1 *Introduction*

Our understanding of both the large-scale properties of our Universe and the processes through which galaxies form and evolve has greatly improved over the last few decades, thanks in part to new observational probes and more refined numerical simulations. While the precision with which we measure the cosmic background radiation, the distribution of matter and the properties of galaxies continues to increase, we are improving our simulations to include more physical processes and to resolve ever smaller scales. We are learning just how deeply cosmology and galaxy formation are intertwined, and the need to model them simultaneously in order to advance both fields is growing rapidly. In this thesis, we investigate how galaxy formation can alter the structure of the Universe on a large range of scales, and how measuring the structure of the Universe can in turn help us to constrain models of galaxy formation.

1.1 Large-scale properties of the Universe

About 13.8 billion years ago, the Universe came into existence in an event we call the Big Bang. From that moment on, it has been continually expanding. As a consequence of the Universe being both isotropic and homogeneous on large scales, the rate of its expansion at any particular time can be related to the current one through a simple function of only four parameters.[1] These are the present-day matter content of the Universe, $\Omega_{\mathrm{m},0}$; its radiation content, $\Omega_{\gamma,0}$; its curvature, $\Omega_{\mathrm{k},0}$; and the contribution of the cosmological constant, $\Omega_{\Lambda,0}$, which we presently refer to mainly as dark energy. Since $\Omega_{\mathrm{k},0}$ is defined such that the sum of these parameters is by definition equal to unity, this means that only three are independent. Determining the values of these parameters with ever-growing precision is one of the main aims of cosmology, as together with the Hubble constant, H_0 (the present rate of expansion), they fully determine the evolution of the Universe as a whole.

Currently, the strongest constraints on these numbers come from observations of the oldest light in the Universe: the cosmic microwave background, or CMB, a relic from the Big Bang. The CMB was last scattered when the Universe was only about $380,000$ years old, at a time when the expansion had cooled the Universe down sufficiently for protons and electrons to combine and form neutral hydrogen and helium (an event called "recombination"), allowing light to travel freely for the first time. It shows us the Universe at the earliest time we could possibly observe its light, therefore informing us about the initial conditions, from which any successful model should be able to explain the properties of the Universe as we see it today.

The CMB is incredibly smooth, indicating a very high level of homogeneity – the relative variations in the density of baryonic matter (or, "normal", visible matter) we see in it are of the order of 10^{-5} (see Figure 1.1). In order for these fluctuations to grow into the galaxies we see today, most matter in the Universe needs to be (cold) dark matter[2], which is indeed what different observations indicate. Through very precise CMB measurements using e.g. the Wilkinson Microwave Anisotropy Probe (WMAP) and Planck satellites, the dark matter fraction and a host of other cosmological parameters (including all parameters mentioned thus far) can be determined with ever greater accuracy (e.g. Hinshaw et al., 2013; Planck Collaboration et al., 2013). These measurements indicate, for example, that the

[1] In more detail, the evolution of the Universe also depends on the equations of state (the ratios of pressure and density) that characterise these parameters. The equation of state of the radiation content depends on the number of relativistic particle species, while that of the cosmological constant depends on the nature of dark energy, both of which are not completely certain yet.

[2] If dark matter is "warm" or "hot", this means that it consists of particles with velocities that are sufficiently high at the time of decoupling to stream out of density fluctuations, thus preventing their growth. This process, called free streaming, sets a minimum scale above which fluctuations can form and depends mainly on the masses of the particles, and their cross sections. While we already know that not all dark matter can be hot, warm dark matter is not yet completely ruled out, although its particle mass is strongly constrained (e.g. Viel et al., 2013).

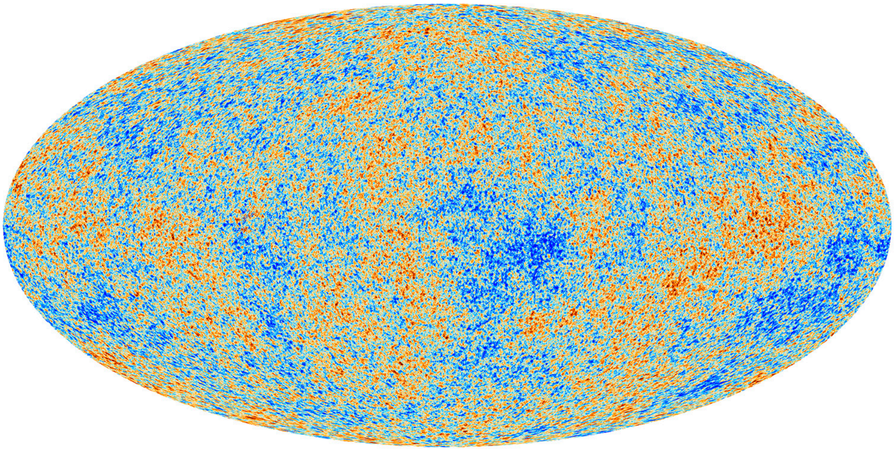

Figure 1.1: The Cosmic Microwave Background (CMB) as captured by the Planck satellite. This is a snapshot of the Universe only $380,000$ years after the Big Bang. It is homogeneous and isotropic on large scales, but very small fluctuations exist nonetheless. The colour scale shows relative differences of order 10^{-5}.

Universe is "flat" (i.e. space is not curved but Euclidean), and although everything we observe directly is baryonic matter, the Universe is in fact dominated by dark matter and dark energy. We refer to the model that contains all these ingredients as the ΛCDM-model, currently the standard model of Big Bang cosmology.

1.2 Testing the standard cosmological model through clustering

As we mentioned, in order for any model of the Universe to be truly successful, it has to be able to explain all that we see on large scales today. This includes, for example, the current (accelerating) rate of expansion, but also the different galaxy populations we observe and the distribution of matter. The latter is the main focus of this thesis: the clustering of matter, i.e. the structure of the Universe. The matter distribution is completely determined by the initial conditions of the Universe; therefore, it is in principle possible to derive all cosmological parameters by examining how matter is organised in the present-day Universe, provided one understands how structure forms and evolves. In what follows, we present a simplified view of the formation of structure, starting from the very first density fluctuations in our Universe.

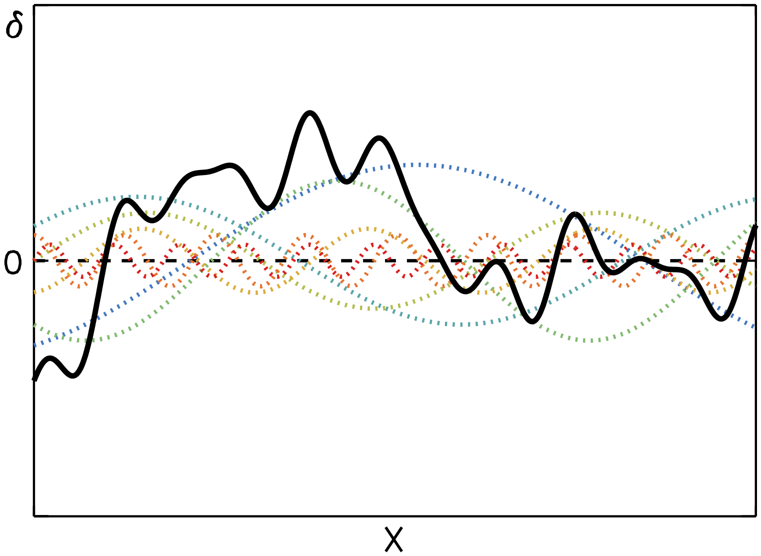

FIGURE 1.2: Illustration of the density field as a field of fluctuations. As a one-dimensional analogue, we show how seven independent harmonic waves with different amplitudes, wavelengths and phases (indicated by dotted lines) together add up to the fluctuation field indicated by the solid curve. Notice that because of contributions of large waves, high-δ fluctuations are often found close together (i.e. they cluster).

1.2.1 Linear structure formation

Let us consider some part of the Universe with a mean density $\bar{\rho}$. At[3] any three-dimensional position $\mathbf{x}$ we can calculate a local density $\rho(\mathbf{x})$, which may differ from the mean. We can now define the density contrast field, or density fluctuations field, as:

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}. \tag{1.1}$$

If δ is positive at some position $\mathbf{x}$, this means that there is a local overdensity. Under the influence of gravity this overdensity will grow[4], attracting more and more matter and thereby forming structure.

In order to understand what happens as these overdensities grow, let us first consider the simplest picture of structure formation: the linear one. Linear structure formation applies when the density fluctuations are very small, i.e. $\delta \ll 1$. This is indeed valid for the early Universe, which was extremely homogeneous and

[3] To be precise, a density can never truly be defined at a singular location: one needs to assume some smoothing scale.

[4] For baryonic matter, overdensities may be stable against collapse due to pressure forces. Dark matter – which dominates the matter content of the Universe – does not feel pressure, however, and is able to form structure more freely. We will briefly return to this point later in this chapter.

therefore contained only small fluctuations.[5] As density fluctuations influence each other through gravity, they do not evolve independently. However, if we consider each density fluctuation as a superposition of plane waves, then in the linear regime these waves *do* evolve independently. Additionally, this view allows us to consider the growth of structure in a statistical sense, at given *scales* instead of at given locations, which is far more meaningful in a largely isotropic and homogeneous Universe.

An illustration of this wave picture is shown in Figure 1.2. A one-dimensional density fluctuation is shown as a solid black line. Each such fluctuation can be uniquely decomposed into harmonic waves with different amplitudes, wavelengths and phases; in this case, into the seven waves shown as dotted lines. Notice that, mainly due to the contributions of the longer waves, high-δ fluctuations tend to cluster – i.e. they are likely to be found close together. As we will see later, this has large consequences for how matter is organised today.

The relation between the spatial fluctuations δ and the density waves $\delta_{\mathbf{k}}$ can be expressed through a discrete Fourier transform:

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}\, e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{1.2}$$

where $\mathbf{k}$ is the wave vector, related to the wavelength by $\lambda = 2\pi/|\mathbf{k}|$. We can now quantify the amount of structure on any Fourier scale $k = |\mathbf{k}|$ by squaring the amplitudes of these density waves, averaging over all waves with the same wavenumber k to obtain a statistic called the matter power spectrum:

$$P(k) \equiv \left\langle |\delta_{\mathbf{k}}|^2 \right\rangle_k . \tag{1.3}$$

Inflationary theory[6] predicts that the primordial power spectrum should scale as a power law:

$$P(k) \propto k^{n_{\mathrm{s}}}, \tag{1.4}$$

with a spectral index n_{s} that is very close to unity, meaning that to good approximation the fluctuations in the gravitational potential were scale invariant.[7]

How the linear density fluctuations evolved from primordial times is best described in terms of the scale factor of the Universe, $a(t)$, which is a dimensionless length scale that gives a measure of the size of the Universe when it has an age t. By definition, $a(t_0) = 1$, where t_0 is the current age of the Universe. Since the Universe is continually expanding, a was smaller in the past, and infinitesimally

[5] Note that a completely homogeneous, uniform density field cannot exist: at the very least, microscopic variations in density must exist due to quantum mechanics. Incidentally, such quantum fluctuations are expected to be the seeds of the variations we see in the CMB – stretched out to macroscopic scales by a process called inflation very shortly after the Big Bang – and consequently of all structure existing today.

[6] Recently, the first direct evidence for inflation was found by the BICEP-2 team in the form of a gravitational wave signal in the CMB, see BICEP2 Collaboration et al. (2014).

[7] The scale-invariant power spectrum is also called the Harrison-Zel'dovich power spectrum. Current CMB measurements by the Planck satellite indicate $n_{\mathrm{s}} \approx 0.96$.

small at the moment after the Big Bang. It is related to the redshift z through $a = 1/(1+z)$; at $z = 1$, the distance between two points in the Universe was thus twice as small as it is now. For reference, the redshift of the CMB/recombination is $z \approx 1100$.

As we mentioned at the beginning of this chapter, the evolution of the Universe is determined by its constituents, most importantly matter, radiation and dark energy.[8] Consequently, the growth of fluctuations at any time depends on which of these constituents dominates. In our simplified picture, a linear density fluctuation in some spherical volume is expected to grow as:[9]

$$\delta \propto \frac{1}{\bar{\rho} a^2}, \tag{1.5}$$

where $\bar{\rho}$ is the mean (energy) density of the dominant component of the Universe at some time t. The amount of matter in the Universe is (to a very high degree) constant, meaning that its density just scales inversely with the volume of the Universe:

$$\bar{\rho}_{\mathrm{m}} \propto a^{-3}. \tag{1.6}$$

Therefore, when matter dominates the Universe, linear fluctuations grow as $\delta \propto a$. The energy density of radiation, on the other hand, does not only scale inversely with the volume, but by an additional factor a since its energy is not conserved due to photons being redshifted during expansion. Hence:

$$\bar{\rho}_{\gamma} \propto a^{-4}, \tag{1.7}$$

meaning that during radiation domination linear fluctuations may grow as $\delta \propto a^2$. Finally, for dark energy, which is a property of space itself and therefore has a constant density, we have:

$$\bar{\rho}_{\Lambda} \propto a^{0}, \tag{1.8}$$

meaning that when dark energy dominates, density fluctuations cannot grow at all: they are damped by the expansion of the Universe.

However, this damping is not exclusive to the Λ-dominated era. This is related to the existence of the horizon, the maximum distance between causally connected regions. If two regions are farther apart than this, i.e. farther than light (and gravity) could have travelled within the age of the Universe, then they could not have been in causal contact. Fluctuations on scales smaller than the horizon can be damped if the Universe expands faster than they collapse, which is the case during the era of radiation domination. In short, this limits the growth of linear fluctuations to at most logarithmic growth – much slower than the otherwise power-law growth[10]. A summary of the growth rates for both sub- and superhorizon fluctuations is shown in the table at the top of the next page. Here λ is the wavelength of a fluctuation and r_{H} is the horizon scale.

Since the densities of each of the constituents of the Universe scales differently with the scale factor of the Universe, it is clear that each dominates in some era.

[8] As observations show the Universe to be almost completely flat, geometrically speaking, these

	γ-dom.	m-dom.	Λ-dom.
$\lambda < r_{\rm H}$	damped	$\delta \propto a$	damped
$\lambda > r_{\rm H}$	$\delta \propto a^2$	$\delta \propto a$	damped

Radiation, which scales most steeply with a, must have dominated when the Universe was very young (i.e. a was very small), followed by matter, followed by dark energy. Indeed, radiation dominated the content of the Universe up to a redshift of $z \approx 3600$ (corresponding to an age of the Universe of approximately $50,000$ years), and dark energy has been dominating since $z \approx 0.4$ (for approximately the last 4.2 billion years), meaning that the Universe was matter-dominated during the majority of its existence, allowing new structure to form.

Combining this insight with the table shown above, we conclude that a special scale exists, namely the scale of the horizon between the radiation and matter dominated eras, $r_{\rm H,eq}$. This scale depends on several cosmological constants, but roughly[11] $r_{\rm H,eq} \sim 10^2\, h^{-1}\,$Mpc. Fluctuations larger than this scale were able to grow before the matter-dominated era, while smaller fluctuations were damped. Afterwards, linear fluctuations on all scales could grow at the same rate.

It can be shown that the damping of subhorizon fluctuations depends on their size: smaller fluctuations were damped more strongly by the expansion of the Universe. This means that the theoretical linear power spectrum, that started out as $P(k) \propto k^{n_{\rm s}}$, changed shape on scales $\lambda < r_{\rm H,eq}$ during radiation domination. Consequently, after the radiation-dominated era the power spectrum roughly looked as follows:

$$P(k,t) \propto \begin{cases} k^{n_{\rm s}-4} & \text{for } \lambda < \lambda_{\rm eq} \\ k^{n_{\rm s}} & \text{for } \lambda > \lambda_{\rm eq}. \end{cases} \tag{1.9}$$

We show the detailed power spectrum of linear fluctuations in Figure 1.3. The exact shape and features of this power spectrum depend on all the cosmological parameters of the standard ΛCDM model, most of which we have already mentioned: besides $\Omega_{\rm m,0}$, $\Omega_{\Lambda,0}$, $\Omega_{\gamma,0}$, $\Omega_{\rm k,0}$, H_0 and $n_{\rm s}$, these are $\Omega_{\rm b,0}$ (the baryonic matter content of the Universe) and σ_8 (the normalisation of the power spectrum). As all of these influence the power spectrum independently, *every one of these* can be determined just by measuring the linear power spectrum to very high precision. This is essentially what we try to do when observing the CMB, which makes it the single most powerful observable for understanding our Universe as a whole.

Note that up until shortly before recombination, *baryonic* subhorizon fluctu-

are the only constituents of consequence.

[9] This approximation is derived by considering the gravitational evolution of a linear fluctuation in an expanding Universe with mean density $\bar{\rho}$, which is not trivial.

[10] Which, in turn, is much slower than the exponential growth in a non-expanding Universe.

[11] The unit shown here for $r_{\rm H,eq}$ is the typical unit of distance used in cosmology. "Mpc" is shorthand for "megaparsec", i.e. one million parsecs (a bit over three million light years), while "h" is the dimensionless Hubble constant, defined as $h \equiv H_0/[100\,({\rm km/s})/{\rm Mpc}] \approx 0.7$.

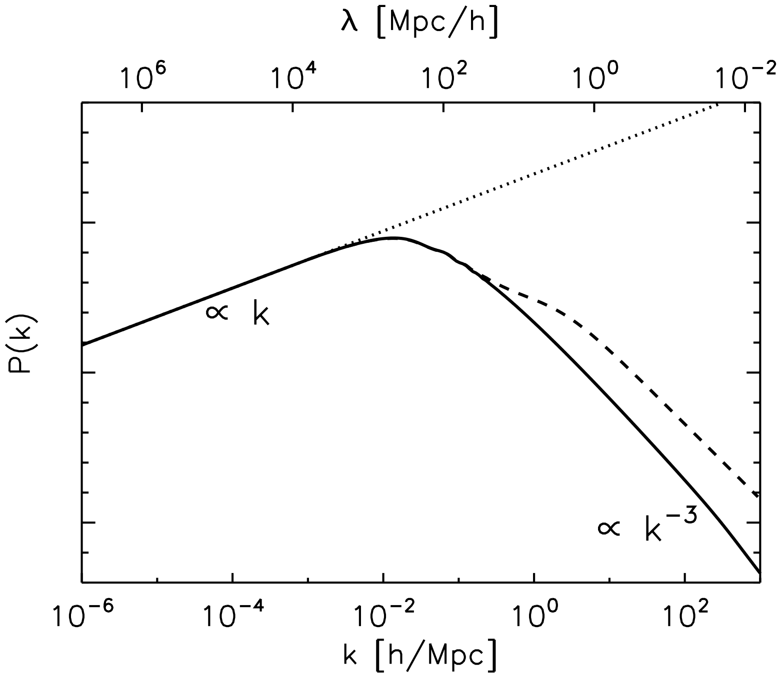

FIGURE 1.3: The theoretical matter power spectrum. The wavelength of fluctuations decreases towards the right. A dotted line shows the corresponding primordial power spectrum for $n_s \approx 1$. At roughly the horizon scale at matter-radiation equality, $r_{H,eq} \sim 10^2\,h^{-1}\,\mathrm{Mpc}$, the power spectrum turns over and the power law index asymptotes to $n_s - 4 \approx -3$. The dashed line shows a correction for non-linear growth at later times, which only affects scales of a few tens of Mpc or less.

ations were unable to grow, even though the matter-dominated era had already begun. This is because photons were capable of dragging baryons along, damping their fluctuations.[12] Therefore, up until the time of the CMB only fluctuations made up of cold dark matter (which is the dominant form of matter in our Universe) were able to grow. Afterwards, when the baryons could collapse, they followed the dark matter perturbations that were already present. The distribution of dark matter therefore dictated where stars and galaxies would form.

1.2.2 Non-linear evolution

As we mentioned before, a successful theory need not (and, within reason, cannot) predict the exact distribution of matter around us in absolute terms. Rather, it

[12]Related to this are the baryon acoustic oscillations (BAO). Gravity and pressure forces (the latter mainly caused by the photons) counteracted one another, causing oscillations in the baryonic fluctuations. These can be seen as small wiggles in the matter power spectrum around $100\,h^{-1}$ Mpc (see Figure 1.3) and are still imprinted on structure today.

predicts how matter is organised in a statistical sense – for example by predicting its power spectrum. Up until now, we only considered what would happen to linear density fluctuations, i.e. fluctuations that are very small ($\delta \ll 1$). However, if we want to predict the clustering of matter not only in the very early Universe, but also today, we need to consider what happens to fluctuations that grow large enough to actually collapse, for which the simplified picture sketched above no longer applies. Without a theory that accurately predicts the amount of non-linear structure as a function of scale in the Universe today, we cannot test our model against observations.

Many useful insights can be gained from taking a perturbative analytical approach to non-linear structure formation. For example, it grants us expressions for the time it takes for a halo to form, the relative density at which it forms, and its final size and mass.[13] However, since non-linear collapse is such a complex process – even when only considering dark matter – clustering predictions nowadays are mainly made by fitting to the results of simulations. These predictions are then compared to clustering measurements from observations in order to learn more about the underlying cosmology.

The general picture of non-linear structure formation looks as follows. As fluctuations grow, they will generally not be spherically symmetric; consequently, they will collapse first along one direction, forming sheets of matter (also called pancakes, see e.g. Zel'dovich, 1970). It is around this time that fluctuations will enter the non-linear regime, meaning the approximations used in the previous subsection are no longer valid. These sheets will collapse along a second direction, forming filaments, which finally collapse to make what we call haloes, forming a cosmic web of filaments with very massive haloes at the nodes (hosting galaxy groups and clusters) and smaller ones throughout. The dark matter collapse then stops as these haloes virialise, meaning that they attain a quasi-static dynamic equilibrium between the internal gravitational forces and the random motions of their particles. Further growth then proceeds through the merging of haloes, especially in clustered environments where the probability of two haloes encountering one another is large.

This non-linear evolution has important consequences for the clustering of matter, as shown by the dashed line in Figure 1.3. It is the haloes that are of most interest to us, as these are the regions where galaxies form.

1.2.3 The role of galaxy formation

Haloes are the highest-density regions of dark matter, which constitute the potential wells into which gas flows. Contrary to dark matter, gas feels pressure and can radiate its energy away as photons, allowing it to cool to the centres of haloes and form stars and galaxies, which merge and grow and evolve.

[13] A formalism generally referred to as (extended) Press-Schechter theory, see e.g. Press & Schechter (1974), Bond et al. (1991) and Sheth, Mo & Tormen (2001).

When we look out into the night sky, we do not see the majority of matter, which is in the shape of filaments and haloes. Instead, we see only the very peaks of the matter distribution, as this is where the galaxies reside (see bottom row of Figure 1.4). Galaxies are therefore biased tracers of the cosmic density field[14], which means that we must understand the complicated physical processes through which galaxies formed and evolved in order to be able to use them to derive cosmological information. The better we understand the galaxy bias, the better we can constrain the large-scale properties of our Universe by measuring how galaxies cluster.

Galaxies are not just important to the structure of the Universe because they are biased tracers; they influence the clustering of (dark) matter as well, which we can measure through the gravitational effect of matter on light (called lensing). Through gas cooling, baryons can attain much higher densities than dark matter, and form more structure on galactic scales than dark matter alone could. The dark matter haloes respond to the formation of galaxies in their centres by contracting somewhat, thereby changing the amount of structure on small scales (e.g. Blumenthal et al., 1986). This needs to be taken into consideration when one tries to predict the clustering of matter based on the relatively simple dark matter only picture of the Universe. Even though dark matter is dominant and baryons trace it initially, they act differently on small scales.

However, galaxy formation is not only more complex, but also more violent than the formation of dark matter haloes. For example, when stars die they may explode as supernovae, potentially heating up large amounts of gas, which prevents this gas from forming stars. Together, supernovae in a galaxy may drive galactic fountains of gas, ejecting the gas out of the galaxy. For small enough galaxies (occupying low-mass haloes), supernovae may even destroy the galaxy altogether. Very massive galaxies may host an active galactic nucleus (AGN) in their centre, heating mass amounts of gas and ejecting it far out of the galaxy. Because of these feedback processes, the pressure of the gas is increased and it will resist forming structure, meaning that the clustering of matter is lower than what would be expected from the simple dark matter picture. If enough gas is driven out of the galaxy, the dark matter haloes may respond in a way opposite to that we would naively expect: expanding on super-galactic scales (e.g. Velliscig et al., 2014; also see Chapters 2 and 3 of this thesis). This can even occur without feedback, due to pressure smoothing of virialised gas on large scales. Many other physical processes involved in the formation and evolution of galaxies may also influence clustering predictions. Currently, clustering measurements are becoming so precise that the need to understand the physics of galaxy formation to comparable accuracy is rapidly increasing.[15] Without it, we cannot test our theoretical

[14] Galaxies are often seen as biased tracers of haloes, which are in turn biased tracers of the entire matter distribution.

[15] Even if we do not fully understand the physics of galaxy formation, we may still be able to self-calibrate our models or marginalise over parameters that describe the effects of e.g. feedback on halo profiles (see e.g. Zentner, Rudd & Hu, 2008; Yang et al., 2013; Zentner et al., 2013).

models of cosmology against observations in a meaningful way.

1.3 Numerical simulations

Because of the complexity and immense range of scales involved in the formation of galaxies, numerical simulations are the only way to test our models at the precision of current and upcoming observations.

Different approaches to cosmological simulations are available. For example, one could run a simulation in which one assumes all matter acts like dark matter (N-body or collisionless simulations, see Chapter 5), making it possible to simulate a given region at a far higher resolution than otherwise possible, and base the formation and evolution of the baryonic component of the Universe on these, generally assuming the dark matter is not affected by the baryons. By solving sets of coupled equations for the evolution of baryons and galaxies and using observations as constraints for the parameters involved, one can very quickly obtain predictions for other quantities on a large range of scales. This is the approach taken by semi-analytic models of galaxy formation (see Baugh, 2006, for a review on the methodology; also, see Chapters 4 and 6 of this thesis).

Another way is to include the baryons in the simulation directly along with the dark matter, solving the gravitational and other physical equations involved simultaneously (hydrodynamical simulations, see Tormen, 1996, for an introduction and Springel, 2010, for a review on the method employed here; also, see Chapters 2 and 3 of this thesis). Because there are more complicated equations to solve, because there are more variables to track and because there are higher densities involved (decreasing the time steps), such simulations are often run at much lower resolution than pure N-body simulations in order to keep the computational time and memory consumption down. However, the trade-off is that fewer approximations have to be made and that the effects of the baryons on the dark matter are modelled explicitly. Since not all the relevant scales can be resolved (yet), the physics of e.g. star formation and supernova feedback have to be modelled in a comparable way to semi-analytics, which brings some uncertainties with it. It is here that much may be gained in coming years, as these physical recipes in both semi-analytical and hydrodynamical simulations are constantly being improved, leading to more realistic representations of our Universe.

The matter distribution in a hydrodynamical simulation from the OWLS project (Schaye et al., 2010) referred to often in this thesis, called *AGN_L100N512* (see Chapter 2), is shown in Figure 1.4. The region shown is $100\,h^{-1}$ Mpc on a side and $10\,h^{-1}$ Mpc thick. The distributions of cold dark matter, gas and stars are shown in separate columns, and from top to bottom cosmic time increases. For $z = 127$, the fluctuations are still linear, and no stars have formed yet. At $z = 6$,

However, this requires us to model the way baryons affect clustering in some way, and may come at the cost of decreasing the statistical significance of the measurements.

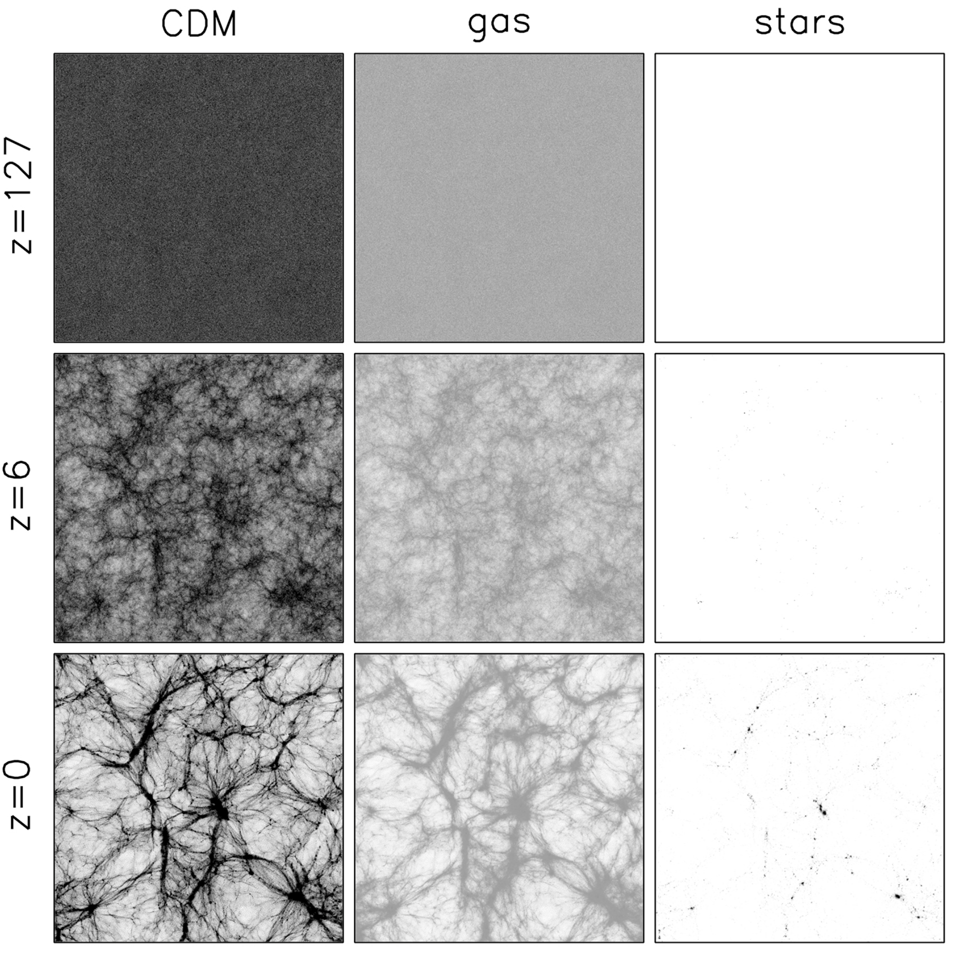

FIGURE 1.4: This figure illustrates the growth of structure from a linear to a highly non-linear state, for a comoving (meaning that we scale out the expansion of the Universe) volume $100\,h^{-1}$ Mpc on a side. The slice shown here, showing the projected mass density, is $10\,h^{-1}$ Mpc thick. Each column shows the evolution of a different component: from left to right, these are cold dark matter, gas, and stars. Each row shows the volume at a different cosmic time. At the starting redshift of the simulation, $z = 127$ (only 12 million years after the Big Bang), no significant structure has formed yet, and density fluctuations are still very small. At $z = 6$ (almost a billion years after the Big Bang), the dark matter is clearly collapsing and starting to form a cosmic web. Gas still traces the dark matter on the scales visible here. Galaxies, visible as small clumps of stars, have started forming at the points of highest density relatively recently. Finally, at $z = 0$ (the present, 13.8 billion years after the Big Bang), all the structure we see today has formed. Note that the gas no longer perfectly traces the dark matter, but is distributed somewhat more smoothly. This is mainly caused by energetic feedback processes associated with galaxy formation, heating the gas. The galaxies themselves, seen in the right-most panel, are clearly biased tracers of the overall mass distribution, having formed where the dark matter densities are highest.

the cosmic web has started to take shape and galaxies have formed in the more massive haloes. The gas still traces the dark matter almost perfectly on large scales. By $z = 0$ (the present time), the cosmic web is more pronounced and all the galaxies we see today have formed, which trace the large-scale structure of the cold dark matter. The gas has been heated by gravitational accretion shocks and by feedback from both supernovae and AGN, and is distributed somewhat more smoothly than the dark matter.

1.4 This thesis

In the near future it will be possible to measure the distribution of galaxies and matter to unprecedented precision. To get the most out of these observations and to avoid unwanted biases, our theoretical models will have to match the accuracy of real-life measurements. The fields of cosmology and galaxy formation are now more tightly tied together than ever before: we need to understand the processes involved in galaxy formation to interpret the clustering of matter and tie our observations to a set of cosmological parameters. Additionally, small-scale clustering measurements – which are less sensitive to cosmology – may help us to constrain our galaxy formation models. We explore all these topics in this thesis.

In **Chapter 2** we investigate the effects of galaxy formation on the clustering of matter through the use of the OWLS suite of simulations (Schaye et al., 2010; Le Brun et al., 2014), in which different physical processes were varied one at a time. We compare the results of hydrodynamical simulations to those of dark matter only models, which are generally used to interpret weak lensing measurements of the matter distribution, and show that feedback from galaxy formation can have much larger effects on the matter power spectrum than previous studies have shown. We also investigate how the clustering of dark matter changes when such processes are included.

Since the clustering of galaxies and the galaxy-galaxy lensing signal may be similarly affected, we also examine the two-point galaxy correlation function and the galaxy-matter cross-correlation in these simulations, in **Chapter 3**. We will show that efficient feedback can change the predictions by $\sim 10\%$, and although this shift is mainly due to the masses of both galaxies and haloes being systematically lowered, significant residual effects remain after correcting for the change in mass.

Next, we explore the validity and consequences of several assumptions that are typically used in models based on the halo model and halo occupation distribution. Specifically, in **Chapter 4** we investigate if and how the shapes and alignments of haloes are reflected in the clustering of galaxies, using the Guo et al. (2011) semi-analytical model run on the Millennium Simulation (Springel et al., 2005). We also ask the question whether it is possible to measure this form of "assembly bias" from galaxy surveys, without knowing the distribution of dark matter. In **Chapter 5** we test the postulate of halo models that all matter resides in haloes, using collisionless simulations from the OWLS project. We calculate the clustering of matter in our simulations and compare it to the clustering of matter within haloes above a certain mass, exploring also the effects of using different halo definitions.

Finally, in **Chapter 6**, we present a fast and accurate clustering estimator for use in semi-analytical models of galaxy formation. Using this halo model based estimator, it is possible to predict the projected galaxy correlation function to an accuracy of $\sim 10\%$ using only a very small subsample of haloes, meaning it can be used efficiently while exploring the parameter space of a model. By using clustering data as a constraint in addition to the usual one-point functions (such as the stellar mass or luminosity functions), degeneracies can be removed, improving both the match of the model to multiple data sets at once and our understanding of galaxy formation. We apply our model to the semi-analytical model of Guo et al. (2013), and show how the best-fit parameters change to bring the model into agreement with the newly-added constraints.