 6 - 15$ becomes ionized.

It is thought that this process occurs via the escape of ionizing photons ($h\nu \geq 13.6$eV) from the sites of star formation into the surrounding IGM. In this way each star forming object will ionize a 'bubble'of hydrogen. As the number density of star forming objects increases, ionized bubbles begin to overlap until the intergalactic medium is ionized and reionization is said to be complete.

The redshift at which reionization takes place is still uncertain, additionally reionization is a patchy process, and proceeds by stages (Gnedin (2000)). Studies of quasar spectra have placed the redshift of reionization as low as 6 (Becker et al. (2001)), the WMAP CMB experiment estimated that the redshift of reionization as $\sim 11 - 17$ (Spergel et al. (2003, 2006)), and numerical studies of reionization in ΛCDM simulations place this epoch anywhere in the range $z \sim 7 - 12$ (Haiman and Loeb (1998); Gnedin and Ostriker (1997); Gnedin (2000)).

The end of reionization marks the time at which the universe takes on the characteristics it has at a redshift of zero, with star forming galaxies merging in a hierarchical manner embedded in a highly ionized intergalactic medium.

Many open questions remain about the nature of the sources that reionized the universe. Current best estimates indicate reionization occured at $z \sim 12$ it is known that the observed number counts of high redshift galaxies and quasars could not produce enough ionizing flux to successfully reionize the universe. Other sources therefore become necessary. Some candidate objects have included population III stars and micro-quasars.

[2]This period has become known as the 'cosmic dark ages' a term originally coined by Sir. Martin Rees

21-cm hydrogren emission (Madau et al. (1997)) provides an ideal tool for probing the epoch of reionization and some of the new generation of radio telescopes (e.g. LOFAR (Zaroubi and Silk (2005))) are targeted at discriminating between these different sources and so explaining the process of reionization.

The Existence of 'Dark' Matter

Over the past century a variety of observations have suggested very strongly that a large fraction of the Universe must be in the form of 'dark' matter. That is: matter that does not emit or scatter enough electromagnetic radiation to be directly detectable.

The first evidence for a large fraction of matter in the universe being dark came from two directions almost simultaneously. On large scales it was discovered that when the virial theorem is applied to galaxy clusters (Zwicky (1933)) the observable mass in the galaxy cluster could not account for the random motions of the galaxies. On smaller scales it was found that in our own galaxy the numbers of visible stars fell short by 30-50% of adding up to the amount of matter implied by their velocities. Over the next decades evidence for a large fraction of the matter in the Universe being dark began to stack up. Rotation curves in galaxies were discovered to be flat out to large radii (Ostriker et al. (1974)), and gravitational lensing implies that there is much more mass in galaxy clusters than can be observed directly (Walsh et al. (1979)).

It is still unclear exactly what dark matter is made of. Many candidates have been suggested, these may be broadly classified into baryonic and non-baryonic candidates. Baryonic candidates are mainly classified under the umbrella term MACHOs (massive compact halo objects), a general name for any kind of astronomical body that emits little or no radiation and is present in interstellar space. Candidates for MACHOs have included black holes, neutron stars, brown dwarfs and very faint red dwarf stars. Studies of BBN have indicated that baryons do not make up enough of the universe to explain all of the missing mass. Non-baryonic candidates may be classified as 'cold','warm' or 'hot', differing from each other by the mean velocity the candidate particles have at high redshift. At least one hot dark matter candidate is known, the neutrino,although this particle is not massive enough to account for the missing mass in the universe (Goobar et al. (2006)). Additionally, hot dark matter candidates do not allow gravitational collapse on small scales due to their free-streaming motion out of small potential wells. For this reason at least some of the dark matter content of the universe is thought to be in the form of cold dark matter.

Evidence for dark matter is so strong that it has become an integral part of the modern cosmological paradigm, and has been found necessary in order for large structures to have formed by gravitational collapse over the lifetime of the universe.

The Modern Cosmological Model

Starting from the assumptions of homogeneity and isotropy and applying Einstein's theory of general relativity the BB theory provides a coherent framework within which we can make progress towards understanding the formation and evolution of galaxies.

Shortly after the BB we have seen how a period of inflation, which imprints tiny density fluctuations into the initial density field, solves many problems with the basic BB picture and provides a natural mechanism by which the seeds of the galaxies we see today can be formed. When the temperature of the universe drops enough that it is neutral then we see that the primordial radiation field is decoupled from the matter field and subsequently evolves independently. We observe this relic radiation at the present day at microwave wavelengths and it provides a strong probe of the properties of the universe.

At later times, the perturbations laid down by inflation begin to collapse gravitationally, dragging gas into their potential wells. This gas cools radiatively and some fraction of it collapses into stars. Stellar feedback processes make the resulting structure of galaxies and the ISM very complex, and very many different feedback processes operate in bringing about a stable equilibrium.

In addition to the CMB, many other observations help in constraining the properties of the Universe. These include large scale galaxy surveys (e.g. Colless et al. (2001); York et al. (2000)), which probe the clustering properties of the galaxies at low redshift and hence give us a second, independent probe of the properties of the primordial density field.

The 'Hubble key project' is an attempt to pin down the Hubble constant, H_0, through the use of Cepheid variable stars in nearby galaxies as standard candles in order to make accurate distance determinations for a sample of nearby (≤ 20Mpc) galaxies (Freedman et al. (2001)). H_0 may then be probed by applying the Cepheid calibration to several secondary distance indicators (e.g. Type Ia supernovae and the Tully-Fisher relation) operating further out in the Hubble flow.

Additionally supernovae used as standard candles can tell us about the expansion properties of the universe (Branch (1998); Knop et al. (2003))

Parameter	WMAP Only	WMAP + 2dFGRS
$100\Omega_b h^2$	$2.233^{+0.072}_{-0.091}$	$2.223^{+0.066}_{-0.083}$
$\Omega_m h^2$	$0.1268^{+0.0072}_{-0.0095}$	$0.1262^{+0.0045}_{-0.0062}$
h	$0.734^{+0.028}_{-0.038}$	$0.732^{+0.018}_{-0.025}$
τ	$0.088^{+0.028}_{-0.034}$	$0.083^{+0.027}_{-0.031}$
n_s	$0.951^{+0.015}_{-0.019}$	$0.948^{+0.014}_{-0.018}$
σ_8	$0.744^{+0.050}_{-0.060}$	$0.737^{+0.033}_{-0.045}$

Table 1.1: Cosmological Parameters as derived from the WMAP 3 year results. The middle column shows the results obtained from the WMAP data alone, the right hand column shows the effect of adding the analysis of the 2dFGRS. Data adapted from Spergel et al (2006). Ω_b and Ω_m were defined by equation 1.4, h is the Hubble constant divided by 100km/s/Mpc, τ is the optical depth to the surface of last scattering, n_s is the slope of the initial density perturbation power law (equation 1.12) and σ_8 is the RMS density fluctuation in spheres of 8Mpc.

These many independent observations paint a coherent picture of a universe born in a BB, and well described by the Friedman equations. The most recent experiments to obtain the values of the constants that describe the universe include the WMAP probe (Spergel et al. (2003, 2006))

Assuming a power-law flat ΛCDM model the WMAP experiment along with other experiments as detailed above places tight constraints on the values of the cosmological parameters. Table 1.1 summarises the current WMAP estimates of these numbers, both from data from the WMAP experiment alone, and from a combination of the WMAP and 2dFGRS survey (Colless et al. (2001)), which reduces the error bars on the measurements.

Combining the WMAP results with supernova catalogues allows us to estimate that the Universe is dominated by dark energy. In combination with the Hubble key project (Mould et al. (2000)) the value of the Hubble constant can be measured with great precision.

Now that we have sketched out the primary features of a BB cosmology, the remainder of this introductory chapter will be concerned with a more detailed treatment of the properties of individual galaxies.

1.2 The Interstellar Medium

Thus far we have sketched the main features of the evolving universe from when it was a fraction of a second old down to the present day. We have not, however, looked in much detail at the properties of the gas *internal* to each galaxy. In this section we describe the properties of the interstellar medium (ISM), and the stellar content of galaxies.

1.2.1 The Components of the Interstellar Medium

According to the models of McKee and Ostriker (1977) (hereafter MO77; see also Efstathiou (2000); Monaco (2004); Krumholz and McKee (2005)) the ISM of the MW consists of at least three separate and distinct gas phases: a hot tenuous medium at a temperature of $\sim 10^6 K$; cold, dense molecular clouds at a temperature of $\leq 100K$ and a warm medium that exists at the boundaries between clouds and the hot medium with $T \sim 10^4 K$. In the MW, the hot medium has a filling factor of 70–80% and the cold clouds account for a few percent of the volume (MO77). Different techniques are used to observe these different phases, with radio observations probing roto-vibrational transitions of molecules (mainly CO), the 21-cm line probing atomic hydrogen, and UV- and X-ray observations probing the hot phase, see e.g. Binney and Merrifield (1998) for an overview and further references. The fact that different techniques are used to observe the different gas phases might exaggerate the degree to which these phases are really distinct.

Observations of star formation in the MW show that most form in groups, either as gravitationally bound clusters or unbound associations, in the most massive of the molecular clouds (Giant Molecular Clouds, hereafter GMCs), with masses $\sim 10^6 M_\odot$, and sizes of order 30-50pc, see e.g. Blitz and Thaddeus (1980) and Lada and Lada (2003) for recent reviews. The relatively small sizes of GMCs currently limit detailed observations of such clouds to nearby galaxies, with recent surveys done in the MW (Solomon et al. (1987); Heyer et al. (2001)), M33 (Rosolowsky and Blitz (2004)) and the LMC (Fukui et al. (2001)).

Blitz et al. (2006) present detailed observations of GMCs in five local galaxies. GMCs are excellent tracers of spiral structure to the extent that they are often used to *define* the location of arms, in a similar way as for example HII regions or massive stars are. There is a good correlation between GMCs and filaments of HI, although the reverse is not true, suggesting that clouds form from HI.

Observations in more distant galaxies are currently limited to measuring the mean surface density of molecular gas, with more detailed observations awaiting new instruments such as ALMA[3]. According to Young and Scoville (1991), the fraction of gas in the molecular phase depends on Hubble type, with early-type spirals tending to be predominantly molecular, and late-types atomic . Optically barred spirals show a clear enhancement of CO emission along the bar. This suggests that the large-scale structure of the galaxy affects the formation of clouds, and hence presumably also the star formation rate. A physical implementation of star formation should therefore aim to track the formation and evolution molecular clouds. A recent review of the status of three phase models of the ISM is presented by Cox (2005)

1.2.2 The Atomic Component of the Interstellar Medium

The material that makes up the majority of the volume of the ISM is a diffuse, atomic gas. Typical densities and temperatures for this phase of matter are approximately $n = 10^{-2.5}$ atoms per cubic centimetre and $T = 10^{5.7}$K (MO77). This diffuse gas phase is formed by the heating action of supernovae, and it is this phase which most supernova remnants expand into and so controls the dynamics of SN remnant expansion (for a more complete discussion on this point see section 3.2.6).

One of the major physical processes that controls the thermal properties of the atomic component of the ISM and hence the phase structure of the ISM is radiative cooling (see e.g. Black (1981); Sutherland and Dopita (1993)). The first studies in this area (Cox and Tucker (1969); Tucker and Gould (1966)) concentrated on simple, optically thin plasmas. Since then both our understanding of atomic processes and the available computing power have increased and allowed a full treatment of collisional ionization rates and other atomic processes. Sutherland and Dopita (1993) performed calculations of the radiative cooling rates of a hot astrophysical plasma in collisional ionization equilibrium. Photoionization (Seaton (1958)), electron collision ionization, charge transfer reactions and recombination radiation (Gould (1978)) were all treated in a self consistent manner for a set of 225 species of ion. The resulting cooling rates are shown in figure 1.2. Sutherland and Dopita (1993) noted that non-equilibrium effects can appreciably change the shape of the astrophysical cooling function, but only analysed these effects for a few, idealised, non-equilibrium situations. The cooling function of a primordial gas is of vital importance to the formation of galaxies as it is this cooling that determines the fraction

[3]http://www.alma.nrao.edu/

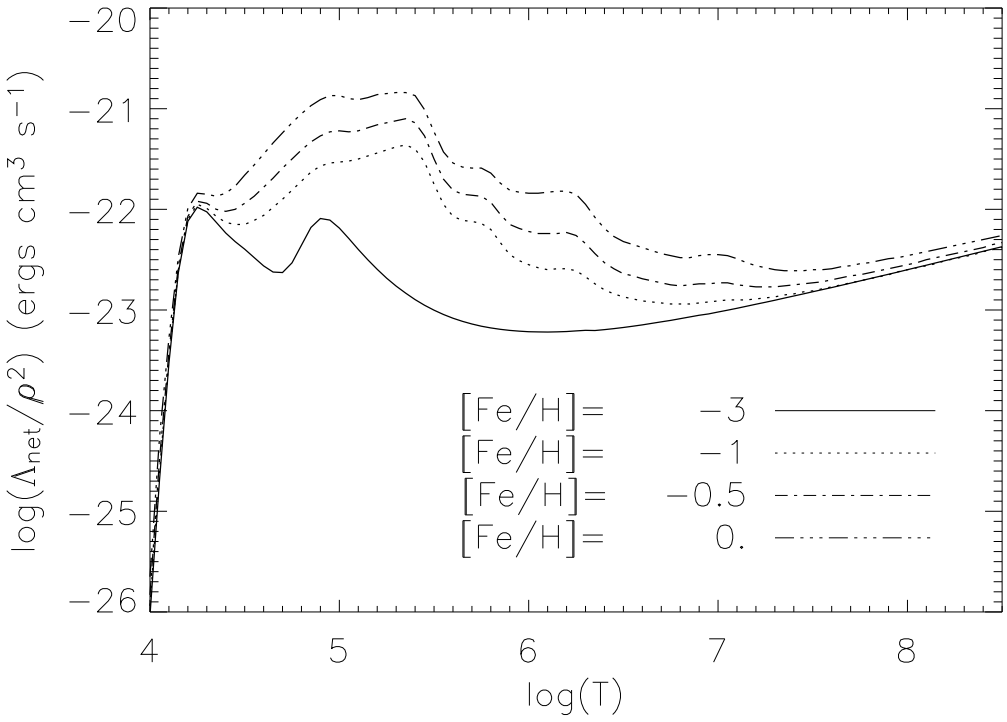

Figure 1.2: Radiative cooling rate of a gas as a function of temperature for a number of different metallicities. Data from Sutherland & Dopita (1993). The ratios between the different metal abundances are assumed to be in solar proportion. The cooling rate falls off below 10^4K due to the lack of molecular hydrogen in these calculations. The mid temperature cooling peaks are due to electron collision ionization (and associated radiation due to recombination) and at high temperatures Bremsstrahlung radiation dominates.

of gas that can cool into galactic disks and controls with great sensitivity the properties of the interstellar medium

1.2.3 The Molecular Component of the Interstellar Medium

It has long been understood that the vast majority of star formation in the Milky Way takes place inside molecular clouds (see e.g. Lada and Lada (2003) and references therein). The study of molecular clouds is therefore of critical importance in our understanding of how galaxies form and evolve. In this section we discuss the formation and properties of

molecular clouds before investigating the mechanism by which they collapse into stars in the next section.

The Formation of Giant Molecular Clouds

Although the early models by Field et al. (1969) assumed that the different phases of the ISM were in pressure equilibrium, modern observations paint a much more complex and dynamic situation, in which the ISM is shaped by turbulence, possibly powered by SNe and the large-scale dynamics of the galactic disk itself, see e.g. Burkert (2006) for a recent review. A galaxy-wide effect seems required to explain the observed Hubble-type dependence of cloud properties.

Yet how GMCs form in this complex environment is not well understood (Blitz and Rosolowsky (2004)). Some authors have suggested that GMCs form by the coagulation of preexisting molecular clouds (Kwan and Valdes (1983); Oort (1954)) while others have argued that GMCs form primarily from atomic gas through instability or large-scale shocks (Blitz and Shu (1980)), a viable mechanism by which this could occur is the formation of convergent flows induced by a passing spiral arm (Ballesteros-Paredes et al. (1999)). Of course, both modes of formation may occur: in high density regions, where the vast majority of hydrogen is molecular it seems likely that molecular clouds form from the coagulation of smaller clouds, whereas in the outskirts of galaxies where the gas is predominantly atomic the compression of gas in spiral density shock waves provides a more plausible formation mechanism.

Observationally, star formation in disks seems to occur only above a given surface-density threshold (Kennicutt (1989)), with star formation dropping abruptly below the threshold even though the gaseous disk may extend far beyond it. Schaye (2004) describes a model in which this threshold arises naturally due to the thermal instability when gas cools from 10^4 K to the cold phase (~ 100K), rendering the disk gravitationally unstable on a range of scales. This suggests that a combination of thermal instability and large-scale dynamics may be responsible for determining when and where GMCs form.

Elmegreen (2000) discusses observational evidence that GMCs are in fact short lived entities that form, make stars and disperse again on their dynamical time scale. This short time-scale alleviates the need for an internal energy source to sustain the observed internal supersonic turbulence, something that had puzzled astronomers for a long time. Pringle et al. (2001) discuss this assumption in more detail, and suggest that GMCs form from agglomeration of the dense phase of the ISM, already in molecular form, when

compressed in a spiral shock. They envisage the pre-existing molecular gas to be in dense 'wisps', in the inter arm regions, formed from atomic gas by shocks, as simulated by Koyama and Inutsuka (2000).

Recent numerical simulations support the view that when clumpy gas is overrun by a density wave, it produces structures that resemble GMCs. Wada et al. (2002) present high-resolution two-dimensional simulations of the evolution of perturbations in a cooling, self-gravitating disk in differential rotation. They show that the disk develops stationary turbulence, even without any stellar feedback. Bonnell et al. (2006) and Dobbs et al. (2006) perform three dimensional simulations of the passage of clumpy cold gas through a spiral shock. Their simulations produce dense clouds, with large internal velocity dispersion reminiscent of the 'supersonic turbulence' seen in GMCs. They note that the velocity dispersion is generated on all scales simultaneously, in contrast to what is usually meant by turbulence where energy cascades from large to small scales. Mac Low and Klessen (2004) review the current state of the art in simulations and models of GMC formation, including references to more recent work. In this picture of clouds, GMCs are temporary structures formed and dissolving in converging flows, do not require an internal source of energy, are not in virial or pressure equilibrium, and need not even be gravitationally bound. They point-out that the relative contribution of galactic rotation and stellar sources to driving the observed turbulence is not clear.

The Properties of Giant Molecular Clouds

As implied by the name, molecular clouds are dense regions consisting primarily of molecular hydrogen that forms via interaction with interstellar dust grains (Cohen (1976b)). Molecular clouds have a small spatial extent and as such are very hard to resolve. Over recent years, however, surveys of molecular clouds have been undertaken in the MW and various nearby galaxies. When it became possible to spatially resolve individual molecular clouds it was found that they do not have a simple structure, but are instead complex and dynamic objects. The study of the structure of molecular clouds is very much an active research area today. Hierarchical temperature and density peaks are known to exist within molecular clouds, with dense cores embedded within larger, less dense, molecular clumps. These less dense clumps are, in turn, embedded in larger and less dense clumps (Dickman et al. (1990); Elmegreen and Falgarone (1996)). Densities decrease by approximately an order of magnitude going from clump to clump (Cesaroni et al. (1994)). This has led to molecular clouds being described as having a 'fractal' struc-

ture (see Elmegreen and Elmegreen (2001) and references therein)

1.3 The Stellar Content of Galaxies

In this section we discuss the formation, properties and observations of the stellar content of galaxies. We first introduce the mechanism by which most stars are thought to form; the collapse of turbulent molecular clouds (section 1.3.1). We then describe the ways in which star formation can be observationally probed in distant galaxies (section 1.3.2). Finally we talk about how massive stars end their lives as supernovae (section 1.3.3). This process is of critical importance in the energetics and dynamics of galaxy evolution and additionally supplies the vast majority of the heavy elements in the universe (1.3.4).

1.3.1 Star Formation in Molecular Clouds

In the older picture of cloud formation, GMCs were long-lived, gravitationally bound, virialised objects. The presence of supersonic turbulence ensures that clouds do not immediately collapse to form stars, as this would predict a star formation rate for the MW which is far higher than observed. Locally unstable clumps collapse to form proto-stars, which built-up their mass to produce the initial mass function (IMF) through competitive accretion (e.g. Bonnell et al. (1997)).

However simulations show that the energy contained in supersonic motions is quickly dissipated even in the presence of magnetic fields (see references in Mac Low and Klessen (2004)). To sustain the turbulence hence requires an energy source, for example from star formation, yet some clouds have turbulence but no current star formation.

The modern picture is one in which clouds are short-lived structures and the turbulence results from the same process that formed the cloud in the first place. Observationally, GMCs turn a small fraction $\epsilon_\star \approx 0.1$ of their mass into stars before they disperse again. This low star formation efficiency of clouds may be due to the fact that they are short lived. The short life times of (star forming) clouds also follows from the small age spread in star clusters (see e.g. Gomez et al. (1992)), and indicates that star formation in a given cloud only lasts for a few million years. The short life-times of GMCs then also suggests that competitive accretion (e.g. Bonnell et al. (1997)) is less important in shaping the IMF (Padoan and Nordlund (2002)).

Turbulence generates a range of substructures inside a GMC, and Padoan and Nordlund (2002) suggest that such 'turbulent fragmentation' builds a mass spectrum of protocores, some of which will collapse under their own gravity to form stars. The resulting IMF is a power-law due to the self-similar nature of the turbulence. Only cores dense enough so that self gravity can overcome their magnetic and thermal energy can collapse. This consideration flattens the IMF at low masses, and prevents very low-mass cores from forming stars. They also argue that the maximum mass is a fraction of the overall cloud mass.

A young stellar population does of course dump a lot of energy into its surroundings through stellar winds, ionisation, and SN explosions. Even if these do not drive the observed turbulence, they may contribute to the destruction of the cloud, and prevent further star formation. Most simulations use such feedback from star formation to regulate the star formation rate.

1.3.2 Observations of Star Formation Rates

In all but the closest galaxies individual stars remain unresolved even with the most modern telescopes. Most information on the star formation properties of galaxies comes from integrated light measurements in the ultra-violet (UV) and far-infrared (FIR) and from observation of the strength of recombination lines, for example Hα. We will now discuss each of these methods, highlighting both the advantages and disadvantages of each one.

Ultraviolet Continuum: Young stars dominate the emission at ultraviolet wavelengths so the luminosity scales almost linearly with SFR (Kennicutt (1998)). The conversion between UV flux in a given band and the SFR can be derived using stellar synthesis models (see e.g. Bruzual and Charlot (1993)). The optimal wavelength range to make UV observations is $1250 - 2500$ angstroms, which is longward of the Lyman alpha forest but still short enough to remove most contamination from older stellar populations. These wavelengths cannot be observed from ground based telescopes for redshifts $z < 0.5$ but are visible from redshifts $1 - 5$. For this reason the most complete UV studies of nearby galaxies have been carried out by balloon, rocket and space experiments.

Far-Infrared Continuum: Interstellar dust in galaxies will absorb a significant fraction of the stellar luminosity. This heats the dust, which then re-emits photons with a black body spectrum characterised by the temperature of the dust. In real galaxies the situation is somewhat complicated by the fact that the dust exists at at least two discrete

temperatures (e.g. Cox et al. (1986)), and so any analysis must take into account contributions from a 'warm' component associated with dust around star-forming regions and another 'cool' component, which is emitted by the diffuse dust throughout the rest of the galaxy. This makes the calibration of the conversion factor between FIR luminosity and SFR a somewhat controversial subject, and only relatively recently have the calibrations used by different authors (Hunter et al. (1986); Lehnert and Heckman (1996); Meurer et al. (1997)) begun to agree to within better than a few tens of percent.

Recombination Lines: Hydrogen atoms that interact with high energy photons can be ionized, and then at a later time recombine with the electrons, emitting photons at specific frequencies. Stars with masses $> 10M_{\odot}$ and lifetimes of < 20Myr contribute significantly to the flux of these ionizing photons, so measuring the strength of the recombination lines provides an estimate of the instantaneous star formation rate. The most commonly used line is Hα, which is both very strong and exists in a part of the spectrum visible from the ground (at least for $z < 0.5$). Calibrations between Hα flux and SFR have been provided by a number of authors (Kennicutt (1983); Leitherer and Heckman (1995)). The Hα line is used very commonly as an SFR indicator due to its high sensitivity. Additionally SFRs measured by this method are very sensitive to the slope of the IMF, and as such provide a good way of constraining its properties (Kennicutt (1983))

At high redshift the Hα line is redshifted outside of the visible window and therefore must be investigated with space-based experiments. There is therefore a lot of interest in using higher order lines as SFR tracers. The integrated strengths of the other hydrogen recombination lines is very low, and they are strongly influenced by absorption in stellar atmospheres. However, the [OII] forbidden-line doublet is sufficiently strong and sufficiently well behaved that it is possible to calibrate between SFR and [OII] luminosity (Gallagher et al. (1989)). SFRs measured by this line are somewhat more uncertain than similar ones calculated using the Hα line, but the [OII] line is visible up to a higher redshift.

Many large surveys of the global SFRs of galaxies have been carried out using all of the techniques summarised above. Hα(Cohen (1976a); Young et al. (1996)), UV(Deharveng et al. (1994); Buat (1992)), FIR(Colless et al. (2001); Moshir et al. (1992)). The absolute SFRs in galaxies (usually expressed in units of $M_{\odot}/\mathrm{yr}$, show an enormous range, from virtually zero in gas-poor elliptical, S0, and dwarf galaxies to $20M_{\odot}/\mathrm{yr}$ in gas-rich spiral galaxies and $100M_{\odot}/\mathrm{yr}$ in high redshift starburst galaxies, although these figures are

somewhat uncertain due to the as yet poorly constrained shape of the stellar initial mass function, which although measured for our own galaxy (Salpeter (1955); Kroupa (1998); Chabrier (2001)), is still unknown in high redshift objects, where the physics of the ISM may be significantly different to the low redshift universe.

1.3.3 The Physics of Supernova Explosions

A supernova event marks the end of life for some stars. There are a variety of different mechanisms by which SN can occur. These include accretion of matter from a companion star (Type Ia) or exhaustion of its own fuel supply (Types Ib, Ic, and II). Each of these mechanisms can take a star to the point where it can no longer support itself against gravity and thus collapses, throwing off its outer layers in a burst of energy that outshines an entire galaxy. This discussion is divided into sections on Type I supernovae and Type II supernovae (see e.g. Woosley and Weaver (1986) for a review). The distinction between supernova types is based on the presence or absence of hydrogen lines in the spectrum. The two types of supernova are almost totally distinct phenomena, the only relations being the approximate equality of the explosion energy (10^{51} erg) and the fact that both produce heavy elements.

Type II supernovae are estimated to occur once every 44 years in our galaxy, and Type I's every 36 years (Tammann (1982)). In this section we will discuss the physics of both Type I and Type II supernovae before summarising the behaviour of the resulting supernova blast wave, and describing the metal yields from different types of supernova.

Type 1 Supernovae

Supernovae that do not show prominent hydrogen lines are classified as Type I. Type I supernovae are not a homogeneous set and following Elias et al. (1985) we will consider Type I supernovae to be split into three subgroups, Type Ia (strong Si II line), Type Ib and Ic (characterised by an absence or presence of a Helium line)

Type Ia supernovae are generally considered to be due to an accreting white dwarf in a binary system that grows to a critical mass (Whelan and Iben (1973)), and ignites carbon or helium. The energy from the burning disrupts the star at high velocity. In general no neutron star or black hole remnant would be expected from this type of explosion.

Type Ib and Ic supernovae are thought to be associated with massive stars (Wheeler and Levreault (1985)), and perhaps represent the collapse of a star massive enough to

have blown away its outer hydrogen envelope. The supernova explosion in this case would proceed in a way similar to the iron core-collapse supernovae discussed in the following section.

Type 2 Supernovae

Type II supernovae are generally believed to be a consequence of gravitational collapse of massive stars ($M > 8M_{\odot}$). Type II supernovae are characterised by the presence of hydrogen lines in their spectra (Branch et al. (1981)). The fate of massive stars depends sensitively on their mass and composition (metallicity) at birth. See e.g. Heger et al. (2003) for a recent review.

The mass range for stars that undergo supernova events fitting our description of a Type II supernovae is bounded on the lower end by the largest mass of star that can become a white dwarf and on the upper end by the most massive stars that will keep an appreciable hydrogen envelope. For isolated stars, this mass can be estimated in a number of ways: by finding the mass of the progenitors of the largest white dwarfs (Iben and Renzini (1983)), with statistical arguments on the occurrence of supernovae (Tammann (1982)) and theoretical models for white dwarf formation (Iben (1985)). The values obtained by each of these techniques are in good agreement at $8M_{\odot}$.

The mass of the most massive star that maintains a hydrogen envelope is uncertain and depends strongly on metallicity, and varies from approximately $30M_{\odot}$ in solar metallicity stars to hundreds of solar masses in metal free stars (Heger et al. (2003)). Stars with masses greater than $8M_{\odot}$ and with a hydrogen envelope will undergo what we observe as Type II supernovae. More massive stars may undergo a violent instability

Stars in the mass range $8 - 11M_{\odot}$ (Barkat et al. (1974)) form, in their later stages of evolution, a core of Oxygen, Neon and Magnesium (Miyaji et al. (1980)). According to the models of Barkat et al. (1974) these O-Ne-Mg cores grow through a complicated series of shell burnings, until two competing effects *electron capture by Ne and Mg* (acting to allow the core to collapse gravitationally) and *Oxygen deflagration* (acting to cause the star to explode) begin to act against each other. Miyaji et al. (1980) found that electron capture is the dominant effect and stars in this mass range have cores that collapse into neutron stars.

More massive stars undergo supernovae via a different mechanism. These stars are massive enough to ignite all six nuclear burning stages (hydrogen, helium, carbon, neon, oxygen and silicon) until an iron core is formed. As the iron core grows its temperature

and density increase until photons have enough energy to destroy heavy nuclei, a process known as photodisintegration. Most importantly:

$$^{56}_{26}\mathrm{Fe} + \gamma \rightarrow 13\,^{4}_{2}\mathrm{He} + 4n\,, \tag{1.18}$$

and

$$^{4}_{2}\mathrm{He} + \gamma \rightarrow 2p^{+} + 2n\,. \tag{1.19}$$

Photodisintegration can, in a very short time, undo what the star has been trying to do its entire life, namely produce elements more massive than H and He. This process of stripping iron down to protons and neutrons is highly endothermic and the core begins to collapse. Under the extreme conditions that now exist the free electrons that had assisted in supporting the core through electron degeneracy pressure are captured by heavy nuclei and the protons that were produced through photodisintegration

$$p^{+} + e^{-} \rightarrow n + \nu_{e}\,. \tag{1.20}$$

Most of the core's support is now gone and it begins to collapse rapidly until supported by neutron degeneracy pressure, at which point it rebounds. The outer layers of the star collapse inwards on the core and rebound, sending shock waves outward. This is a Type II supernova.

The late stage evolution of stars massive enough to shed their hydrogen-envelopes is more complex still and depends upon the rotation speed, metallicity and mass of the star. Depending on mass and metallicity different instabilities become important, including the 'pair instability' (Fowler and Hoyle (1964)). The resulting supernova events can have a large range of effects on the supernova remnant from complete disruption of the star (Heger and Woosley (2002)) to the formation of a massive black hole (Fryer et al. (2001)).

Main sequence stars with masses greater than $8M_{\odot}$ are short lived so Type II supernovae are observed only in regions that have recently formed stars and therefore rarely in elliptical galaxies (Tammann (1974)).

The Evolution of a Supernova Remnant

In the classical model of Supernova Remnant (SNR) evolution (Chevalier (1977); Gull (1973); Woltjer (1970)) there are four main stages or phases through which a SNR passes: Free Expansion, Adiabatic, Snowplough and Dispersal. We simplify the discussion of SNR evolution by treating a supernova as an injection of some amount of energy, ϵ_0, to an amount of mass M_0, which acquires an initial velocity, v_0. At time t the resulting blast

wave has a radius R_s. We treat the system as spherically symmetric and consider how it interacts with the surrounding ISM of density n_{ism} hydrogen atoms per cubic centimetre (Woltjer (1972)) we can treat each of these phases in turn:

I. Free Expansion: The supernova ejecta will sweep up gas from the ISM as it travels. As long as the mass swept up is very much less than the mass in the ejecta:

$$\frac{4\pi}{3} n_{ism} m_H R_s^3 < M_0 \,, \tag{1.21}$$

then the expansion of the remnant will be essentially free and much of the development will depend on the initial conditions.

II. Adiabatic: The swept up mass will increase until it exceeds M_0, we then enter the phase where the behaviour of the expanding shock wave can be described by the self similarity solution of Sedov (see section 3.2.6 for a more detailed discussion of this phase). During this phase the shock wave has a radius $\propto t^{2/5}$, other properties of the shock wave are discussed more fully in section 3.2.6. This solution for the behaviour of the shock wave assumes that the gas is adiabatic. In reality, a hot gas will cool radiatively so this regime is valid whilst

$$\int \left(\frac{d\epsilon}{dt}\right)_{\rm rad} dt < \epsilon_0 \,. \tag{1.22}$$

By assuming that all of the mass internal to the shock wave is contained in a thin shock wave it can be shown (Woltjer (1972)) that the time at which the integrated radiative energy loss equals $\frac{1}{2}\epsilon_0$ is

$$t_{\rm rad} = 1.1 \epsilon_0^{4/17} n_{ism}^{-9/17} \,, \tag{1.23}$$

and assuming that the adiabatic relation for the velocity still holds:

$$v_{\rm rad} = 2.0 \times 10^7 n_{ism}^{2/17} (\epsilon_0 / 1 \times 10^{50})^{1/17} {\rm cm/s} \,. \tag{1.24}$$

This evaluates to 200km/s and depends only very weakly on the parameters of this particular supernova. This velocity is taken to mark the end of phase II and the beginning of phase III.

III. Snowplough: When the radiative losses become dominant the behaviour of the shock wave will change. Matter passing through the shock wave cools rapidly and the density becomes high. In this phase we can consider the shock wave to be a thin, dense shell ploughing through the ISM. Pressure forces are no longer important and the shock moves with a constant radial momentum. The equation of motion simply becomes (Chevalier (1977)):

$$(4\pi/3) n_{ism} m_H R^3 v = C_1 \,, \tag{1.25}$$

where C_1 is a constant. Integrating with respect to time we obtain that

$$R = R_{\rm rad}\left(\frac{8}{5}\frac{t}{t_{\rm rad}} - \frac{3}{5}\right)^{1/4}, \tag{1.26}$$

where the subscript $\rm rad$ represents the values at the transition from stage II to stage III.

IV. Dispersal: After the shell has slowed down for some time its velocity becomes comparable to the random thermal motions of the ISM (around 10km/s). The shell disperses and loses its identity. The matter in the shell is mixed in with the material in the surrounding ISM, depositing the heavy elements that formed during the supernova explosion in the ISM.

The model as introduced so far neglects a number of important physical processes, including the presence of magnetic fields in the swept up shell, the pressure of cosmic rays within the shell, the inhomogeneity of the ISM and thermal conduction between the hot gas of the supernova remnant and the surrounding cold molecular clouds (this process is discussed in section 3.2.7). These additional physical processes mean that the problem of understanding the evolution of a SNR is not yet closed and is still an active area of research.

1.3.4 Metal Enrichment

As discussed previously in this section, stellar nucleosysnthesis and supernova explosions create elements heavier than the H and He formed by BB nucleosysnthesis (see section 1.1.2).

In astronomy any elements heavier than helium are described as 'metals'. The fraction of metals in a sample of gas is usually quantified using its 'metallicity', defined for Iron as

$$\left[\frac{\rm Fe}{\rm H}\right] \equiv \log_{10}\left(\frac{N_{\rm Fe}}{N_{\rm H}}\right) - \log_{10}\left(\frac{N_{\rm Fe}}{N_{\rm H}}\right)_{\odot}, \tag{1.27}$$

Where N represents the number density of a particular species, and the subscript $\odot$, is the value for the sun. A star with a metallicity equivalent to that of the sun therefore has a metallicity of 0. The observed range of metallicities in the MW is -4.5 (old, extremely metal poor stars) to +1 (very young, metal rich stars).

Calculating the expected metal yield from a supernova explosion is not an easy task, and is a function of the mass, metallicity and rotational speed of the star. Additionally each of the different supernova mechanisms will produce different proportions of metals. There is still uncertainty about the precise metal yields from all types of supernovae.

Various authors have investigated the yields from Type II supernovae (Woosley and Weaver (1995); Thielemann et al. (1996); Limongi et al. (2000)) and although much work has been performed in this area, the resulting supernova yields are still somewhat uncertain (Travaglio et al. (2004)) due to our only approximate understanding of the precise mechanism by which core collapse supernovae are triggered. The dominant product from these supernovae are thought to be the α elements, defined as the $Z > 22$ multiples of He (e.g. Mg, Si, S, Ca), which are synthesised by alpha capture during SNII.

The nucleosysthesis expected in type I supernovae has been calculated by several groups (e.g. Nomoto et al. (1984)). The dominant products from these supernovae are the iron peak nuclei (elements with an atomic number near that of iron, and high binding energy per nucleon).

The differing yields from different supernova types may, therefore, be used to probe the star formation history of the universe. Since Type II supernovae occur on a very short timescale, the relative abundances of the α and Fe elements can provide clues as to the history of star formation in a given object (Tissera et al. (2002))

1.4 Motivation for This Thesis

In this section we provide a brief overview of the contents of each chapter.

1.4.1 Computational Cosmology

Cosmological structure formation is a complex and non-linear process, therefore a thorough understanding of both simulation algorithms and their shortcomings is of crucial importance in carrying out scientific investigation in this field. This chapter is primarily concerned with numerical issues affecting cosmological simulation. We begin by providing an overview of the simulation techniques used in cosmological structure formation simulations, before carrying out numerical tests of two simulation codes. We then introduce the process by which cosmological initial conditions are created and using this knowledge probe the numerical effects that alter the properties of a halo as a function of its mass resolution.

1.4.2 Statistical Modelling of the Interstellar Medium

Most computational models of star formation and supernova feedback are crude at best and rely on straight fits to empirical laws. In this chapter we introduce in detail a novel,

physically motivated prescription for the simulation of a multiphase ISM on galactic scales. This model could find many applications, two of which are discussed in the following two chapters.

1.4.3 The Interstellar Medium in Isolated Galaxies

In chapter 4 we investigate the properties of the interstellar medium in an isolated disk galaxy using the model introduced in chapter 3. We find that the model described in chapter 3 can reproduce many of the observed properties of quiescent galaxies, and we undertake a short investigation into the effects of changing some of the physics of the ISM.

1.4.4 Simulating Galaxy-Galaxy Interactions

Mergers and other gravitational interactions between galaxies are an essential part of hierarchical galaxy formation scenarios, and it is likely that the vast majority of galaxies in the observable universe have been shaped by collision events (Toomre (1977)). This suggests that in order to understand the properties of present day galaxies we require a thorough investigation of the dynamics and energetics of galaxy-galaxy interactions. In this chapter we analyse the properties and behaviours of previously quiescent disk galaxies as they undergo gravitational interactions with one another. It is found that many of the properties of the resulting remnant, including density and brightness profiles, the age gradient and the makeup of the ISM match very well with the corresponding properties in local early-type galaxies. Additionally during the merger strong star formation is observed in the tidal features of the interacting galaxy pair.