Chapter 1

Introduction

1.1 Basic elements of cosmology

The complex distribution of galaxies, groups and (super-)clusters that constitutes the large-scale structure of the universe originated from small, seed fluctuations in the cosmic density field, as evidenced by the minute anisotropies observed in the cosmic microwave background (e.g. Bennett et al. 2003). These primordial fluctuations can be traced back to a (near-)Gaussian density field, consistent with the theory of inflation (Guth 1981) that predicts that the universe expanded over many orders of magnitude in size within a tiny time interval shortly after the “Big Bang”, about 14 billions years ago. The geometry of the universe is believed to be dominated not by matter, but by “dark energy” (expressed by a cosmological constant, \( \Lambda \)) which has driven the cosmological expansion to accelerate over relatively recent times (see Carroll et al. 1992). This is confirmed by the latest studies of distant supernovae (Perlmutter et al. 1999; Riess et al. 2004). The power spectrum of the microwave background, combined with information on the spatial distribution of galaxies obtained from large galaxy redshift surveys, have placed important constraints on the cosmic matter budget, the density distribution at recombination \( (z \approx 1000) \), and the biasing between galaxies and the dark matter (e.g. Lahav et al. 2002; Tegmark et al. 2004). The currently favoured model for structure formation is the so-called cold dark matter model (“\( \Lambda \)CDM”), which explains how bound objects such as galaxies and clusters formed on increasingly larger spatial scales due to the growth of density fluctuations from gravitational instability in an expanding universe. According to such models, the universe became (re-)ionized by the first light of stars, protogalaxies and/or quasars at \( z \approx 7 - 20 \) (e.g. Loeb & Barkana 2001; Fan et al. 2002; Kogut et al. 2003). The best current observational evidence for galaxies and quasars in the early universe is found at about 1 billion years after the Big Bang \( (z \approx 6 - 7) \), confirming that the first galaxies and quasars developed within a couple of hundred million years (e.g. Kodaira et al. 2003; Kneib et al. 2004; Eyles et al. 2005; Bouwens et al. 2006).

1.2 Structure formation

The theoretical framework that is used to describe the formation of the large-scale structure and the galaxies it contains is based on the following two tenets (White & Rees 1978):

(i) the distribution of the dominant mass component at all scales developed purely from gravitational clustering through the collapse and merging of dark matter halos (PS; Press & Schechter 1974). The resulting hierarchy of formed structures is a consequence of the fact that most of the power of the initial mass fluctuations was at small scales.

(ii) galaxy formation is driven by the dissipation and collapse of gas in the cores of dark matter halos.

This theoretical framework aims at providing a detailed explanation of the most fundamental issues in cosmology concerning the formation and evolution of galaxies and the large-scale structure.
Figure 1.1 — Predictions for dark matter halo abundances in semi-analytical models and N-body simulations. **Left panel:** Correction factors for the halo abundance as a function of mass to be applied to the extended PS formalism as derived by Sheth & Tormen (1999) (from Somerville & Primack 1999). **Right panel:** Halo abundance as a function of redshift, plotted as the mass fraction in objects of mass $M$. Points are from the Millennium Run N-body simulation (Springel et al. 2005). Solid lines are the analytic fitting function from Jenkins et al. (2001). Dotted lines give the PS model predictions at the minimum and maximum redshift (taken from Springel et al. 2005).

For example, which of the galaxies observed at high redshift are the progenitors of local galaxy populations and how and when did they form? Which of the local galaxies host the remnant black holes that once powered high redshift active galactic nuclei (AGN)? What are the effects of AGN on the star formation history? How and when did clusters of galaxies form, and what role does galaxy environment play in the evolution of the galaxies themselves? The formation and evolution of structure in the universe can, in principle, be derived analytically from the power spectrum of density fluctuations, but not beyond the point where linear theory breaks down due to collapse and merging of structures that distorts the form of the fluctuation spectrum. The non-linear regime can be studied using semi-analytical approximations, or experimentally through numerical N-body simulations. The semi-analytic approach (White & Frenk 1991, see also Somerville & Primack 1999 for a review) is used to study the complicated feedback loops that exist between gas cooling, star formation, merging, supernovae, AGN, dust extinction, etc. The semi-analytical prescriptions for star and galaxy formation are applied to halo merger trees based on the PS formalism. The inherent hierarchical nature of the merger-tree formalism is in contrast with the dissipative monolithic collapse model in which galaxy assembly is sudden rather than gradual (Eggen et al. 1962). There is, however, mounting evidence that at least certain types of galaxies may have formed in short, massive starbursts at very high redshifts (e.g. Mobasher et al. 2005). The PS formalism was extended to relate halos of a given mass at a specific redshift, to the masses of their progenitor halos prior to that redshift (Bond et al. 1991; Bower 1991). A later modification was proposed by Sheth & Tormen (1999) in order to account for a discrepancy between the original PS formalism and results from numerical simulations leading to overpredictions in the halo abundances at small scales and underpredictions at large scales (Fig. 1.1).

The semi-analytical approach has been successful in reproducing some of the key local observables (e.g. luminosity functions, Tully-Fisher relation, sizes, colours, metallicities), but reproducing all of those simultaneously remains challenging (e.g. Kauffmann et al. 1999; Cole et al. 2000). It has been
suggested that a powerful test to discern between passively evolving galaxies that formed in a relatively early burst of short duration and galaxies that formed later through hierarchical processes, lies in the study of the shape of the K-band luminosity function at \( z \sim 1 \) (Kauffmann & Charlot 1998). Such a test was recently performed by Somerville et al. (2004) who concluded that although the hierarchical model provides a better match to the data than the monolithic model at least qualitatively, there is a quantitative disagreement between the number of (red) galaxies observed at \( z \gtrsim 1.5 \) and the number expected from the model. A better knowledge of galaxies at \( 1.5 < z < 2.5 \) is of crucial importance for providing better constraints for either model.

Recent results from large N-body simulations coupled with semi-analytical post-processing have provided numerous predictions for the properties of the large-scale structure and galaxies over a wide redshift range (see right panel of Fig. 1.1; Springel et al. 2005, and references therein). These large \( \Lambda \)CDM simulations will form a navigable frame of reference for current and future observational cosmologists. For example, it is predicted that quasars at \( z \sim 6 \) lie in the center of very massive dark matter halos of \( \sim 10^{12} M_\odot \) surrounded by many fainter galaxies, and that these halos will evolve into massive clusters of \( \sim 10^{15} M_\odot \) at \( z = 0 \) (Springel et al. 2005). The simulations indicate that the quasar progenitors form at \( z \sim 17 \). The simulations are also successful in reproducing the clustering of galaxies observed in local redshift surveys, and furthermore they predict that ‘baryon wiggles’, reflecting the acoustic oscillations in the mass power spectrum of the \( \Lambda \)CDM model, should be imprinted on the galaxy distribution out to \( z \sim 3 \). These wiggles may be used to constrain the nature of the dark energy (Blake & Wall 2002b). The power spectrum analyses of large galaxy redshift surveys (the Two-degree Galaxy Redshift Survey and the Sloan Digital Sky Survey) have recently shown the baryonic oscillations to be present in the distribution of galaxies in the local universe (Eisenstein et al. 2005; Cole et al. 2005).

1.3 Recent advances in the study of clusters of galaxies

Despite the rapid advancements in our understanding of the evolution of galaxies concerning e.g. their star formation histories, morphologies and clustering at \( z \lesssim 6 \) (e.g. Papovich et al. 2004; Bouwens et al. 2004a, 2006; Giavalisco et al. 2004a; Franx et al. 2003; Barmby et al. 2004; Ferguson et al. 2004; Lotz et al. 2004, 2005; Conselice 2003; Ouchi et al. 2004; Porciani & Giavalisco 2002; Lee et al. 2005), the study of the evolution of clusters of galaxies has progressed at a much slower pace. Clusters of galaxies are the most massive and largest structures in the universe, that are the product of billions of years of gravitational growth. They typically have masses of \( 10^{14-15} M_\odot \), and contain hundreds of galaxies over a region of several Mpc across. Because mass condensations on these scales are extremely rare (Kaiser 1984), it is believed that they formed relatively late in the history of the universe and are thus even more rare at high redshifts (e.g. see Mo & White 2002). Also, clusters and their progenitors (‘protoclusters’\(^1\)) are much harder to distinguish at high redshift because the density contrast between clusters and the field decreases with lookback time, and thus the study of cluster formation is a very challenging field.

The most distant clusters known, at roughly half the Hubble time (\( z \sim 1 \)), contain significant populations of relatively old galaxies, as well as younger star-forming galaxies (e.g. Dressler et al. 1999; van Dokkum et al. 2000; Demarco et al. 2005; Goto et al. 2005; Homeier et al. 2005; White et al. 2005; Stanford et al. 2005). Clusters have a ‘red sequence’, or colour-magnitude relation (CMR) consisting of massive, predominantly early-type galaxies (see Fig. 1.2). Its slope is believed to originate from a mass-metallicity relation, in the sense that more massive galaxies are better able to retain their metals and therefore appear redder than less massive galaxies (see Kodama & Arimoto and refer-

\(^1\)See chapter 10 for a working definition.


**Figure 1.2** — One of the most distant and massive clusters known: Cl1252–2927 at $z = 1.24$. Left: ACS/WFC $i_{775}+z_{850}$-band image. Right: Color-magnitude diagram of early-type galaxies within the central 2’ of the cluster, with elliptical galaxies indicated by (filled) circles and S0 galaxies by squares. The lines indicate fits to the CMR for ellipticals only (solid line) and for all early-types (dashed). The dot-dashed line represents the relation for the Coma Cluster, transformed to these bandpasses at $z = 1.24$ assuming no evolution. The figures were taken from Blakeslee et al. (2003).

ences therein). Interestingly, some parts of clusters are remarkably old even at $z \sim 1$ as evidenced by the following. First, the tight scatter in the CMR for cluster early-types suggests formation redshifts of $z_f > 2.5$ with average luminosity-weighted ages of 2–4 Gyr (e.g. Ellis et al. 1997; Stanford et al. 1998; Stanford et al. 2005; Blakeslee et al. 2003; Mei et al. 2006). Second, analysis of the cluster morphology-density relation with HST suggests that while the relative fractions of S0s and spirals in clusters evolve strongly from $z \approx 1$ to $z \approx 0$, the elliptical fraction exhibits no significant evolution (Postman et al. 2005). It has been suggested that an excess of spiral galaxies observed at $z \sim 1$ can account for the deficit of S0 galaxies through extensive periods of merging at $z \lesssim 0.5$. Third, although many high redshift clusters ($z \gtrsim 0.5$) show signs of substructure and filamentary, diffuse X-ray emission, all clusters have hot virialized cores (core radii of a few hundred kpc) which are also traced by the distribution of the rest-frame optical stellar light, as well as the mass distribution determined from weak-lensing analysis (e.g. Jee et al. 2005a,b).

The most distant clusters currently known have been found mostly from wide-field surveys at optical (e.g. Gladders & Yee 2005) or X-ray wavelengths (e.g. Rosati et al. 1998; Mullis et al. 2005). The optical surveys can target high redshift clusters quite efficiently by looking for a cluster red sequence, provided that enough area is being surveyed to sufficient depth. However, as the 4000Å break shifts towards the near-infrared for $z \gtrsim 1.2$ the selection becomes more and more difficult. Recently, Stanford et al. (2005) discovered the second highest redshift cluster ($z = 1.41$) yet detected from a concentration of objects with high photometric redshifts. The survey data used in this detection comes from the infrared Spitzer Space Telescope, indicating the versatility of this new space telescope. The ‘discovery power’ of X-ray surveys becomes increasingly less with redshift, since the X-ray surface brightness profile of clusters is proportional to $(1 + z)^{-4}$. Even worse, structure formation predicts that the X-ray signature of thermal gas only becomes apparent during the epoch of cluster virialization, which is believed to have occurred by $z \sim 1–1.5$ for the most massive clusters currently known. In the very near future, the search and study of clusters is likely to experience a significant revolution through sensitive, large area surveys targeting the Sunyaev-Z’eldovich (SZ)
signature of hot intracluster gas that scatters the photons of the CMB and modifies the incident CMB spectrum. The strength of the SZ effect is redshift independent and relies only on the presence of a hot medium above a certain mass threshold that is limited only by the sensitivity of the instruments. Large samples of galaxy groups and clusters are expected to be discovered at both low and high redshift (e.g. Larouque 2003, see Carlstrom et al. 2002 for a review) that could significantly advance the study of (massive) structure formation.

1.4 Searches and discoveries of forming clusters in the early universe

Given the relatively old ages of the stellar populations of elliptical galaxies in massive galaxy clusters at a distance of roughly half the Hubble time \((z \sim 1)\), an interesting epoch of cluster formation and evolution could lie at even higher redshifts. Are such structures partially virialized at \(z > 2\), and do they consist of several galaxy groups in sub-halos that eventually merge and give rise to a cluster red sequence? Do these structures lie at the nodes of filaments in the large scale structure as observed in N-body simulations? In any case, the progenitors of clusters must possess intense star formation. This star formation and possibly enhanced AGN activity at \(z \gtrsim 2\) could be responsible for `pre-heating' (Tozzi & Norman 2001) as well as chemical enrichment of the intra-cluster medium (e.g. Arnaud et al. 1992; Tozzi et al. 2003; Maoz & Gal-Yam 2004; Ettori 2005). Simulations predict that the ages of galaxies depend strongly on environment, with ellipticals in the most massive halos being \(> 1\) Gyr older on average. Up to 50\% of the stellar mass in these galaxies is presumably formed by \(z \sim 4\), although it was only assembled into a single galaxy at \(z \sim 1\) (De Lucia et al. 2006).

Are we already witnessing some of these phenomena at \(z \gtrsim 2\)? Overdensities of galaxies have been discovered out to \(z \approx 6\) (e.g. Pascarelle et al. 1996; Steidel et al. 1998, 2005; Keel et al. 1999; Kurk et al. 2000; Pentericci et al. 2000; Francis et al. 2001; Möller & Fynbo 2001; Venemans et al. 2002, 2004; Venemans et al. 2005; Shimasaku et al. 2003; Kurk et al. 2004; Ouchi et al. 2005; Stiavelli et al. 2005). These structures are all overdense compared to the field, but their derived physical properties are generally highly uncertain. These objects have estimated galaxy overdensities close to or in excess of the requirements for gravitational collapse, group- or clusterlike masses of \(10^{13} - 15\) \(M_\odot\), projected sizes of several to tens of comoving Mpc, and in some cases measured velocity dispersions of several 100 km s\(^{-1}\) determined from emission line galaxies. Their topologies and masses indicate that they may constitute `filaments', `sheets', `proto-groups' and `protoclusters'. While some of these structures have been found as by-products of wide field surveys using broad or narrow band imaging and spectroscopic follow-up, in other instances, these large-scale structures were traced by a luminous or powerful radio galaxy or quasar that facilitated in pinpointing the overdense region. In the following section, we will focus on the latter group, since it forms the basis of the current thesis.

1.4.1 Probing the emergence of cosmic structures around distant radio galaxies

Luminous radio galaxies are amongst the most massive forming galaxies at high redshift \((z \gtrsim 1)\). They form a bright envelope in the K-band Hubble redshift diagram (De Breuck et al. 2002; Rocca-Volmerange et al. 2004), indicating that radio galaxy hosts statistically have baryonic masses of up to \(10^{12} M_\odot\) over a wide redshift range. This is confirmed by detailed studies of individual objects. Pentericci et al. (2001) and Zirm et al. (2003) measured the stellar hosts of radio galaxies at \(1 < z < 3\), finding that they often possess \(r^{1/4}\)-law light profiles, indicating that they are massive, passively evolving elliptical galaxies. Villar-Martín et al. (2006) presented the first rest-frame optical-near-infrared spectral energy distribution of a radio galaxy at \(z = 2.5\), which suggested the existence of a reddened, evolved stellar population of age \(> 1.8\) Gyr \((z_f \gtrsim 6)\) and mass \((5 \pm 2) \times 10^{11} M_\odot\). The galaxy itself is embedded in a giant (> 100 kpc), perhaps rotating, nebula of emission line gas which
is a common feature in high redshift radio galaxies (e.g. van Ojik et al. 1997; Venemans et al. 2002; Reuland et al. 2003). Some radio galaxies show signs of vigorous starbursts with star formation rates as high as several hundreds to a thousand $M_\odot$ yr$^{-1}$ (e.g. Dey et al. 1997; Papadopoulos et al. 2000; Stevens et al. 2003; Zirm et al. 2005). An example of the host galaxy of radio source MRC 1138–262 at $z = 2.16$ is shown in Fig. 1.3, showing several characteristic features of radio galaxies at rest-frame UV wavelengths (see figure caption for further details). From the study of the clustering properties of radio galaxies it has been found that they are associated with the densest environments that may be virializing at any cosmic epoch (e.g. Negrello et al. 2006, and references therein).

The theory of structure formation predicts that, in principle, the most massive galaxies at any epoch are associated with the most extreme peaks in the large-scale structure. Could distant radio galaxies trace the progenitors of the galaxy clusters seen in the local universe? If so, this would provide us with a unique tool for studying cluster formation in the early universe. Over the past decade, this hypothesis has been tested by searching for companion galaxies in the vicinity of radio sources. Many radio galaxies in the redshift range $1.5 < z < 2$ have been found to be associated with overdensities of relatively red galaxies (e.g. Sánchez & González-Serrano 1999, 2002; Thompson et al. 2000; Hall et al. 2001; Barr et al. 2003; Best et al. 2003; Wold et al. 2003; Bornancini et al. 2006), suggesting clusters of Abell richness class 0–1. In a pioneering study by Pentericci et al. (2000) and Kurk et al. (2000), a narrow-band filter was used to search for Ly$\alpha$ companion objects in a $7' \times 7'$ field around a radio galaxy at $z = 2.16$ using the Very Large Telescope (VLT) in Chile. Follow-up spectroscopy of a large number of candidate Ly$\alpha$ excess galaxies, augmented by a sample of near-infrared selected candidate H$\alpha$ excess galaxies (Kurk et al. 2004) showed a large structure of galaxies within 1000 km s$^{-1}$ of the radio galaxy. The mass of the system is $\sim 10^{14}$ $M_\odot$, estimated from the overdensity of Ly$\alpha$ emitters relative to the field. The radio galaxy protocluster program was expanded by means of a Large Program with the VLT to carry out similar studies towards other luminous radio galaxies in the redshift range $2.0 < z < 5.2$. The program resulted in the discovery of six new structures of Ly$\alpha$ galaxies with masses in the range of $10^{14}$–$15$ $M_\odot$ (Venemans et al. 2002, 2004; Venemans et al. 2005, 2006). The velocity dispersions of the systems decrease with increasing redshift, roughly as predicted for forming clusters by numerical simulations (Venemans et al. 2006). The work of Venemans et al. has provided strong evidence that distant radio galaxies are tracers of rich environments in the early universe.

1.4.2 The ACS high redshift cluster/protocluster survey

The Advanced Camera for Surveys (ACS; Ford et al. 1998) on the Hubble Space Telescope (HST) is a unique instrument for studying galaxy and cluster evolution, due to its unprecedented sensitivity, relatively wide field of view ($3.4' \times 3.4'$), and high spatial resolution ($\sim 0'1$). This has motivated an extensive program to study the properties of massive galaxy clusters in the early universe, using broad-band imaging with the HST/ACS Wide Field Channel (WFC), as part of the Guaranteed Time Observations (GTO). The survey covers a large number of X-ray selected clusters in the redshift range $0.8 < z < 1.4$ (Blakeslee et al. 2003; Mei et al. 2006; Holden et al. 2005; Postman et al. 2005; Goto et al. 2005; Jee et al. 2005a,b; Homeier et al. 2005, 2006).

The cluster program is complemented by the study of protocluster candidates around high redshift ($z > 2$) radio galaxies and quasars (Miley et al. 2004; Overzier et al. 2006a,b; Zheng et al. 2006; Zirm et al. 2005), the results of which are presented in this thesis.
Figure 1.3 — ACS/WFC g275+i014 detail image of the radio galaxy MRC 1138–262 at \( z = 2.16 \). The system consists of a large (> 100 kpc) conglomeration of sub-galactic clumps embedded in a region of diffuse emission. The total star formation rate in the clumps and the diffuse component are about equal. Comparison with rest-frame optical light indicates that the mass of the UV continuum component seen in this image is only a small fraction of the total mass of the system, suggesting that the MRC 1138–262 system has elements of both monolithic and hierarchical formation (Miley et al., in prep.). Contours show the radio emission at 4.5 GHz.

1.5 This thesis

This thesis presents additional evidence for associations of star-forming galaxies that may have developed early on in the history of the universe, especially in the close vicinity of luminous active galactic nuclei (radio galaxies and quasars). These structures form an interesting class of relatively rare and massive objects. A systematic study of these structures may shed light on the origin of galaxy clusters. Here, I analyse the environments of radio galaxies at large and small scales, and study the morphological and spectral properties of several galaxy overdensities between \( z = 2 \) and \( z = 6 \) and investigate their relation to the epoch of cluster formation, and to the properties of the emerging large-scale structure in general. The structure of this thesis is as follows.

Chapter 2 – In this chapter an analysis of the angular correlation function, \( w(\theta) \), of radio sources in
the 1.4 GHz NVSS and FIRST radio surveys is presented. Below \( \sim 6' \) the signal is dominated by the size distribution of classical double radio galaxies. A high amplitude measured for the cosmological clustering suggests that powerful radio galaxies probe significantly more massive structures compared to normal galaxies, quasars as well as radio galaxies of average power. This is consistent with powerful radio galaxies being associated with massive galaxies in relatively rich environments at high redshift. Their clustering scalelength \( (r_0) \) at \( z \sim 1 \) is close to that measured for extremely red objects (EROs) associated with a population of old elliptical galaxies at similar redshifts, and we propose that EROs and radio galaxies may be the same systems seen at different evolutionary stages. Depending on the underlying model for their evolution from \( z \sim 1 \) to \( z = 0 \), the clustering of radio galaxies could be in agreement with both \( \Lambda \) CDM hierarchical predictions for massive early-type galaxies, and with passive evolution into a present-day population of clusters.

Chapter 3 – This chapter reports of the use of the Chandra X-ray observatory to study 5 radio galaxies at \( z \sim 2 - 3 \). The goals were to (i) study the nature of their non-thermal X-ray emission, (ii) investigate the presence of hot gas, and (iii) look for overdensities of active galaxies near high redshift radio galaxies. We detected unresolved X-ray components towards the radio nuclei, and their X-ray luminosities implied that the nuclei are surrounded by obscuring material with H\( \text{I} \) column densities of \( \sim 10^{22} \text{ cm}^{-2} \). We found extended emission coincident with the radio hotspots or lobes, which can be explained by the Inverse-Compton scattering of photons that make up the cosmic microwave background (CMB). The magnetic field strengths of \( \sim 100 - 200 \mu \text{G} \) that we derive agree with the equipartition magnetic field strengths. The relative ease with which the lobe X-ray emission is detected is a consequence of the \((1 + z)^4 \) increase in the energy density of the CMB. For one of the lobes, the X-ray emission could also be produced by a reservoir of hot, shocked gas.

We detected no diffuse emission and derive upper limits of \( \sim 10^{44} \text{ erg s}^{-1} \), thereby ruling out a virialized structure of cluster-size scale at \( z \sim 2 \). The average number of soft X-ray sources in the field surrounding the radio sources is consistent with the number density of AGN in the Chandra Deep Fields, and analysis of their angular distribution shows no evidence for rich large-scale structure associated with these radio galaxies, in contrast to what was found for the radio galaxy PKS 1138–262 at \( z = 2.2 \).

Chapters 4 & 5 – Here deep ACS and VLT \( K_S \)-band observations are presented of fields around the radio galaxy TN J1338–1942 at \( z = 4.1 \). We study in detail 12 spectroscopically confirmed companions previously found through their excess Ly\( \alpha \) emission by Venemans et al. (2002), and conclude that the Ly\( \alpha \) emitters (LAEs) are young (a few \( \times 10^7 \) yr), dust-free galaxies based on small sizes, steep UV slopes (\( \beta \approx -2 \)) and blue UV-optical colours with star formation rates (SFRs) of \(< 14 \ M_\odot \text{ yr}^{-1} \). We derive stellar masses of a few \( \times 10^8 \ M_\odot \), and estimate the LAE AGN fraction to be minimal.

We selected 66 Lyman break galaxies (LBGs) at \( z \sim 4.1 \) (`\( g_{475} \)-dropouts'), six of which are in the LAE sample. Their SFRs, sizes, morphologies, UV slope-magnitude and \((i_{775} - K_S) \) vs. \( K_S \) colour-magnitude relations are all similar to those found for LBGs in the 'field'. We quantify the number density and cosmic variance of \( z \sim 4 \) LBGs, and show that the field of TN J1338–1942 is richer than the average field at the \( 3 - 5\sigma \) significance. The angular distribution is highly filamentary, with about half of the objects clustered in a 4.4 arcmin\(^2\) region that includes the radio galaxy and the brightest LBGs. The generally fainter LAEs appear to favour regions that are devoid of LBGs, while LBGs detected in the rest-frame optical \( (K_S) \) tend to lie in the richest region. This may suggest a forming age- or mass-density relation. We find an excess signal in the angular correlation function at separations of \( \theta < 20'' \), corresponding to the typical halo size of dark matter halos hosting bright LBGs recently shown to be statistically present in large LBG field samples. The large galaxy overdensity, its corresponding mass overdensity and the sub-clustering at the approximate redshift of TN J1338–1942 suggest the
assemblage of a > 10^{14} M_\odot structure, possibly a ‘protocluster’.

**Chapter 6** – In this chapter we focus on the properties of the host galaxy of the radio source TN J1338–1942 at z = 4.1. TN J1338 is the dominant galaxy in the protocluster in terms of size and luminosity and therefore seems destined to remain the brightest cluster galaxy. The high spatial resolution ACS images reveal several kpc-scale features within and around the radio galaxy. The rest-frame continuum light is aligned with the radio axis and is resolved into two clumps with luminosities of \sim 10^9 L_\odot and sizes of a few kpc. The estimated nebular continuum, scattered light, synchrotron- and IC-scattering contributions to the aligned continuum light are only a few percent of the total observed flux which is likely dominated by forming stars with a star formation rate of \sim 200 M_\odot yr^{-1}. A simple model in which the jet has triggered the star formation is consistent with the available data. A small, linear feature in the z_h850 aligned light may be indicative of a large-scale shock associated with the advance of the radio jet. The rest of the aligned light also seems morphologically consistent with star formation induced by shocks associated with the radio source, as seen in other high-z radio galaxies. An unusual feature is seen in Ly\alpha emission. A wedge-shaped extension emanates from the radio galaxy perpendicularly to the radio axis. This “wedge” naturally connects to the surrounding asymmetric, large-scale (\sim 100 kpc) Ly\alpha halo. We posit that the wedge is a starburst-driven superwind. The shock and wedge are examples of feedback processes due to both the active galactic nucleus and star formation in the earliest stages of massive galaxy formation.

**Chapter 7** – Here ACS observations are presented, of the most distant radio galaxy known, TN J0924–2201 at z = 5.2. This radio galaxy has 6 spectroscopically confirmed Ly\alpha emitting companion galaxies, and appears to lie within an overdense region. Although the radio galaxy shows some continuum emission aligned with the radio axis, its basic properties (half-light radius and UV star formation rate) are comparable to the typical values found for Lyman break galaxies at z \sim 4 – 5. The Ly\alpha emitters are sub-L^* galaxies, with deduced star formation rates of 1 – 10 M_\odot yr^{-1}. One of the Ly\alpha emitters is only detected in Ly\alpha, and the lack of continuum emission could be explained if the galaxy is younger than \sim 2 Myr and is producing its first stars.

Observations in V_606\check{z}_{850} were used to identify additional LBGs associated with this structure. In addition to the radio galaxy, there are 22 V_606-break galaxies with z_{850} < 26.5, two of which are also in the spectroscopic sample. We compare the surface density of \sim 2 arcmin^{-2} to that of similarly selected V_606-dropouts extracted from the Great Observatories Origins Deep Survey (GOODS) and the Hubble Ultra Deep Field (UDF) parallel fields. We find evidence for an overdensity (> 99% confidence), based on a counts-in-cells analysis applied to the control field. The excess is suggestive of the V_606-break objects being associated with a forming structure around the radio galaxy.

**Chapter 8** – The angular clustering has been measured from a sample of 506 i_{775} dropout galaxies obtained from deep ACS fields to study clustering at z \sim 6. For our largest and most complete subsample (L \geq 0.5L^*_{\text{rest}}), we detected clustering at \sim 94% significance. We derived a (co-moving) spatial correlation length of r_0 = 3.6^{+1.3}_{-1.0} h_{72}^{-1} Mpc and bias b = 3.6^{+1.3}_{-1.2} using an accurate model for the redshift distribution. No clustering could be detected in the much deeper but significantly smaller UDF sample. We compare our findings to Lyman break galaxies at z \sim 3 – 5 at a fixed luminosity. Our best estimate of the bias parameter implies that i_{775} dropouts are hosted by dark matter halos having masses of \sim 10^{11} M_\odot, consistent with the typical mass of halos hosting V_606 dropouts at z \sim 5. We evaluate a recent claim by Lee et al. (2005) that at z \geq 5 star formation might have occurred more efficiently compared to that at z = 3 – 4. This may provide an explanation for the very mild evolution observed in the rest frame UV luminosity density between z = 6 and 3. Although our results are consistent with the star formation efficiency also being higher at z \sim 6, our errors are too large to find conclusive evidence for this.
Chapter 9 – A five square arcminute region around the luminous radio-loud quasar SDSS J0836+0054 (z = 5.8) hosts a wealth of associated galaxies, characterized by very red (1.3 < i775 - z850 < 2.0) colour. The surface density of these z ~ 5.8 candidates is approximately six times higher than the number expected from deep ACS fields (see chapter 8). We also find evidence for a substructure associated with one of the candidates. It has two very faint companion objects within two arcseconds, which are likely to merge. The finding supports the results of a recent simulation that luminous quasars at high redshifts lie on the most prominent dark-matter filaments and are surrounded by many fainter galaxies. The quasar and star formation activity from these regions may signal the build-up of a massive system.

Chapter 10 – This chapter attempts to provide new constraints on the scenario for the formation of galaxy clusters, based, in part, on the observational evidence presented in this thesis. The chapter is structured in 3 parts. In part I, we compile the first overview of observational evidence for overdensities of galaxies between z = 2 and z = 6. The overdensities, estimated from the number densities of star-forming galaxies (Ly\text{em} emitting galaxies and Lyman break galaxies) relative to random fields, are on the order of a few. If these structures were to collapse under the influence of their own gravity, their masses would be \( \sim 10^{14} \) to \( 10^{15} \) M\(_\odot\). Because this is comparable to the masses of clusters of galaxies in the local universe, we define the term ‘protocluster’ as being an object that meets the requirements for forming a bound object on the mass scale of a cluster prior to, or at, the present epoch, but which has not yet collapsed and virialized at the epoch corresponding to its observed redshift. In part II, we use simple theoretical descriptions for the growth of overdensities in a ΛCDM universe to study the evolution of the sample of candidate protoclusters compiled in part I. Using very conservative estimates of the overdensities, we find that the majority of the structures are likely to collapse within a finite time. We identify several structures as meeting the requirements for virialization at \( z \approx 0.5 \), whereas others are expected to have fully collapsed by the present epoch. In part III, we use a simple model for the star formation history of cluster red sequence galaxies to demonstrate that the observed star formation rates in protocluster fields can explain the build-up of the stellar mass in the red sequence galaxies of relatively nearby clusters.

References

Demarco, R., Ford, H. C., Rosati, P., MPLA, 20, 1409, 2005