1-1-2003

The clustering of galaxies around damped absorbers.

Nicolas F. Bouche
University of Massachusetts Amherst

THE CLUSTERING OF GALAXIES AROUND DAMPED ABSORBERS

A Dissertation Presented

by

NICOLAS F. BOUCHÉ

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2003

Department of Astronomy

THE CLUSTERING OF GALAXIES AROUND DAMPED ABSORBERS

A Dissertation Presented

by

NICOLAS F. BOUCHÉ

Approved as to style and content by:

James D. Lowenthal, Chair

Neal S. Katz, Member

Stephen E. Schneider, Member

Guy Blaylock, Member

Ronald L. Snell, Department Chair
Department of Astronomy

To Tim and his invaluable support,

ACKNOWLEDGMENTS

I would like to express my special thanks to

Dan McIntosh, Kelly Holley-Bockelmann and especially Sojourner Jackson for their help in proof reading the manuscript,

James Lowenthal, for his continuous moral (and financial) support,

Ari Maller, Houjun Mo, and Neal Katz for very helpful discussions,

the UMASS Department of Astronomy,

Max Pettini, and the Institute of Astronomy at Cambridge,

the committee members,

my parents for supporting my interest in astronomy,

and every one else whom I might have forgotten, but have directly or indirectly contributed to this work.

This research was partly based on observations made with the NASA/ESA *Hubble Space Telescope* obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under NASA contract NAS 5-26555.

This research was mainly based on observations made the Kitt Peak National Observatory, at the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under cooperative agreement with the National Science Foundation.

This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

ABSTRACT

THE CLUSTERING OF GALAXIES AROUND DAMPED ABSORBERS

SEPTEMBER 2003

NICOLAS F. BOUCHÉ

B.Sc., UNIVERSITÉ CATHOLIQUE DE LOUVAIN-LA-NEUVE, BELGIUM

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor James D. Lowenthal

We measured the clustering of galaxies around damped absorbers, all at a redshift of approximately three. For this purpose, we used wide-field images that covered a total area of one square degree. We selected galaxies that were close to the absorbers (in a redshift slice) using bayesian photometric redshift techniques. In the redshift slice centered on the absorbers, we detected the clustering of galaxies at the 87% confidence level. Due to the small sample of absorbers, we could not conclude whether the clustering was larger or weaker than that of the galaxies with each other. As a cross-check, in three other redshift slices that were not centered on the absorbers, there was no evidence for clustering.

We performed the same analysis on simulated data sets from numerical simulations. There, we found that the clustering of galaxies around the simulated absorbers is weaker than that of the galaxies with each other, by a factor of approximately two. This is consistent with damped absorbers being associated with low-luminosity galaxies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

INTRODUCTORY WORDS

What is the purpose of studying distant and seemingly useless galaxies? This is a question that many people ask me. After all, the thesis you are about to read will have little impact on the lay person, even on the well educated men and women of Amherst, MA. So, I think it is important to give a clear answer right from the start.

My opinion has always been that astronomy has and will continue to revolutionize our views of the world surrounding us, thanks to pioneers such as Galileo, Copernicus, Kepler, Hubble and others.

Astronomy today is a continuation of the work begun by these pioneers, and one can be sure that more surprises and revolutions are around the corner. However, I think the best answer to the question is given by C. Sagan (1996):

> We succeeded in taking that picture [from deep space], and, if you look at it, you see a dot. That's here. That's home. That's us. On it, everyone you ever heard of, every human being who ever lived, lived out their lives. The aggregate of all our joys and sufferings, thousands of confident religions, ideologies and economic doctrines, every hunter and forager, every hero and coward, every creator and destroyer of civilizations, every king and peasant, every young couple in love, every hopeful child, every mother and father, every inventor and explorer, every teacher of morals, every corrupt politician, every super-star, every supreme leader, every saint and sinner in the history of our species, lived there on a mote of dust, suspended in a sunbeam.
>
> The earth is a very small stage in a vast cosmic arena. Think of the rivers of blood spilled by all those generals and emperors so that in glory and in triumph they could become the momentary masters of a fraction of a dot. Think of the endless cruelties visited by the inhabitants of one corner of the dot on scarcely distinguishable inhabitants of some other corner of the dot. How frequent their misunderstandings, how eager they are to kill one another, how fervent their hatreds. Our posturings,

> our imagined self-importance, the delusion that we have some privileged position in the universe, are challenged by this point of pale light.
>
> Our planet is a lonely speck in the great enveloping cosmic dark. In our obscurity–in all this vastness–there is no hint that help will come from elsewhere to save us from ourselves. It is up to us. It's been said that astronomy is a humbling, and I might add, a character-building experience. To my mind, there is perhaps no better demonstration of the folly of human conceits than this distant image of our tiny world. To me, it underscores our responsibility to deal more kindly and compassionately with one another and to preserve and cherish that pale blue dot, the only home we've ever known.

If more world leaders were interested in astronomy, there would be less wars and pettiness between countries. As Sagan said, there is *only one* small speck of dust around, and the study of astronomy reminds us to share it and its resources more evenly.

CHAPTER 1

INTRODUCTION

In this chapter, I present the main objective of this work, and attempt to reach non-scientists and scientists who are not in this field. Section 1.1 presents the subject of quasar absorption lines and defines the focus of this thesis, namely hydrogen gas clouds called 'damped Ly-α absorbers'. Section 1.2 explains the motivation for this project and lays out our goals and the structure of this work. Finally, section 1.3 defines a few technical terms.

1.1 Quasar absorption lines

Quasi-stellar objects (QSOs or quasars) are among the most luminous objects in the universe. They are powered by very massive black holes that reside in the center of galaxies: the gas and stars that are slowly vanishing into the black hole are heated to high temperatures and a good fraction of the matter (10%) is converted into energy via Einstein's equation $E = mc^2$. This process makes quasars so luminous that they can be seen from very large distances.

Thus, scientists can perform detailed spectroscopic analyses of these very distant quasars. Quasars are so far away that the light observed from Earth today was in fact emitted when the universe was only 1 billion years old (7% of the universe's current age). Because of the vast distance covered by the light reaching us and because hydrogen is very abundant, intervening gas clouds are likely to intercept the light on its way to Earth. Fig. 1.1 illustrates a situation where a distant quasar is observed through one of today's large telescopes. Small and large clouds are represented as

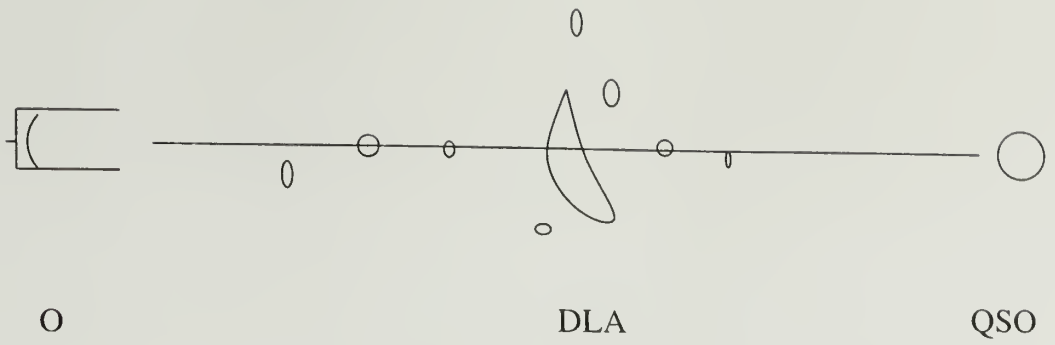

Figure 1.1 A bright quasar in a distant part of the universe is observed by the observer O's telescope [picture not to scale]. The quasar light path will likely intercept clumps of gas (represented as ellipses) that will be visible through distinctive absorption features in the quasar spectrum. These gas clumps may or may not be associated with galaxies. Small gas clouds will produce small absorptions, while larger or denser clouds will be seen as strong absorptions. See Fig. 1.3 to see the effect of the gas on a quasar spectrum. See Fig. 6.3 for a more realistic distribution of the gas from computer simulations by N. Katz.

ellipses. The light that travels through 90% of the universe is likely to intersect gas clouds on its way to the observer's telescope.

One major phenomenon happens to the quasar light because of the expansion of the universe: its spectrum[1] is shifted towards longer wavelengths, an effect known as redshifting. For instance, a quasar at a redshift z of six has its spectrum shifted by a factor of $1 + z = 7$. Thus, its emission lines are redshifted by the same factor. This phenomenon is shown for 10,000 quasars from the two degree field survey (2dF; Outram et al., 2001) in Fig. 1.2. As the redshift increases (vertical axis), the emission lines are shifted more and more towards the right.

Fig. 1.3 shows the spectrum of a distant quasar, QSO 0201+1120, which is at a redshift $z_{\rm QSO} = 3.61$. The spectrum reveals a few emission lines. For instance, the strongest emission line, which is clearly visible at ~ 5600Å, is Lyman-alpha (Lyα) and is due to the QSO itself. The mechanism that produces this strong emission line is the following: the high energy photons that are coming from the outskirts of the black

[1] A spectrum decomposes the light from the quasar in wavelength.

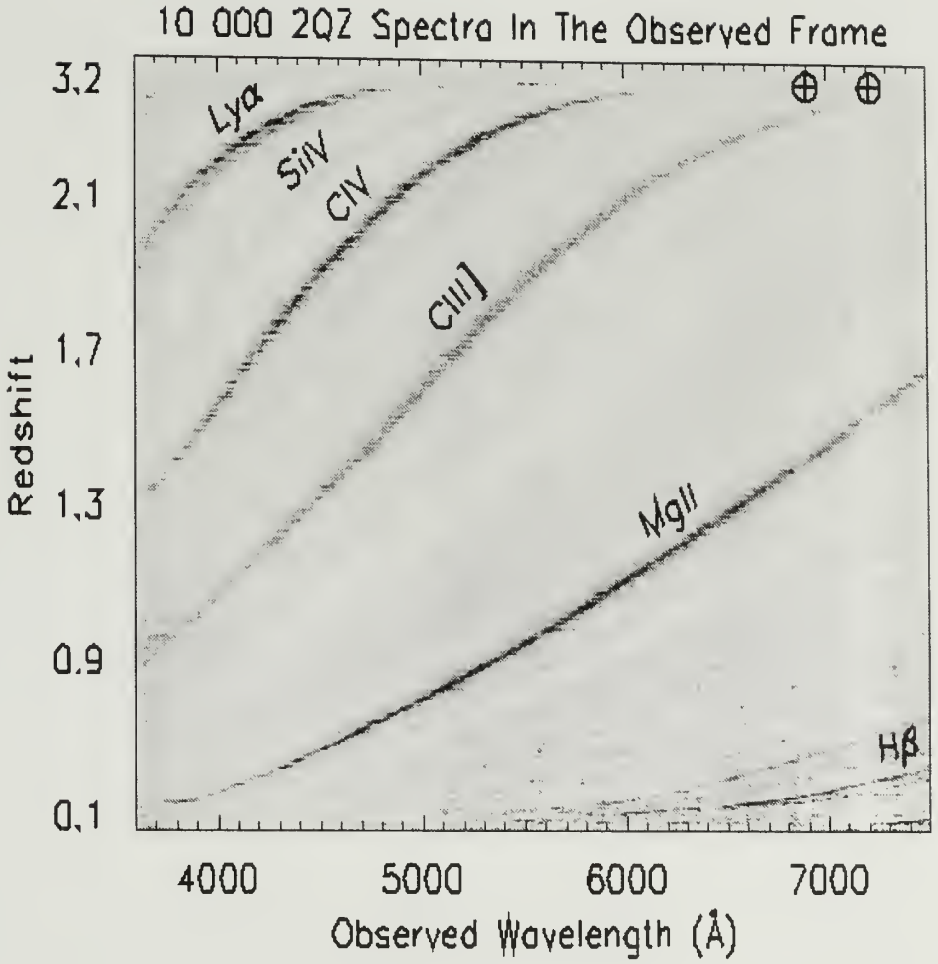

Figure 1.2 From the 2dF quasar redshift survey, the spectra of 10,000 QSOs are shown sorted by increasing redshifts (vertical axis). The emission lines (labelled) are being shifted towards longer wavelengths as the redshift of the QSO increases. Prominent emission lines are shown (reproduced from http://www.2dfquasar.org).

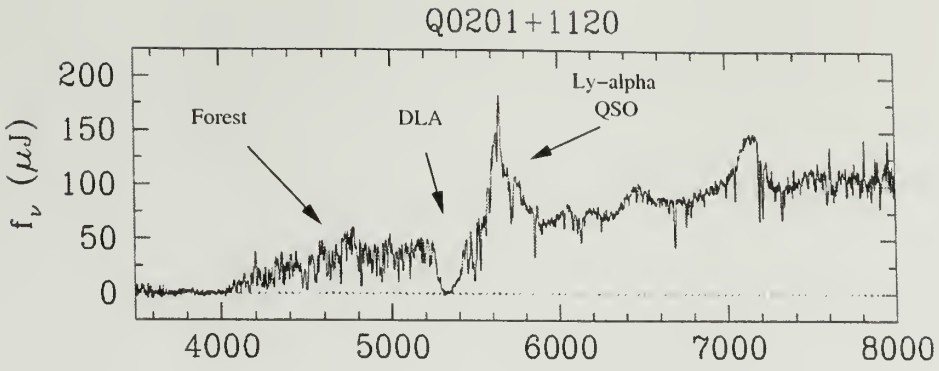

Figure 1.3 Spectrum of the quasar 0201+1120 which is at a redshift of $z_{\rm QSO} = 3.61$. The quasar light appears in emission. The strongest emission line is the redshifted Lyα emission line from the QSO itself. Intervening hydrogen gas will absorb light at the same transition Lyα. However, because the gas is closer to us, its redshift is lower than that of the QSO and it will absorb the QSO light to the left (blue-ward) of the QSO Lyα line. Thus, the numerous absorptions blue-ward of the QSO Lyα emission line are caused by diffuse, small, gas clouds along the line of sight that form the so-called Lyα 'forest'. A clear example of a larger cloud (DLA) causing the strong absorption at 5150Å is shown. All the features to the right of Lyα from the QSO are due to other atoms, called 'metals'.

holes excite hydrogen atoms in the vicinity of the quasar. Hydrogen is being ionized by the radiation, and electrons are set free. This continuous process is balanced by the recombination of electrons and hydrogen nuclei. During this recombination process, the electron goes back to the ground state of hydrogen, and emits light at a few particular wavelengths, each giving rise to an emission line. One of the most prominent lines is the so-called Lyman-α transition.

The spectrum in Fig. 1.3 also reveals many absorption lines. Each of these absorption lines are associated with one of the clouds that intercept the line of sight as depicted in Fig. 1.1. The clouds contain hydrogen atoms, some fraction that are neutral with their electrons in the ground state. Such an electron is very likely to absorb a photon, and create a Lyman-α absorption. Because the intervening clouds are closer to us, i.e. at redshifts lower than that of the QSO, their absorption will occur at a less redshifted position, i.e. to the left of the QSO Lyα line as shown in

Fig. 1.3. These small gas clumps are numerous, and thus will produce a large number of small absorptions. This is the so-called 'forest' in Fig. 1.3. Larger clouds, which are rarer, will produce larger absorptions. The absorption lines, to the right of Lyα of the QSO, are due to elements other than hydrogen, usually referred to as metals.

The Lyα forest, which is interesting in itself, is fairly well understood (e.g. Davé et al., 1999; Efstathiou, Schaye, & Theuns, 2000). Its properties (sizes, velocity distribution) can be reproduced by theoretical models and cosmological simulation. Briefly, most of the Lyα lines arise from small fluctuations in the matter density that constitutes the ionized intergalactic medium, rather than from distinct clouds. However, the rarer and larger absorbers (referred to as 'damped' absorbers) are not understood at all: after more than twenty years of research on damped clouds, their size, their relation to galaxies and their role in galaxy formation are still not understood. These rare absorbers are the focus of this thesis.

1.1.1 Damped Ly-alpha clouds, definition

Damped Ly-alpha absorbers (DLAs) are strong absorption lines seen in quasar spectra when a QSO shines through parts of hydrogen gas clouds. The total amount of hydrogen in the absorbing gas is related to the strength of the absorption, which in turn is related to the total amount of neutral hydrogen (HI), called *column density* (see section 1.3), along the line of sight.

Why are they called 'damped'? Every absorption line has a shape that is the combination of two components. The first component is due to the motion of the atoms in the gas and is referred to as the thermal component. The second component is intrinsic. From simple classical physics, it turns out that any transition behaves just like a damped oscillator. The damping is actually caused by the finite life-time of the atomic level associated with the transition. This intrinsic component produces a profile with large 'wings'. For small absorptions (small column densities), the thermal

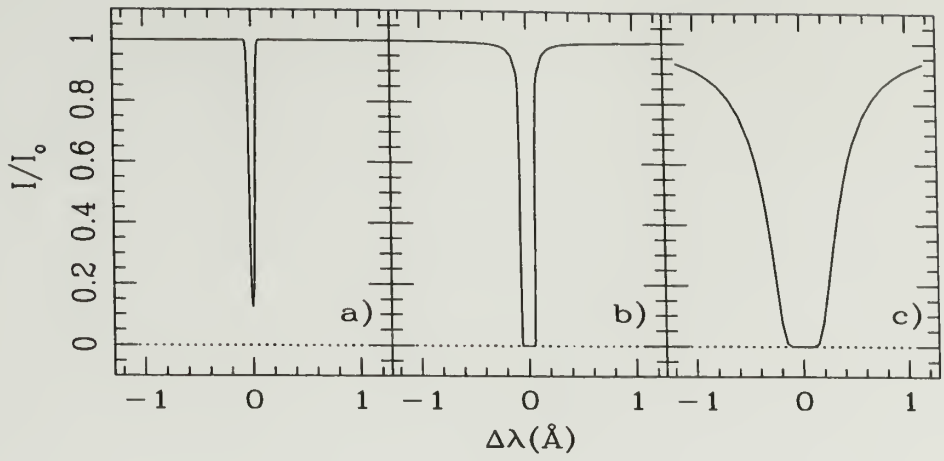

Figure 1.4 Modeled line profiles for increasing strength or column density, from low (a) to high (c). The strength is caused by the number of hydrogen atoms participating in the absorption. One sees that as the strength, or column density, increases, the 'damped wings' start to dominate the profile. The weak absorbers are more numerous and create the forest seen in Fig 1.3, while the stronger 'damped absorbers' are much rarer. The dotted line shows the saturation level.

component dominates and the wings are not noticeable. However, as the absorption gets stronger, the absorption line becomes saturated, and the 'damped wings' of the intrinsic component become more and more significant, and the absorption is said to be 'damped'. Fig 1.4 illustrates three absorption lines with increasing column densities.

Damped absorbers are, in fact, rare. Indeed, out of the tens of thousands of QSOs known, there are more or less three hundred DLAs known with redshifts up to $z \sim 4$. Chapter 2 will present an overview of what we know about them.

1.2 Thesis motivations: why study DLAs?

There are several reasons why DLAs are important. First, in the context of galaxy formation, there are only two classes of objects that we can observe at low and high redshifts, galaxies and absorbers. Bright galaxies are found by direct imaging with large telescopes, while absorption line clouds are seen only in quasar spectra.

Thus, studying absorption line clouds is key to understanding the process of galaxy formation (and evolution) at low and high redshifts.

Second, DLAs contain the largest reservoir of neutral hydrogen at high redshifts. They contain more hydrogen than all the absorption lines in the 'forest' combined, as shown in chapter 2. This large reservoir of neutral hydrogen makes them prime candidates to make a large number of stars. Thus, DLAs may be the precursors of galaxies, and they could play a significant role in galaxy formation.

Third, DLAs make up the only class of objects in which we can measure the abundances of heavy elements at very high redshifts. Thus, irrespective of their true nature, they provide us with the closest measure of global chemical enrichment in the universe. This explains why so many researchers are studying the abundance (and evolution) of heavy elements in DLAs (see section 2.6).

In summary, despite numerous studies, the role of DLAs in galaxy formation is still not understood. Will they evolve into large galaxies? Are they gas falling into young galaxies? Where do the heavy elements observed in DLAs come from? Since they have so much neutral gas, they ought to form stars. If DLAs indeed make stars, the abundance of heavy elements, which are produced in stars, ought to be connected with the global star formation history of the universe. It is possible to measure both the star formation history and the abundance of heavy elements in high redshift DLAs, and they do not match: The amount of metals inferred from the star formation history far exceeds the abundance of heavy elements observed in DLAs.

Originally, Wolfe et al. (1986) proposed that DLAs are large galaxies in formation with a thick gaseous disk . This is often referred to as the standard 'H I disk paradigm'. The main argument in favor of this theory is that the amount of gas contained in DLAs at high redshifts corresponds to today's stellar content (e.g. Lanzetta et al., 1991) as shown in section 2.3. This similarity implies that, between the early universe and

now, the neutral hydrogen gas has been converted into stars, a process that mainly occurs in large spiral galaxies. If that was the case, star formation should be highly visible in DLAs, but it is not (see chapter 2 § 2.5 and chapter 3).

1.2.1 Our scientific goals

Since the nature of DLAs is still not known and remains in debate, the scientific goal of this thesis is the following: **to constrain the nature of DLAs** using two particular approaches.

The first approach is to try to directly image a DLA using the *Hubble Space Telescope* to search for a signature of star formation. Indeed, a region with strong star formation (i.e. with numerous young stars emitting strong UV radiation) will produce strong emission lines (Lyα and Hα). The strength of the emission gives an estimate of the star formation rate. Chapter 3 describes the data and our results of such a study of a DLA at $z = 0.656$.

The second approach is indirect. We investigate the matter distribution around four DLAs at $z \sim 3$, corresponding to 85% of the age of the universe when galaxies were still in formation. We ask, "Are DLAs found in over- or under-dense regions?" The answer to this question will yield clues as to whether or not DLAs are associated with massive galaxies. Massive objects tend to cluster under gravity. Thus, if DLAs are massive, galaxies would tend to cluster around DLAs. The scale at which this clustering occurs and the amount of excess galaxies (or overdensity) will yield a constraint on the mass of DLAs. Therefore, the clustering of galaxies will provide clues to the nature of the damped absorbers. Practically speaking, if one takes images of the sky with the DLA in the center, the distribution of galaxies, which are used as tracers of the matter distribution, will or will not be clustered around the DLA. These high redshift galaxies are often called 'Lyman break' galaxies (LBGs).

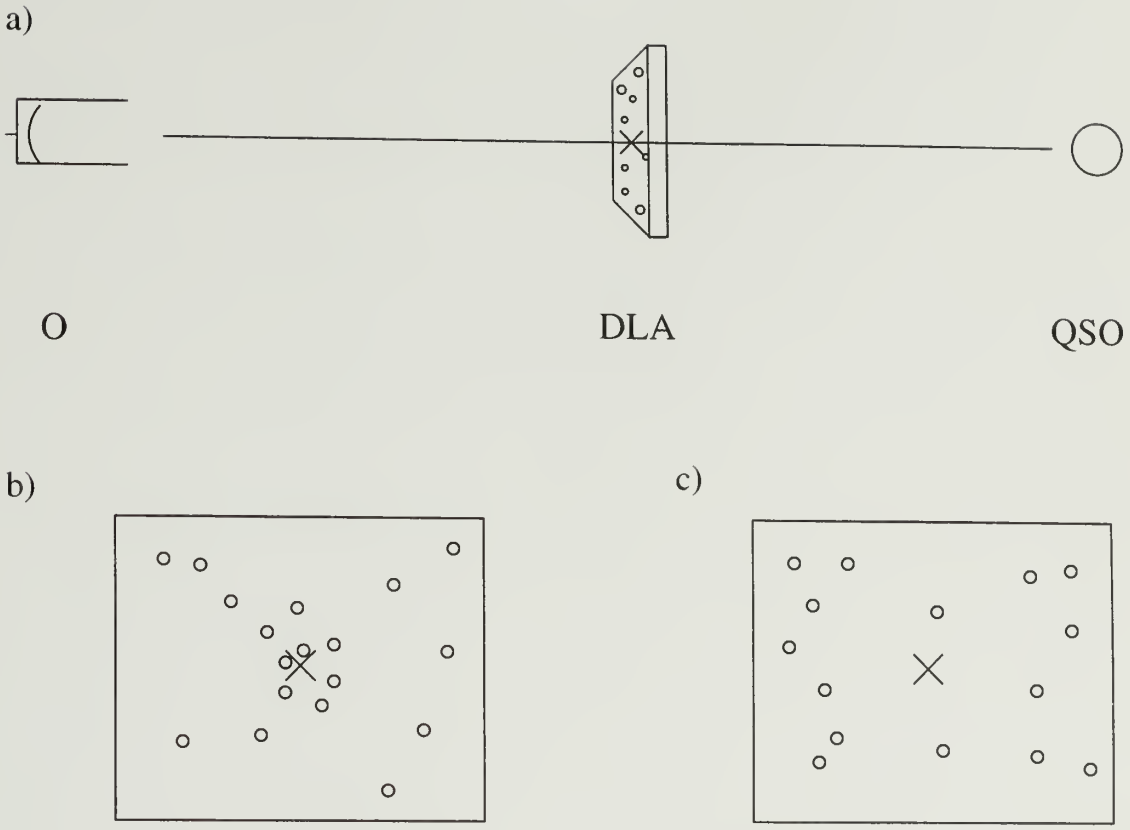

Figure 1.5 The DLA/LBG connection. Panel (a) shows the profile view. A quasar in the background (shown on the right) reveals the presence of a DLA (shown by the cross). Galaxies that are close in distance to the DLA are shown as circles within the survey region depicted as a rectangle. Panels (b) and (c) show the front view of the distribution of the galaxies. An example of galaxies that are clustered around the DLA is shown in panel (b), while an example of galaxies that are not clustered around the DLA is shown in panel (c). Two alternative scenarios that have opposite consequences. In the former, DLAs would reside in over-dense regions, in the latter, in underdense region.

Fig. 1.5 illustrates the situation. Panel a) is a profile view (similar to Fig. 1.1) that shows a quasar in the background with galaxies distributed along the line of sight, and a DLA at some distance shown by the cross. Galaxies are represented by circles. To study the clustering of galaxies that might be due to the presence of the DLA, we need to select galaxies that are close in distance (or redshift) to the DLA. In other words, we need to select galaxies in a narrow slice along the line of sight, represented by the rectangle in Fig. 1.5, and remove all the galaxies that are in the foreground or in the background of the DLA.

Panels b) and c) of Fig. 1.5 show a front view of the galaxies within the narrow redshift range. The DLA is represented once again by a cross in the middle of the field of view. Panel b) shows an example of clustering centered on the DLA, i.e. the DLA resides in an over-dense region of galaxies. On the other hand, panel c) shows an example of anti-clustering, i.e. the DLA resides in an underdense region. These two scenarios lead to very different conclusions about the nature of DLAs. The former would indicate that DLAs are likely to be massive, or associated with a massive structure, while the latter would indicate that DLAs are not part of a massive structure.

The remainder of this work is organized as follows. Chapter 2 summarizes the current knowledge on Lyman break galaxies and on DLAs and results from similar investigations (§ 2.10). For readers unfamiliar with the current knowledge of astronomy, Appendix A gives an overview of the Big Bang theory in the context of galaxy formation. Chapter 3 describes our study of a DLA at $z = 0.656$, based on the first approach. The following three chapters are focused on the second approach, the LBG/DLA connection. Chapter 4 presents the four DLA fields and the data reduction. Chapter 5 shows the method used to select galaxies in a redshift slice centered on the DLA, and chapter 6 contains our results of the clustering analysis.

1.3 Definition of some technical terms

We end this chapter with the definition of a few relevant technical terms:

- **Magnitude:** During the Greek civilization, stars were classified into six classes by Hipparchus (2nd century B.C.), the first class being the brightest that he could see. The faintest that he could see defined the sixth class. Later, observers realized that the eye is sensitive to light logarithmically. The result is the following definition:

$$m = -2.5 \log[\text{flux}(\text{erg s}^{-1} \text{ cm}^{-2})]. \tag{1.1}$$

In practice, the light coming from a star is filtered by the Earth's atmosphere, and by the sensitivity of the photographic plate or the digital detector. Thus, the observed magnitude m_R is the convolution of the observed flux density (flux per unit wavelength) $F_\lambda(\lambda_o)$ with the filter sensitivity R, normalized to a standard-source flux density $g_\lambda^R(\lambda_o)$:

$$m_R = -2.5 \log \left[\frac{\int_0^\infty F_\lambda^{obs}(\lambda_o) R(\lambda_o) \mathrm{d}\lambda_o}{\int_0^\infty g_\lambda^R(\lambda_o) R(\lambda_o) \mathrm{d}\lambda_o} \right]. \tag{1.2}$$

Note the negative sign, and that it is not a flux measurement. Rather it is a flux ratio between the source flux F_λ and the reference flux g_λ.

- **AB-magnitude:** The above definition depends on the reference flux density $g_\lambda^R(\lambda_o)$. A better definition uses a constant flux density as the reference, i.e. $g_\lambda^R(\lambda_o) = 3631$ Jy:

$$m_{\mathrm{AB},\lambda} = -2.5 \log [f(\mathrm{Jy})] - 48.6 \tag{1.3}$$

where f is the flux density expressed in Jy; 1Jy=10^{-23}erg s^{-1} cm^{-2} Hz^{-1}.

- **Color:** The difference in magnitude or the ratio in the flux between two filter pass bands. In the case of U and B, it is denoted as $U - B = m_U - m_B$. A 'red' galaxy will have more flux in the B band, and have $U - B$ more positive.

- **Column density:** The integral of the volume density n along the line of sight, $N_X = \int n_X \, dl$, where n_X is the number density of atoms in cm^{-3}. Physically, it is the number of atoms per unit area, expressed in cm^{-2}

- **Line profile:** An absorption (or emission) line in a spectrum is characterized by its profile $I(\nu)$, which is a function of the cross section (σ_ν) and of the column density N_X of the element X: $I(\nu) = I_o \, e^{-\tau_\nu}$, where I_o is the continuum, and $\tau_\nu = N_X \sigma_\nu$ is the optical depth.

- **Equivalent width:** The area of the line, i.e. the integral of the absorption profile, normalized to the continuum: $W = \frac{\lambda^2}{c} \int \frac{I_o - I}{I_o} d\nu = \frac{\lambda^2}{c} \int (1 - e^{-\tau_\nu}) d\nu$, expressed in Å.

- **Metals:** Since 99% of the baryons in the universe (see Appendix A) is made of hydrogen and helium, astronomers call anything else 'metals'. The metal abundance is the ratio of the column density of a metal X to that of hydrogen ($\frac{N_X}{N_H}$).

- **Metallicity:** The metallicity is the logarithm of the abundance compared to the metal abundance of the Sun: $[X/H] = \log(\frac{N_X}{N_H}) - \log(\frac{N_X}{N_H})_\odot$. E.g., $[X/H] = -1$ is a metallicity of 1/10th of the solar abundance.

- **Ions:** Neutral species are noted X I, single ionized species are noted X II, and so on for higher ionization states.

- **Comoving coordinates:** As the universe expands, its density decreases and distances between galaxies are being stretched. Thus, it is natural to express

distances on a scale that does not depend on the expansion. Such a scale is referred to as the 'comoving' coordinate system. For instance, at a redshift z, a physical scale of one Mpc corresponds to $1+z$ Mpc comoving. Note that any bound system, such as atoms or galaxies do not expand, and thus will have the same physical size at all redshifts.

- **Seeing:** The atmosphere blurs the light coming from stars. Good seeing ($< 1''$) means that the atmosphere did not smear the data as much as in bad seeing ($> 1.5''$). The seeing affects the size of the point spread function (PSF). It is quantified from the full-width at half maximum (FWHM) of the PSF for stellar objects.

Throughout this work, we adopted the following cosmological parameters: $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$[2] and $H_o = 100h$ km s^{-1} Mpc^{-1} consistent with the recent results of the cosmic micro-wave background (Bennett et al., 2003). Thus, at $z \sim 3$, $1''$ corresponds to $\sim 21.5h^{-1}$ kpc and $1'$ to $\sim 1.29h^{-1}$ Mpc, both comoving. At that redshift, the Hubble constant is $H(z) \sim 4.46H_o$, where $H_o = 100h$km s^{-1} Mpc^{-1}. Thus, at redshift $z \sim 3$, a redshit interval of $\delta z = 0.1$ corresponds to $67h^{-1}$ Mpc in comoving coordinates.

Now that all the necessary information for the reader to follow the next chapters had been presented, the remainder of this work will be more technical.

[2]see Appendix A.

CHAPTER 2

SCIENTIFIC BACKGROUND

It is important to understand fully the significance of our study in the context of the general scientific background. In this chapter, we first review in § 2.1 what we know about the galaxies used as tracers of the matter distribution, namely the so-called 'Lyman break' galaxies. In § 2.2 we present a general overview of those approaches. For the reader interested in knowing more about them, each of the following sections is dedicated to one particular approach. The most relevant section of this chapter is section 2.10, which presents previous results on the clustering of galaxies around damped absorbers. Section 2.11 summarizes the available information on DLAs.

2.1 Lyman break galaxies

To investigate the matter distribution around DLAs, we needed tracers of the matter structure. Lyman break galaxies (LBGs) are good candidates for this purpose for they are very luminous and easily detectable at $z \sim 3$, our redshift of interest. They are easily detectable because of the following two reasons. First, starlight emitted in these galaxies is shifted towards the red by a factor of $1 + z = 4$. Thus, astronomers see only the intrinsic ultra-violet (UV) and blue starlight. Because young stars have a large number of massive (and hence luminous) stars that emit lots of UV photons, a young galaxy will be very luminous in the UV as well. Second, almost all the UV radiation at shorter wavelength than the transition Lyman limit ($\lambda_{rest} <$ 912Å) emitted by the galaxy is very likely to ionize hydrogen in the intergalactic

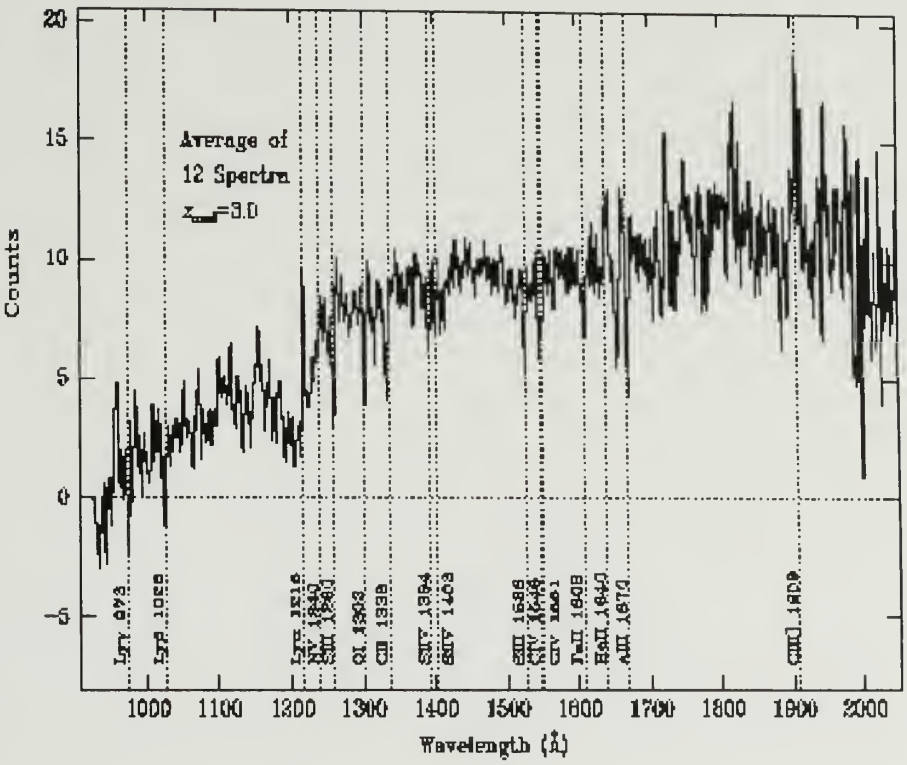

Figure 2.1 The averaged spectra of 12 LBGs as a function of rest-frame wavelength. It shows that all the flux shorter than 912A is absorbed by the inter-galactic medium producing a 'break' in the spectral energy distribution. Several other features are shown. (Reproduced from Lowenthal et al., 1997)

medium and be absorbed. This produces a break in the observed galaxy spectrum at rest-wavelengths $\lambda_{rest} < 912$A (see Fig. 2.1) that, for galaxies at $z = 3$, occurs in the optical part of the spectrum. The result is that in order to find distant galaxies at $z = 3$, one only needs to look for galaxies with such a strong break.

In contrast to DLAs, many properties of these Lyman break galaxies are better known since the first observations of Steidel & Hamilton (1993). Many studies of high-redshift galaxies have been undertaken (e.g. Lowenthal et al., 1997; Steidel et al., 1999), and LBG properties (see Giavalisco, 2002, for a review) are well determined:

- they are compact, their half-light radius $< 5h^{-1}$kpc (Lowenthal et al., 1997),
- they are luminous, they have luminosities 1-2 magnitudes brighter than a typical galaxy today,
- they are star-forming, their star formation rates are $10-50h^{-1}$ M$_\odot$ yr^{-1}, (Lowenthal et al., 1997; Shapley, 2001), and
- they are strongly clustered (e.g. Adelberger et al., 1998; Giavalisco & Dickinson, 2001; Porciani & Giavalisco, 2002, and section 6.3.2).

Their stellar mass, however, is still highly uncertain.

2.2 Overview of the properties of DLAs

As discussed in section 1.1.1, DLAs are the strongest QSO absorption systems with column densities ($N_{H\,I}$) greater than 2×10^{20} atoms per cm^2. This threshold is somewhat arbitrary: it is high enough to ensure the gas is entirely neutral[1], and it also corresponds to the neutral hydrogen column density that one would observe if a quasar shone through our Galaxy.

[1]The exact threshold at which the gas becomes neutral depends on the radiation (UV) surrounding the gas. In a low star formation region with a weak UV radiation, the threshold will be lower.

Previous observational results on DLAs can be summarized as follows:

1. There is little evolution of the total hydrogen gas content in DLAs, or $\Omega_{\rm DLA}(z)$, from $z = 4$ to $z = 1$ (Péroux et al., 2001; Rao & Turnshek, 2000) nor in the number of DLAs per unit redshift dn/dz from $z = 4$ to $z = 0$ (Storrie-Lombardi et al., 2000; Rosenberg & Schneider, 2003);

2. The velocity profiles are asymmetric. The velocity width distribution of DLAs is, however, not unique and can be explained with many different models (e.g. Prochaska & Wolfe, 1997; Haehnelt, Steinmetz, & Rauch, 1998);

3. Few DLAs have been directly imaged. The detected ones show various morphological types: some are large disks, while others are compact and amorphous objects of various sizes and luminosity (Le Brun et al., 1997; Kulkarni et al., 2000; Rao & Turnshek, 2000; Pettini et al., 2000);

4. Number counts indicate that their cross sections are small ($< 20h^{-1}$ kpc);

5. Low redshift ($z < 1$) DLAs have heterogenous chemical properties (Pettini et al., 2000). Furthermore, chemical evolution models do not agree with the observed metal enrichment in DLAs: the gas in DLAs shows little metallicity evolution (Pettini et al., 1999) from z = 3 to z = 0.3;

6. DLAs have small amounts of dust;

7. Few simulations can reproduce all the observed properties of DLAs, given their small cross-section ($10 - 15$ kpc);

These results are inconsistent with the standard DLA/H I disk paradigm (presented in section 1.2) that DLAs are large gaseous disk galaxies (Wolfe et al., 1986, 1995), for they indicate that DLAs (i) do not participate in the overall chemical enrichment of the universe and hence do not trace star formation in any straightforward

way; and (ii) are not characteristic of a particular type of galaxy. Rather they are merely characteristic of a particular type of region: namely, one with a large neutral column density (e.g. Khersonsky & Turnshek, 1996). Thus, it is quite reasonable to see DLA clouds as a heterogenous type of object.

In the following sections, we review each of the aspects mentioned above.

2.3 Redshift evolution of the density of DLAs

The answer to the question "Does the comoving number density $n_{\rm DLA}(z)$ of absorbers evolve with redshift?" should provide important information on the nature of damped absorbers. For instance, if this evolution closely follows the evolution of the number of galaxies, that would indicate that both populations are similar.

Unfortunately, the number density $n_{\rm DLA}(z)$ is not directly observable, because we have a biased sample of DLAs: we see only the ones with a QSO shining behind it.

A more natural tool to investigate the evolution of $n_{\rm DLA}(z)$ is the derivative of n: the number of absorbers per unit redshift (per unit line-of-sight) $\frac{\rm dn}{\rm dz} = l(z)$ which gives the probability per unit redshift of having an absorber along a given line-of-sight and is directly observable (Bahcall & Peebles, 1969). Unfortunately, as explained below and in section B.1, $\frac{\rm dn}{\rm dz}(z)$ does not shed more light on the evolution of $n(z)$. However, $\frac{\rm dn}{\rm dz}(z)$ is directly related to the column density distribution $f(N_{\rm H\,I})$ (see Eq. B.3) which in turn yields the total (hydrogen and helium) gas content in DLAs, $\Omega_{\rm DLA}(z)$ (from Eq. B.5). Appendix B summarizes the relationships between $\frac{dn}{dz}(z)$, $f(N_{\rm H\,I})$ and $\Omega_{\rm DLA}(z)$. Next, we present what we know about these three quantities.

Some surveys have yielded $\frac{dn}{dz}(z) = l_o\ (1+z)^{\gamma}$, where $\gamma = 1.1 \pm 0.5$ and $l_o = 0.055 \pm 0.03$ [2] (Lanzetta, Wolfe, & Turnshek, 1995; Storrie-Lombardi et al., 2000). Fig. 2.2 shows $\frac{dn}{dz}$ for different surveys. From the 1σ error bars at $z \sim 1$ in Fig. 2.2.

[2]Note that l_o for the Lyα forest is ~ 30.

one can see that the slope of $\frac{dn}{dz}(z)$ is poorly constrained, as discussed in Storrie-Lombardi et al. (2000). These results are consistent with no-evolution prediction ($\gamma = 0.5$) shown by the continuous line (see Eq. B.2).

Tytler (1987) first reported the surprising fact that the column density distribution of all absorption systems, damped and non-damped, can be described by a single power law ($f(N) \propto N^{-\beta}$) over 10 orders of magnitude in column densities (N) with $\beta \simeq 1.5$ (see also Storrie-Lombardi et al., 2000). Thus, among QSO absorption systems, DLAs are the rare 'heavy-weights' at the upper end of the distribution of values of H I column densities (see Fig 2.3).

One can integrate the column density distribution ($f(N, z) \propto N^{-\beta}$) and derive (see section B.3) the total gas content in DLAs $\Omega_{\rm DLA}(z)$, shown in Fig. 2.4 from the recent DLA survey Ellison et al. (2001). The result is that the rare DLAs with the largest column densities dominate the contribution to $\Omega_{\rm DLA}(z)$. That is, the DLA contribution dominates all the other absorbers. This implies that DLAs account for the largest reservoir of neutral hydrogen (H I) at high redshifts (Wolfe et al., 1986; Lanzetta et al., 1991; Lanzetta, Wolfe, & Turnshek, 1995; Wolfe et al., 1995; Storrie-Lombardi et al., 1996, 2000; Péroux et al., 2001).

As already pointed out in section 1.2, it is important to compare the amount of gas at redshift $z = 3$, $\Omega_{\rm DLA}(z = 3)$, to the amount of stars at $z = 0$, $\Omega_{\rm stars}(z = 0)$, and to the amount of H I today at $z = 0$. From section A.6.1, the amount of stars, which is represented by a filled star in Fig. 2.4, is $\log \Omega_{\rm stars}(z = 0) = -2.45$ [3] and is comparable to $\Omega_{\rm DLA}(z = 3)$. On the other hand, the amount of gas at redshift $z = 0$ is $\log \Omega_{\rm H\,I}(z = 0) = -3.4$ [4] (From the local H I surveys of Rosenberg & Schneider, 2002) and is represented by a filled square in Fig. 2.4. Thus,

---

[3]Recently, Bell et al. (2003) found $\log \Omega_{\rm stars}(z = 0) = -2.56$, using $h = 0.65$.

[4]converted to $h = 0.65$.

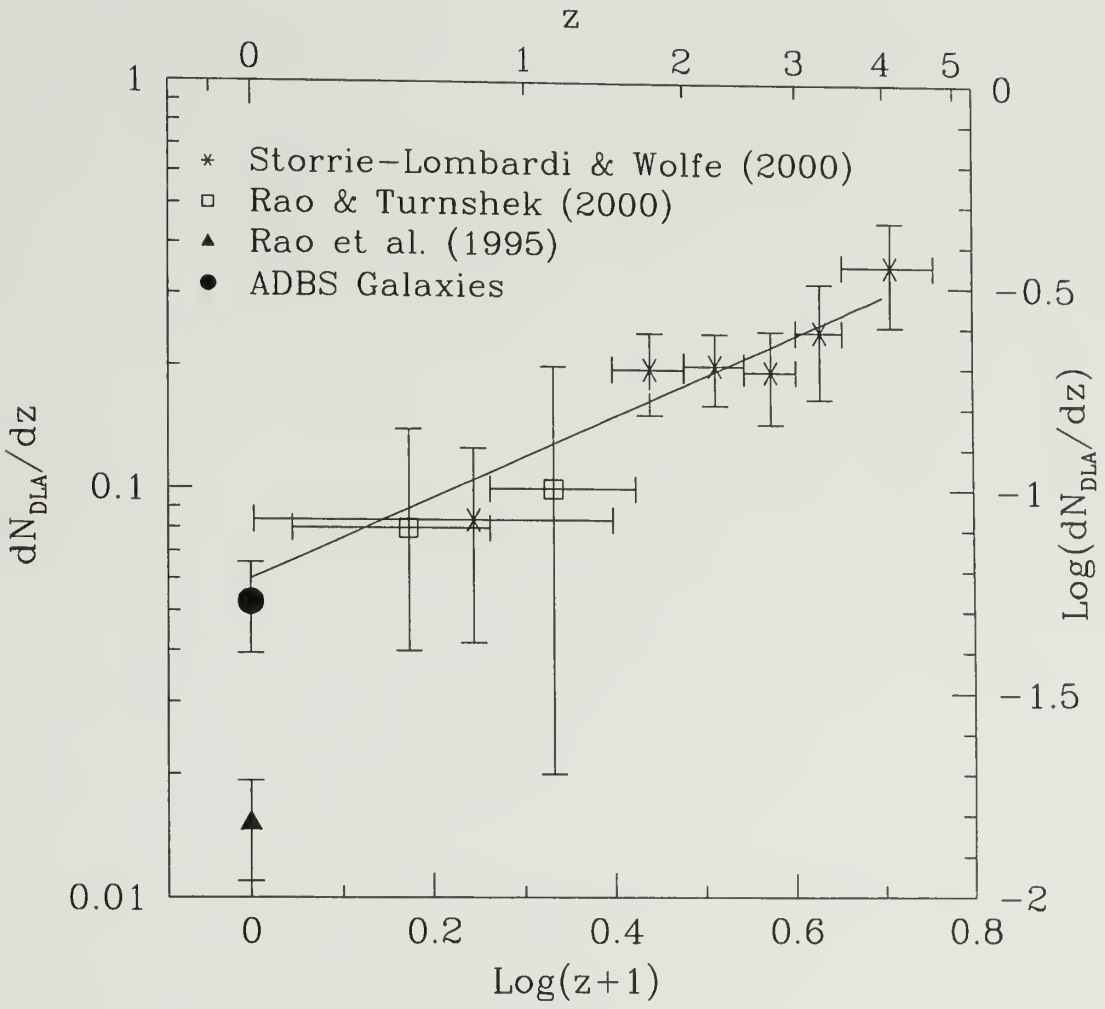

Figure 2.2 The number density damped absorbers per unit redshift $\frac{dn}{dz}(z)$ from different surveys. The line shows the no evolution prediction $\frac{dn_{DLA}}{dz}(z) \propto (1+z)^{0.5}$. The recent survey of the local $z = 0$ H I distribution of Rosenberg & Schneider (2003) is represented by the filled circle. All the surveys are consistent with the no evolution model. (Reproduced from Rosenberg & Schneider, 2003)

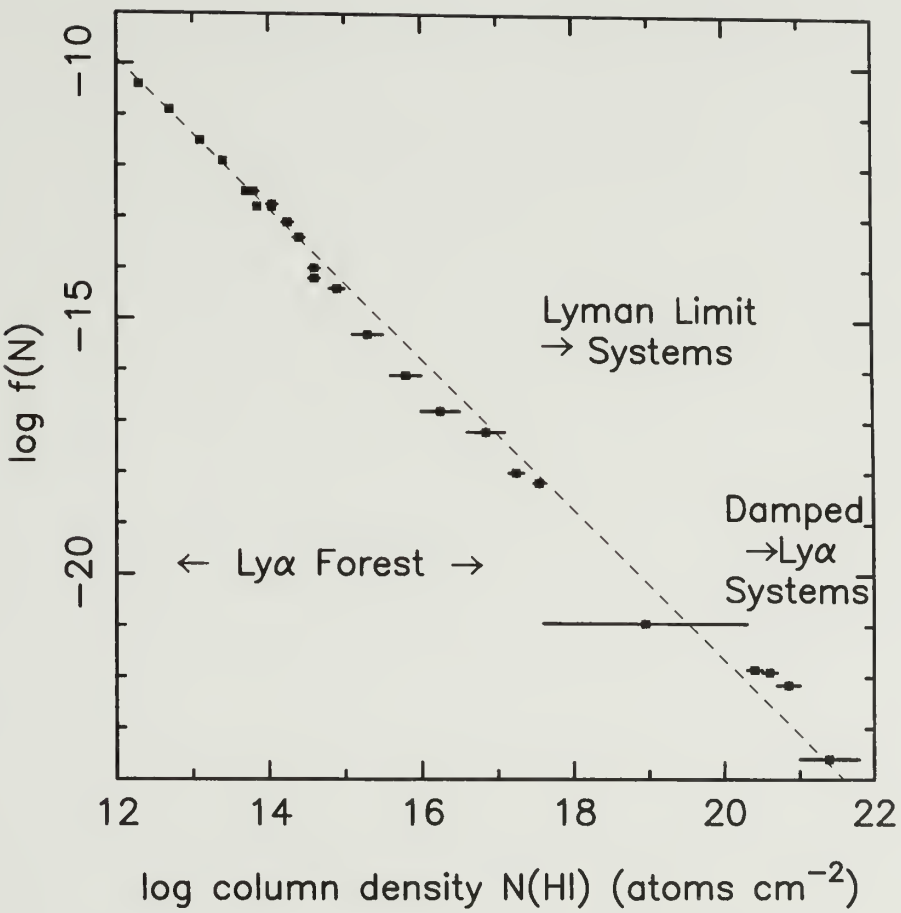

Figure 2.3 The column density distribution of neutral hydrogen (for all absorption systems in 30 QSOs) spans 10 orders of magnitude from $\log N_{\rm H\,I} = 12$ to 22. It can be fitted by a power law $f(N_{\rm H\,I}) \propto N_{\rm H\,I}^{-1.5}$, shown by the dashed line. (Reproduced from Storrie-Lombardi et al., 2000)

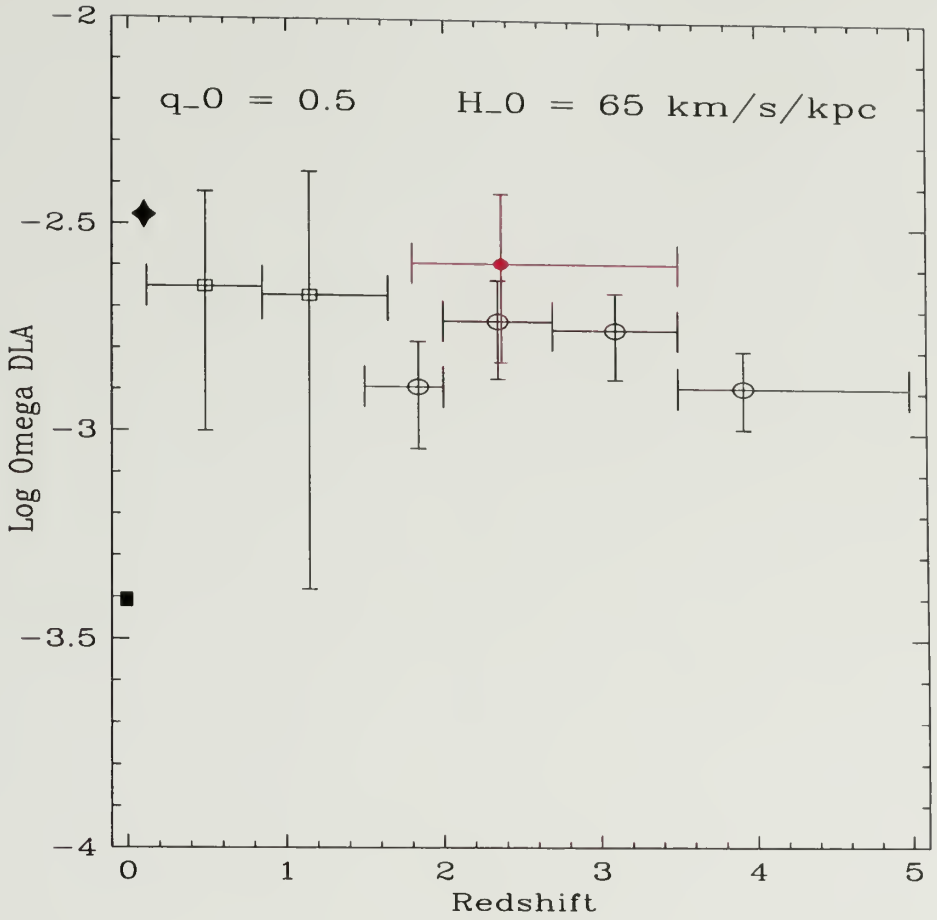

Figure 2.4 Estimates of the mass density of neutral hydrogen gas traced in damped Lyα absorbers. The squares are from Rao & Turnshek (2000), the open circles are from Storrie-Lombardi et al. (2000), while the filled red circle is from the recent survey by Ellison et al. (2001). There is no indication of changes with redshift, i.e. no evolution is seen in the total H I content from $z = 4$ to $z = 1$. At $z < 1$, the results of Rao & Turnshek (2000) are also consistent with little evolution, due to their large error bars. The filled square represents the H I content today, $\log\left[\Omega_{\rm H\,I}(z=0)\right] = -3.4$ (from Rosenberg & Schneider, 2000). The filled star represents the stellar content at $z = 0$, $\log\left[\Omega_{\rm stars}(z=0)\right] \simeq -2.5$. (Reproduced from Ellison et al., 2001)

$$\Omega_{\rm DLA}(z=3) = 5-6 \times \Omega_{\rm H\,I}(z=0) \simeq \Omega_{\rm stars}(z=0). \qquad (2.1)$$

In addition, Fig. 2.4 shows that there is little evolution of $\Omega_{DLA}(z)$ from $z = 4$ to $z = 1$ (Storrie-Lombardi et al., 2000; Ellison et al., 2001; Péroux et al., 2001). When compared to models, the behavior of $\Omega_{\rm H\,I}(z)$ with redshift provides an indication of the evolution of the star formation rate in the universe (e.g. Pei, Fall, & Hausser, 1999; Boissier, Péroux, & Pettini, 2003, also discussed section 2.6).

Unfortunately, no model can account for the slow evolution of $\Omega_{DLA}(z)$. For instance, Boissier, Péroux, & Pettini (2003) compared DLA properties ($\frac{{\rm d}n}{{\rm d}z}$; $\Omega_{\rm DLA}(z)$) with galaxy evolution models based on extrapolating the properties of present galactic disks to early times. They concluded that DLA statistics are well reproduced with their model up to $z < 2$ and that while, low surface brightness (LSB) galaxies are small, LSBs are necessary at $z < 2$ to account for $\frac{{\rm d}n}{{\rm d}z}$ because of their large cross-section and large neutral hydrogen gas content. At $z > 2-3$, their model does not reproduce the observations. They interpret this discrepancy as evidence for the existence of an additional population of absorbers, of generally lower column density, which dominates the number of absorbers at $z \simeq 3$. They speculate that these could be sub-units that will merge with each other and with more massive galaxies at $z \leq 2$. This is in contrast to the H I-disk paradigm where all absorbers are thought to be large spiral disks.

In a similar study, Jimenez, Bowen, & Matteucci (1999) concluded that large disks consume neutral gas too quickly to explain the evolution of $\Omega_{\rm DLA}(z)$, and that the consumption of H I gas by LSBs fits the data better. They remark that this further suggests that LSB disks may dominate the DLA population at high-redshift.

2.4 The kinematics of DLAs

The main (albeit indirect) evidence supporting the disk hypothesis of Wolfe et al. (1986, 1995) comes from measurement of the absorption profiles of metals associated with the DLA. Indeed, if in DLAs, hydrogen is predominantly neutral, other atomic species attain low-ionization states (e.g. Viegas, 1995). As mentioned in section 1.1.1, the absorption profile is nothing more than a profile convolved with the velocity distribution of the gas. For Lyα, the profile width is dominated by its intrinsic broadening. Because the line is not saturated for ions, such as SiII, AlII, FeII, and NiII, the width of these lines is dominated by the velocity distribution of the gas and is, therefore, generally considered a suitable tracer of the velocity distribution of the neutral gas.

It turns out that the line profiles are observed to be highly asymmetric. For example, the study of Prochaska & Wolfe (1997), reproduced in Fig. 2.5, shows the velocity profiles of seventeen DLAs. Fourteen of them have asymmetric profiles. Prochaska & Wolfe (1997) argued that those asymmetric profiles are consistent with what one would expect from lines of sight intercepting thick rotating gaseous disks.

Unfortunately, more than one model can explain the observed kinematics of DLAs. For instance, Haehnelt, Steinmetz, & Rauch (1998) showed that the kinematics asymmetries can be caused by random sampling of irregular density and velocity fields of individual halos, or by intrinsically asymmetric configurations, which arises when two or more clumps collide.

2.5 Emission searches: constraints on star formation in DLAs

As mentioned in section 1.2.1, if DLAs are related to large disks or if they are producing metals (§ 2.6), then they presumably are forming stars. Thus signatures of star formation ought to be detectable. In particular, young massive stars, which produce strong UV radiation, ionize the surrounding gas. Since some of the ionized

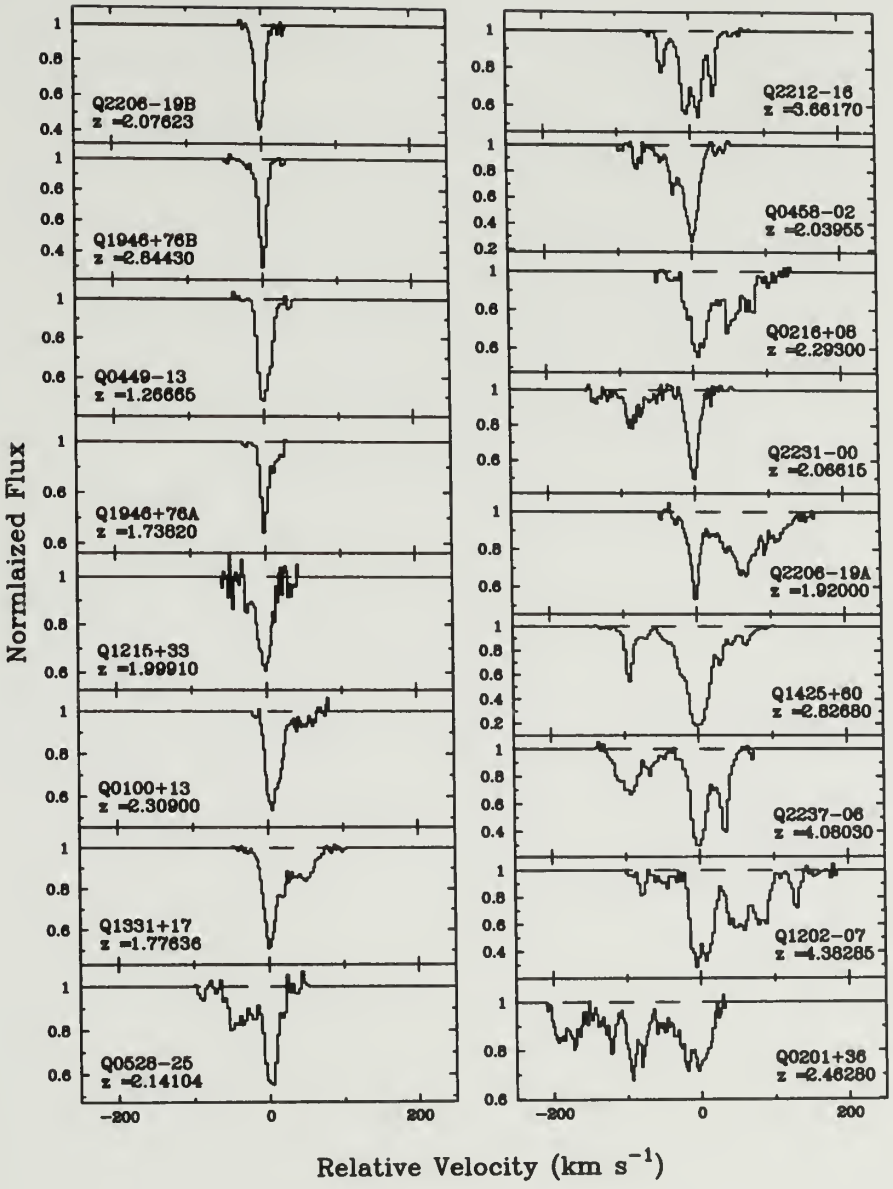

Figure 2.5 Velocity profiles of a sample of 17 DLAs. 14 DLAs have asymmetric profiles. According to Prochaska & Wolfe (1997, 2002), the asymmetric profile distribution is consistent with lines of sight intercepting thick rotating gaseous disks. (Reproduced from Prochaska & Wolfe, 1997)

gas recombines, recombination lines such as the Lyα(1216Å) and Hα(6563Å) atomic transitions are good tracers of the UV radiation. If one assumes a mass distribution for all the stars, called 'mass function', one can convert the UV radiation into a global star formation rate (SFR) (Kennicutt, 1983). SFRs are expressed in solar masses per year: $M_\odot$ yr^{-1}. Kennicutt (1983) and Kennicutt (1998) calibrated the conversion between the observed Hα flux and the star formation rate:

$$\mathrm{SFR}\ (M_\odot\ \mathrm{yr}^{-1}) = L_{\mathrm{H}\alpha}/1.26 \times 10^{41} \mathrm{erg\ s}^{-1}. \tag{2.2}$$

Since the strength of the Lyα line emission is proportional to Hα, the strength of the Lyα line emission is also a measure of the star formation rate. However, Lyα emission may be suppressed by dust extinction, and such measurements should be regarded as lower limits.

Several imaging studies have attempted to directly detect damped Lyα absorbers via this line emission signature or via continuum emission. All the efforts to directly image DLAs can be classified in three categories: A) broad-band imaging to search for continuum emission from stars; B) narrow-band imaging; and C) spectroscopic studies. Both (B+C) are attempts to detect line emission (Lyα or Hα).

2.5.1 Broad-band imaging

With the advent of rest-UV imaging and spectroscopy from the *Hubble Space Telescope (HST)*, data have become available on intermediate redshift $0.1 < z < 1.8$ DLA systems (Le Brun et al., 1997; Rao & Turnshek, 1998; Boissé et al., 1998). Space based observations were crucial for this since Lyα is not visible at $z < 1.75$ from the ground. Le Brun et al. (1997) showed that DLAs display a wide range of morphologies and surface brightnesses: in their sample of eight detections, three were spirals, three others had luminous compact morphologies, and two were compact, low surface brightness galaxies.

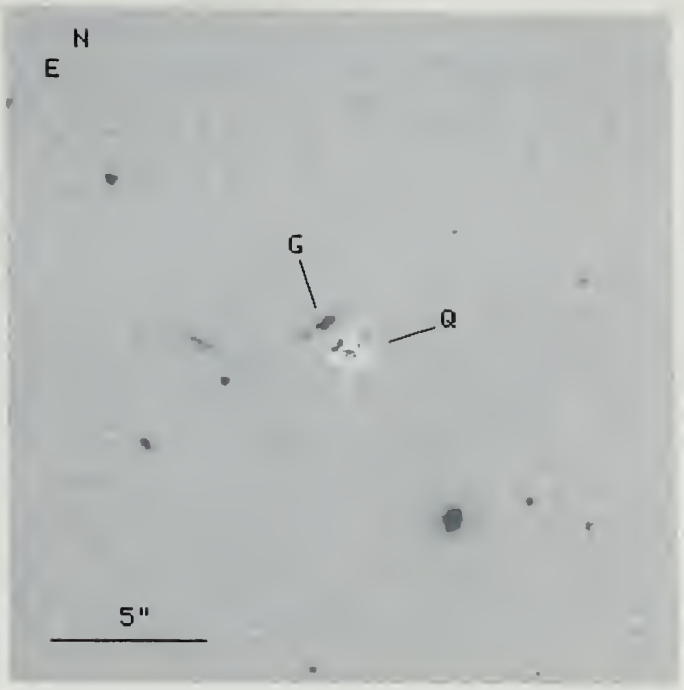

Figure 2.6 Rest-frame B-band image of the field of Q0058+019 (PHL 938) containing a DLA at $z = 0.61251$ using *HST+WFPC2* with the filter F702W. A model of the point spread function has been subtracted from the QSO image revealing a galaxy (labeled 'G') 1.2″ to the NE of the QSO position. Given the proximity, Pettini et al. (2000) argue that the galaxy is likely to be the DLA. The candidate absorber appears to be a low luminosity ($L \simeq 1/6L^*$) disk galaxy seen almost edge-on with an inclination of $i = 65\,\mathrm{deg}$, at a projected distance of 10 h_{50}^{-1} kpc from the QSO sight-line. (Reproduced from Pettini et al., 2000)

More recently, Pettini et al. (2000), for instance, detected a highly inclined, low luminosity ($L \simeq 1/6L^*$) galaxy $6h^{-1}$ kpc away from the QSO sight-line ($z_{abs} = 0.612$) also using *HST* (see Fig 2.6).

2.5.2 Early search for Lyα using narrow-band imaging

Similar, to broad-band imaging, early attempts to detect diffuse Lyα emission from DLAs at $z > 2$ using deep narrow-band imaging (e.g. Lowenthal et al., 1995) did not reveal the absorber, but unveiled a few companion Lyα emitters. In contrast, 'blank sky' Lyα searches at the time yielded null results. These results hinted at the clustering of galaxies around DLAs and prompted Wolfe (1993) to calculate the cross-correlation between galaxies and DLAs at $z = 2.6$. He came to the important

conclusion that indeed no-clustering can be ruled out, i.e. Lyα emitters are clustered around DLAs.

2.5.3 Ground-based narrow-band imaging

Ground-based surveys (photometric and spectroscopic) for Hα emission around $z > 2$ DLAs have been mostly unsuccessful (Bunker et al., 1999; Teplitz, Malkan & McLean, 1998), except in a few cases: the surveys of Bechtold et al. (1998) and Mannucci et al. (1998) have revealed multiple objects in the DLA fields. The objects are separated by more than several tens of arcseconds from the QSO line-of-sight. These Hα emitting objects appear to have a wide range of SFRs, 6–90 $M_\odot$ yr^{-1}. Kulkarni et al. (2000) suggested that the relatively large separations of these emission line objects from the quasar indicate that they are not the DLA absorbers themselves, but star-forming regions in a group or cluster also containing the DLA. The main limitation in these ground-based searches was that none have been able to probe the regions very close ($< 2''$ or 11.7 kpc at $z = 2$) to the quasar line-of-sight to rule out large spiral disks at high redshift with confidence.

In addition, ground-based observations using the WIYN telescope of three low redshift DLAs ($z < 0.3$) showed dwarf and/or low surface brightness galaxies, with confirmed redshifts (Rao & Turnshek, 1998; Lane et al., 1998).

Of at least ten DLAs at $z \gtrsim 2$ that have been searched for Lyα emission, only a very few show confirmed detections (Djorgovski et al., 1996; Fynbo, Burud, & Møller, 2000). Fynbo, Møller, & Warren (1999) reported the detection of an extended Lyα emission from a DLA at $z = 1.93$ with a flux corresponding to a SFR $\sim 10 M_\odot$ yr^{-1}. Their study, however, was focused on a DLA with absorption redshift close to the emission redshift of the QSO, $z_{\rm em} = z_{\rm abs}$, which are likely associated with the QSO host-galaxy.

2.5.4 Space-based narrow-band imaging

Since at redshifts $z > 0.5$, the Hα transition is redshifted into the infrared part of the spectrum, only the recent development of infrared detectors has made Hα searches possible at redshifts $z > 0.5$. For example, Kulkarni et al. (2000) took advantage of the high spatial resolution of *HST* with its infrared camera *NICMOS-NIC2* to search for Hα from a $z = 1.89$ DLA, and reported a possible detection of an Hα emission feature, 2–3 h_{70}^{-1} kpc in size, 0.25″ from the absorber. They suggested that a faint, compact, somewhat clumpy object, rather than a thick, spiral disk, was responsible for this DLA. The Hα flux implied a star formation rate of $4h_{70}^{-2}$M$_\odot$ yr^{-1}(3-σ upper limit).

Also using *HST/NICMOS*, we searched for diffuse Hα emission of a DLA at $z = 0.656$, a study that we present in chapter 3 (see also Bouché et al., 2001).

2.5.5 Spectroscopic surveys

Another way to search for line emission is via spectroscopy. For instance, using long-slit spectroscopy and Fabry-Perot imaging of ten DLAs, Lowenthal et al. (1995) searched for the redshifted Lyα emission. Reaching a flux limit corresponding to $0.1 - 0.7$ M$_\odot$ yr^{-1}, none of these absorbers were detected.

The spectroscopic survey of Bunker et al. (1999), which searched for redshifted Hα emission in 2.5″ slits around 6 quasars with $z > 2$ DLAs, reached a 3-σ detection limit of 6–18 M$_\odot$ yr^{-1} and also failed to detect any redshifted Hα emission.

2.6 The metallicity of DLAs and chemical evolution models

As mentioned in section 1.2, the mean metallicity of DLAs, for each epoch z, is very important as it is the closest measure of the global degree of metal enrichment of the universe reached at each time, irrespective of the nature of the absorbers.

The global rate of star formation of the universe is known to rise from $z = 0$ to $z \simeq 3$ and to peak at $1 < z < 4$ (e.g. Steidel et al., 1999). In other words, the rate of star formation peaked very early in the universe, 1-2 Gyr after the Big Bang, then declined until the present day. Naturally, since metals are produced in stars, the star formation history is directly related to the global interstellar metallicity.

Models (e.g. Pei, Fall, & Hausser, 1999) that connect this star formation history with the history of gas consumption and metal production predict that the global interstellar metallicity rises with time, from a low value at high redshifts (> 4) to a near-solar value at redshift zero (e.g. Lanzetta, Wolfe, & Turnshek, 1995; Pettini et al., 1997; Pei, Fall, & Hausser, 1999; Kulkarni & Fall, 2002), Fig. 2.7(top) shows the mean metallicity of interstellar gas as a function of redshift from the models of Pei, Fall, & Hausser (1999). If their models are in rough agreement with observations, the models predict a steady incline. At redshifts $z > 1$, however, little chemical evolution is seen in DLAs from $z = 3.5$ to $z = 1$. Whether or not the amount of metals seen in DLAs is evolving with time is still a matter of debate. However, recent studies (Savaglio, 2000; Vladilo, 2000; Prochaska & Wolfe, 2002, 2003) point to a mild increase of the metallicity with decreasing redshift.

From Fig. 2.7, one sees that the typical metallicity of DLAs is about $\sim 1/10$th of solar (with a large scatter spanning two orders of magnitude), and the chemical evolution is much less than the evolution expected from chemical evolution models.

Fig. 2.7(bottom) shows the predicted comoving density of metals from the models of Pei, Fall, & Hausser (1999). From Pettini et al. (1999); Pettini (2003), the total amount of all the metals that are observed in the intergalactic medium (in DLAs and in the Lyα forest) is $\log \Omega_m = \log[7 \times 10^{-6}] = -5.2$, and is represented by the filled square in Fig. 2.7. The amount of metals in the intergalactic medium is consistent with the predictions of Pei, Fall, & Hausser (1999). However, their model

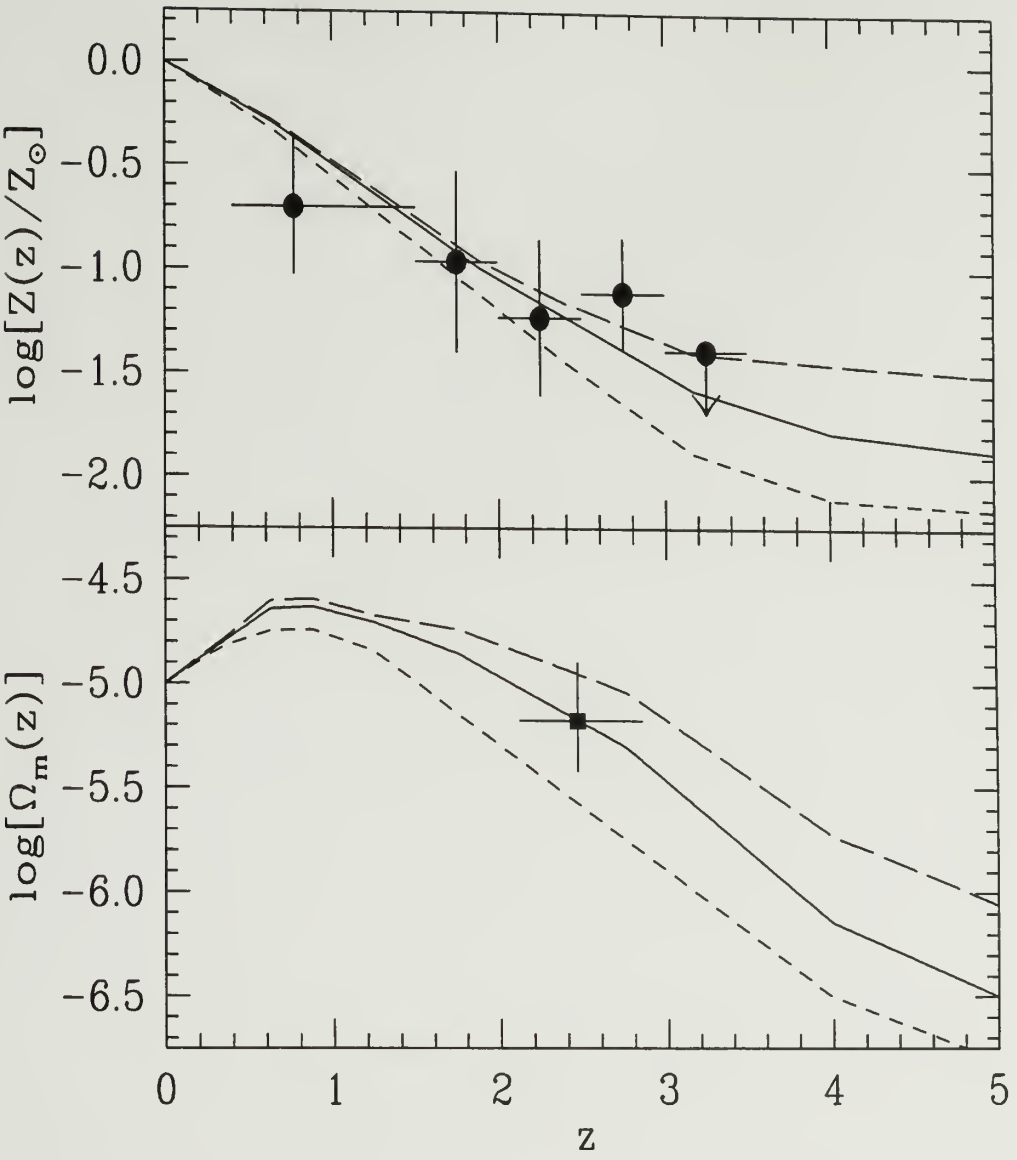

Figure 2.7 Mean metallicity of interstellar gas in galaxies in units of the solar value (upper panel) and comoving density of heavy elements in galaxies (lower panel). The solid, short-dashed, and long-dashed curves are the best, minimum, and maximum solutions of the models of Pei, Fall, & Hausser (1999), respectively. In the upper panel, the data points at $z > 1.5$ are from Pettini et al. (1997), while the data point at $z = 0.77$ is from Boissé et al. (1998). In the lower panel, the filled square is from Pettini (2003).

over-predicts the star-formation history and the amount of $\Omega_{\rm DLA}$ at $z = 3$ by a factor of 2 to 6.

If instead of using the predictions of Pei, Fall, & Hausser (1999), one uses the observed star formation history (Steidel et al., 1999), the amount of metals produced at $z = 2.5$ is expected to be 10 times more than what is seen in all absorption systems (damped and non-damped), as pointed out by Pettini et al. (1999); Pettini (2003) and others. This is sometimes referred to as the 'missing metal' problem (Wolfe et al., 2003a). One way to resolve this discrepancy is to assume that DLAs contain a large amount of metals locked in dust grains, a phenomenon called dust depletion.

But, the models of Pei, Fall, & Hausser (1999) already tend to over-produce the amount of dust. Indeed, their models predict that as much as 20% of optically selected QSOs are missed due to dust absorption. A large amount of dust would be associated with a large fraction of molecular hydrogen—because the latter forms onto dust grains—and results from Ledoux et al. (2003) indicate that it is not the case. Indeed, Ledoux et al. (2003) found that $\sim 60-70\%$ of DLAs had a molecular fraction smaller than 10^{-6}.

Other observations are also not consistent with DLAs hiding a large fraction of metals on dust grains, as shown in the next section. Because of metal depletion onto dust grains, recovering the intrinsic abundances is a complex task (e.g. Vladilo, 2002a,b; Savaglio, 2000).

2.7 The dust content of DLAs

A good tracer of dust depletion, as explained below, is the Chromium to Zinc (Cr/Zn) ratio. For historically reasons, iron has been widely used to estimate the amount of metals in the intergalactic medium. In galactic stars, the abundances of Zn and Cr follow closely the abundance of Fe (Pettini et al., 1994). The situation is very different, however, in the interstellar medium. Extensive surveys of the local

interstellar medium have shown that Zn abundances are close to the solar abundances, i.e. Zn is only slightly depleted. On the other hand, only 1% of Cr remains in the gas phase because Cr is mostly locked up in dust grains (summarized by Pettini et al., 1994). Thus, the ratio (N(Cr II)/N(Zn II)) of the abundance of this two species give a measure of dust depletion, and N(Zn II)/N(H I) measure the metallicity.

Fe is also prone to dust depletion, but not as much as Cr. Because Zn or Si are not prone to depletion, they are preferred tracers of the metals. When Zn is not available, one can attempt to correct for dust depletion of Fe. However, there is not yet a consensus about the procedure to correct for dust depletion.

The reason for this very different behavior is that dust grains are formed by condensation on oxygen silicate in the outflows of red giants. If the outflow is thermally stratified, each element will begin to condense at a certain temperature. Elements with a higher condensation temperature such as Cr will start to condense earlier and thus will be more depleted(e.g Morton, 1974). On the other hand, elements that have a lower condensation temperature such as Zn will not be as much depleted [5].

If dust depletion can be a complication in finding the 'true' metallicity of the damped absorbers, the total amount of dust in DLA appears to be small. Pettini et al. (1994, 1997, 1999) found that the amount of dust in DLAs is about 10% of the value found in the Milky Way, and deduced an extinction of $A_V = 0.02 - 0.05$ or $A_{1500} = 0.1 - 0.2$. This mild extinction is consistent with Ledoux et al. (2003) who found that $\sim 60 - 70\%$ of DLAs had a molecular fraction smaller than 10^{-6}.

In addition, current samples of DLAs towards radio-selected QSOs show no significant difference from optically-selected samples (Ellison et al., 2001). Furthermore, contrary to expectations if dust obscuration is important, Prochaska & Wolfe (2002) found no dependence of the quasar magnitude on the inferred dust opacity of DLAs.

---

[5]The origin for the very different chemical properties lies in their different electron configurations. For instance, the election configuration of Cr is [Ar]$3d^5$ $4s^1$ (a half full 3d orbital with a half full 4s).

This is consistent with the models of Boissier, Péroux, & Pettini (2003), which predict negligible dust absorption.

2.8 The connections between DLAs and LBGs

Originally, Wolfe et al. (1986) assumed that DLAs arise from HI disks associated with large spiral galaxies. Using the local luminosity function of spirals (see Appendix B.4, Eq. B.8), and assuming that disk sizes are proportional to their luminosity (known as the Holmberg relation, Holmberg, 1975), Wolfe et al. (1986) constrained the number of DLAs per unit redshift $\frac{\mathrm{d}n}{\mathrm{d}z}$ to be $\geq 1 \times 10^{-2}(1+z)^{\gamma}$ up to $z \simeq 3$, assuming a typical disk size of $R^* = 23$ kpc. Even though the prediction of $\frac{\mathrm{d}n}{\mathrm{d}z}$ is in agreement with observations (§ 2.3), it was an extrapolation of the local ($z = 0$) luminosity function to high redshifts ($z > 2$), and therefore was highly uncertain.

Fynbo, Møller, & Warren (1999) repeated the analysis of Wolfe et al. (1986), but using the luminosity function of LBGs at $z = 3$ (Steidel et al., 1999), and assuming that the HI disks (or halos) associated with LBGs give rise to DLAs. They found that a disk size $R^* = 27h^{-1}$ kpc is required to account for the observed $\frac{\mathrm{d}n}{\mathrm{d}z}$. Using the Holmberg relation (Holmberg, 1975), (see section B.4), Fynbo, Møller, & Warren (1999) inferred an impact parameter distribution $b(L)$, and argued from it that 50-80% of DLAs at $z = 3$ must be small and fainter than $m_R = 27$. This result is somewhat uncertain for two reasons. The faint end of the luminosity function is not well known, and moreover the Holmberg relation may not hold at high redshifts.

Schaye (2001) used arguments similar to Fynbo, Møller, & Warren (1999) (without using the Holmberg relation) and found that the number density of LBG (up to $m_R = 27$ or $0.1L^*$) accounts for the observed $\frac{\mathrm{d}n}{\mathrm{d}z}$ of DLAs, provided that every LBG has a physical cross section πb^2 with $b = 19h^{-1}$ kpc ($3.5''$).

All these constraints on the cross-section of DLAs extrapolate the LBG luminosity function to faint magnitudes and therefore are highly uncertain. If DLAs could be

made of the numerous galaxies populating the faint end of the luminosity function. DLAs and LBGs are much different: DLAs are metal poor ($Z \sim 1/10$), with a star formation rates $< 1 M_\odot\ yr^{-1}$, while LBGs have strong star formation rates $10 - 50 M_\odot\ yr^{-1}$ and seem to be more metal rich (Pettini, 2003, $Z \sim 1/3$). It may not be too surprising that LBGs and DLAs have distinct properties given their very different selection. LBGs are selected from magnitude limited samples, while DLAs are selected from their HI cross-section.

This picture finds theoretical support from galaxy formation models and is consistent with the strong luminosity constraints presented in § 2.5 and in chapter 3. For instance, Mo, Mao, & White (1999) argued that at least at $z \sim 3$ both DLAs and LBGs are drawn from the same parent population, but DLAs should be biased towards objects with large angular momentum, and LBGs should be small objects with substantial star formation and low angular momentum.

2.9 DLAs in theoretical models

The current galaxy formation model, sometimes referred to as the cold dark matter (CDM) model, is the following (see also Appendix A). In the very early universe, small gaussian quantum perturbations on very small scales were stretched out to cosmological scales by inflation. The small fluctuations grew under the influence of gravity. Small regions collapsed before larger ones. Gravitationally bound regions are referred to as dark matter halos. These dark matter halos lie in filaments and sheets, which constitute the large scale structure. As the dark matter collapsed, the baryonic matter fell into these halos, cooled, and settled in galaxies. Some baryons were left over from this process, such as the ones in the intergalactic medium. Most of this intergalactic medium accounts for the Lyα forest. In this scenario, DLAs could be neutral gas left-over from the galaxy formation process, or be associated with proto-galaxies.

Attempts to simulate DLAs in numerical simulations include: Katz et al. (1996a) presented the first predictions of DLA properties based on 3-D hydrodynamic simulations, and showed that a large range of structures and morphologies, rather than a single uniform type of galaxy, can give rise to a damped absorber. These results are similar to those of Haehnelt, Steinmetz, & Rauch (1998) presented in section 2.4.

Later, Gardner et al. (2001) extended the results of Katz et al. (1996a) by developing a method to compensate for the resolution limitations of their simulations. They extrapolate the relation between the absorption cross section and the halo circular velocity to lower mass halos, using a halo mass function. They found that (1) there was an anti-correlation between the H I column density and the projected distance to the nearest galaxy, with DLAs arising out to $10 - 15$ kpc; and (2) the simulations reproduced the observed abundance of DLAs if the smallest halos ($M <\sim 10^{10} M_\odot$) did not contribute to DLA.

Recently, Nagamine, Springel & Hernsquist (2003) used an improved numerical code to simulate the universe and found that their simulation can reproduce the observed statistics of DLAs ($dn/dz; f(N); \Omega_{\rm DLA}$). They claimed that their treatment of supernovae feedback leads to an estimate of the relation between DLA cross-sections and halo mass that requires no threshold.

In the semi-analytical models of Maller et al. (2000), DLAs arise from the combined effects of massive central galaxies and a number of smaller satellite galaxies within $100h^{-1}$ kpc in virialized halos, rather than only the central galaxies.

Unfortunately, none of these studies include a clustering analysis of DLAs. In Chapter 6, we present results on the cross-correlation between DLAs and LBGs in numerical simulations.

2.10 The clustering of galaxies around DLAs: previous results

Our approach to measure the clustering of galaxies around DLAs, presented in section 1.2.1, is most interesting in the following context. Mainly three hypotheses can explain the kinematics of DLAs: (1) they are thick and large massive disks (Wolfe et al., 1986, 1995; Prochaska & Wolfe, 1997); (2) they are dwarf gas rich low surface-brightness galaxies (Matteucci et al., 1997; Jimenez, Bowen, & Matteucci, 1999; Mo, Mao, & White, 1999; Boissier, Péroux, & Pettini, 2003); or (3) they arise from the combined effect of a massive central galaxy and a number of smaller satellites (Maller et al., 2000) or filaments (e.g. Haehnelt, Steinmetz, & Rauch, 1998). Both in case (1) and (3), a DLA would lie in an over-dense region because it would be associated with a massive galaxy. For instance, if the clustering of galaxies around DLAs is strong, it would be a signature for these scenarios. On the other hand, if the clustering is weak, it would favor the low mass scenario (2) [6].

Note that, (2) and (3) are most compatible with the current view of galaxy formation, outlined in the previous section and in § 6.3.1. Only (1) can readily address the asymmetric profiles, but does not fit any galaxy formation theory.

Already, several previous studies point to an over-density of galaxies near DLAs. As mentioned in section 2.5.2, Wolfe (1993) combined several studies of Lyα emitters around DLAs and found evidence for a correlation between emitters and DLAs at a mean redshift $\overline{z} = 2.6$.

Francis & Hewett (1993) reported the discovery of super-clustering of sub-DLAs at $z \sim 2.3$: A total of four H I clouds are seen in a QSO pair separated by $8'$, each pair being at the same velocity. They concluded that in order to cover both lines of

[6] A more detailed description of the expected clustering of DLAs will be presented in chapter 6 (§ 6.2.8) in the context of correlation functions.

sight, the absorber must be at least $6h^{-1}$ Mpc (for $q_o = 0.5$) or that we are seeing a 20 Mpc super cluster.

Recent results from narrow-band imaging of the Francis & Hewett field shows that spectroscopically confirmed Lyα-emitting galaxies are clustered at the redshift of the strongest H I cloud ($\log N_{\rm H\,I} = 20.1$) towards Q2138-4427 (Fynbo et al., 2003).

Other evidence of clustering include the work of Ellison et al. (2001) and d'Odorico, Petitjean, & Cristiani (2002). Ellison et al. (2001) found that the DLA at $z_{\rm abs} = 3.37$ towards Q0201+1120 is part of a concentration of matter that includes at least four galaxies (including the DLA) over transverse scales greater than $5h^{-1}$ Mpc. d'Odorico, Petitjean, & Cristiani (2002) showed that out of ten DLAs in QSO pairs, five are matching systems within 1000km s^{-1}. They concluded that this result indicates a highly significant over-density of strong absorption systems over separation lengths from ~ 1 to $8h^{-1}$ Mpc.

On the other hand, in a study similar to ours, based also on the Lyman break technique, Gawiser et al. (2001) studied the clustering environment of a DLA at $z = 4$ towards BR 0951-04 and did not find any evidence for cross-correlation between LBGs and the DLA.

Adelberger et al. (2003) reached a similar conclusion in their extensive study of the cross-correlation between QSO absorption lines and galaxies. In their combined sample of four DLA fields, they found two galaxies within $265''(5.7h^{-1}$ Mpc) and within $|\Delta z| < 0.0125$ ($\sim 8h^{-1}$ Mpc) of the DLAs whereas 5.96 were expected if the LBG-DLA cross-correlation is the same as the LBG auto-correlation. They concluded that this lack of galaxies near the four DLAs is significant at $> 90\%$ level and shows that LBGs were not clustered around DLAs on these scales. They argued that this is evidence that DLAs and LBGs "do not reside in the same parts of the universe".

It should be noted that both Gawiser et al. (2001) and Adelberger et al. (2003) are limited to scales smaller than $6h^{-1}$ Mpc due to the size of their detector. Because, the

detector needs to be larger than the clustering scale of interest, they are not sensitive to scales $> 4 - 5h^{-1}$ Mpc.

2.11 Conclusions

The lack of evolution of the gas content in DLAs and of their metallicity (Pettini et al., 1999) calls into question the idea that DLAs are unbiased tracers of global star formation. Indeed, over the redshift interval $(1 < z < 4)$, one would have expected a decrease in Ω_{DLA} as the gas is converted into stars as the observed star formation rate (Steidel et al., 1999) is constant over that redshift range.

This led Rao & Turnshek (2000) to conclude that DLAs and the galaxies that dominate the star formation density are different populations. Since LBGs are dominating the star formation at high redshifts. Furthermore. it is difficult to reconcile the low metallicities of high-redshift DLAs (typically 1/10 of solar at $z \sim 2.5$, Pettini et al., 1997) with the higher metallicities of stars in galaxies today: no global chemical evolution is seen in DLAs from $z = 3.5$ to $z = 0.3$ (Pettini et al., 1999), which indicates that DLAs do not necessarily trace the population responsible for the bulk of star formation. This is consistent with the remarks at the end of § 2.8.

It is remarkable that despite the numerous observations and the large amount of information available on DLAs, no model that can predict all the observations (kinematics, $\frac{dn}{dz}$, $\Omega_{\rm DLA}(z)$, etc.) has emerged. Many results (including ours in chapter 3) are, however, consistent with DLAs being gas rich low surface brightness. Some models require that a large fraction of DLAs be LSBs (e.g. Boissier, Péroux, & Pettini, 2003).

At this point, there are no predictions of the clustering of galaxies around DLAs. A few studies (Wolfe , 1993; Francis & Hewett, 1993; Ellison et al., 2001; d'Odorico, Petitjean, & Cristiani , 2002) showed an enhancement of galaxies near the DLAs. Two studies (Gawiser et al., 2001; Adelberger et al., 2003), however, are consistent with

no clustering on scales $<< 5h^{-1}$ Mpc. In chapter 6, we present the first theoretical predictions (§ 6.3.3) and the first measurement (§ 6.4) of the angular clustering of galaxies around DLAs on scales as large as $5 - 8h^{-1}$ Mpc.

CHAPTER 3

H$-\alpha$ IMAGING OF AN ELUSIVE DAMPED LY-ALPHA CLOUD AT $Z = 0.6$ TOWARDS QSO 3C336

In this Chapter, we present our study of a low redshift DLA at $z = 0.656$ with the *Hubble Space Telescope* and its Infra-Red camera NICMOS to search for Hα emission from star formation. As mentioned in Chapter 2 (section 2.5), regions with strong star formation will produce strong emission lines (e.g. Hα), and the strength of the emission is a measure of the global star formation rate. Here, we searched for Hα from a DLA at $z = 0.656$ using the *Hubble Space Telescope*.

This chapter resulted in the publication of Bouché et al. (2001).

3.1 Introduction

This DLA ($z_{abs} = 0.656$) towards the QSO 3C336 (z=0.927) has, so far, eluded detection entirely despite extensive ground-based searches (Steidel et al., 1997). This QSO line-of-sight is one of the richest known for $z < 1$ absorption line studies, with 6 metal line systems in the interval $0.317 < z < 0.892$. For that reason this quasar field was the target of both a very deep 24000 s *HST-WFPC2* image and a 2160 s Keck/NIRC image (Steidel et al., 1997). Five galaxies associated with the metal line systems were identified in the WFPC2 image and their redshifts confirmed spectroscopically using the Keck/LRIS (Steidel et al., 1997). The only unidentified absorber is a DLA at $z = 0.656$ with $N(\mathrm{H\,I}) = 2 \times 10^{20}$ cm^{-2} and [Fe/H] $= -1.2$. There is no galaxy detected with luminosity $L > 0.05 L_K^*$ near the QSO line-of-sight and as close as $0.5''(\sim 2h^{-1}$ kpc). Two unlikely candidate galaxies exist at large impact parameters from the QSO. The first is a relatively faint, $M_K = -21.43$ ($m_K = 20.77$), late

type spiral located 14.3″ ($\sim 65h^{-1}$ kpc) NE of the QSO. Taking into account the estimated disk inclination and the position angle, Steidel et al. (1997) estimated it would require a disk extent of at least $\sim 120h^{-1}$ kpc to intercept the QSO line-of-sight. The second candidate is a galaxy without a confirmed redshift. If one assumes this galaxy is at the redshift of the DLA, it would have an impact parameter of $41h^{-1}$ kpc and $L_B = 0.04L_B^*$ (similar to that of the SMC).

This large impact parameter raises the question of whether DLA absorption can arise in dense H I regions far from the centers of galaxies, perhaps in regions that have little or no current star formation, as seen locally in mergers (Hibbard & Yun, 1999) and as pointed out by Rao & Turnshek (2000). Alternatively, a separate absorbing galaxy could be situated beneath the QSO on the plane of the sky.

The goal of this project was to detect *any* Hα emission as close as $\sim$ 0.05″ (0.24 h^{-1} kpc) of the QSO to test further the alternate hypothesis. Hα at the redshift of this DLA matches one narrow-band filter of the *HST-NICMOS* camera 1 and, therefore, enables us to put strong constraints on the SFR of the absorber. We can already say there is no L^* spiral galaxy close to the line-of-sight, for Steidel et al. (1997) did not find anything brighter than $0.05L_K^*$. On the other hand, one might expect a dwarf or LSB galaxy with significant star formation such as found by Le Brun et al. (1997) and Rao & Turnshek (2000) for other DLAs (see section 2.5).

In the next section, we describe the observations and the data reduction. The results are given in section 3.3, and we discuss them in section 3.4. Our conclusions are presented in section 3.5. Throughout, we adopt $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 100h$ km s^{-1} Mpc^{-1}; thus 1″ at $z = 0.656$ corresponds to $4.85h^{-1}$ kpc.

3.2 Observations and data reduction

3.2.1 Observations

The observations were carried out with the Near Infrared Camera and Multi-Object Spectrometer (NICMOS) using Camera 1 (NIC1) with Camera 2 (NIC2) in Attached Parallel mode on 1997 October 3. NIC1 (0.043″ pix^{-1}) was chosen to enable us to over-sample the point spread function (PSF) (FWHM $\sim$ 2 pix or 0.09″) and hence to resolve emission as close as possible to the QSO. The exposure times were 2.8hr (5 orbits) and 0.7hr for the filters F108N and F110M respectively (details are listed in Table 3.1). The pixel size is 0.043″ which at the redshift of the DLA corresponds to a physical size of $\sim 0.21h^{-1}$ kpc pixel^{-1}.

For NIC1, we used the narrow-band (NB) F108N filter ($\lambda_c = 1.081\mu$m; FWHM = 0.0094μm) and the broad-band (BB) F110M filter ($\lambda_c = 1.101\mu$m; FWHM= 0.19μm). F108N was chosen to match the wavelength of Hα at the redshift of the DLA (z = 0.656) and the BB filter was chosen to subtract continuum emission. For NIC2, the filter F160W ($\lambda_c = 1.596\mu$m; FWHM= 0.400μm) was chosen but unfortunately, the field turned out to be in an especially empty region of the WFPC2 field (Steidel et al., 1997). The NIC2 images will not be discussed further here.

A reference star (GSC2044.810 from the STScI Guide Star Catalog) to measure the PSF was chosen within 100″ of QSO 3C336. GSC2044.810 has colors $J-H \sim 0.6$ and $H-K \sim 0.6$, which are similar to that of a typical QSO at $z \sim 1$, i.e. $B-V$=1.05 and $V-J$=1.68 (Hyland & Allen, 1982). The star (V=12.59) is $\sim$ 5.5mag brighter than the QSO, which allowed us to shorten exposure times.

Five dithered images of both the QSO and the reference star were taken in both NB and BB filters in order to improve the sampling of the PSF and to avoid any systematic noise from the detector. Each exposure was taken in the MULTIACCUM mode, which allows 25 non-destructive readouts over the entire dynamic range for

each pixel. This allows improved cosmic ray rejection. Exposures were chosen to be short enough to avoid detector saturation.

3.2.2 Data reduction

The images of the QSO and the reference star were reduced, shifted and coadded with the Nicred[1] package (McLeod, 1997). The zeroth read was subtracted from each raw image to remove any pedestal level, and dark subtraction was performed using the pipeline dark files provided by STScI. The photon arrival rate in each pixel was computed by fitting a straight line to the MULTIACCUM readouts. Cosmic rays were rejected by searching for a jump between successive readouts so that full information for each pixel is recovered. The images were then flat-fielded with the pipeline flat files provided by STScI. A sky frame was constructed from the dithered observations and subtracted from the images. Each image was magnified by a factor of 2 after masking the residual cosmic rays and bad pixels. Finally, the dithered images were registered and coadded – with a weight proportional to the inverse noise in a reference region close to the QSO image– to produce the final image (hereafter "high SNR" images). The final result improved the SNR by a factor of ~ 8 compared to images reduced by the STScI pipeline *calnica* process. The FWHM of the PSF is 4.17 pixels (in the magnified images), which corresponds to $0.090''$.

In order to perform absolute photometry on the QSO flux, i.e. to use the multiplying factors (PHOTFLAM) provided by STScI and to convert counts to flux units, we reduced the data using the dark and flat files listed as the Reference Files (used in the calibration of the P330E and G191-B2B 7691 data) prescribed by the NICMOS handbook v.3.0 (hereafter "calibrated" images). To check for any correlation noise produced by the magnification process, we reduced the data without magnifying the pixels (hereafter "small calibrated" images) with the same dark and flat files and we

[1]see http://cfa-www.harvard.edu/castles/Nicred for more information.

confirmed our noise figures (measured around the QSO) in the calibrated images. In the NB QSO field, the SNR was ~ 125 and ~ 15, respectively for the high SNR, and calnica images. In the small calibrated and calibrated images, the SNR was ~ 95. We used the high SNR images for PSF subtraction, while the noise properties were measured in the calibrated images. The calibrated image is shown in Figure 3.1.

3.2.3 Photometry and noise properties.

The total fluxes were obtained by measuring the curve of growth for each of the calibrated QSO images. We converted the fluxes measured within $0.5''$ radius aperture to nominal infinite aperture fluxes by multiplying them by 1.15 as prescribed by STScI [2].

In the NB calibrated image, the noise beyond the PSF is constant within a $2''$-radius circle (from 10 to 50 pixels) around the QSO. Beyond a radius of $2''$, the noise increases due to the poorer first quadrant of the NICMOS detector.

Our 3-σ detection limit for a point source is given by the 3-σ rms per resel (a resel or resolution element is a $0.09''$ diameter aperture, corresponding to the FWHM of the PSF) measured in the quadrant of the image that includes the QSO. For the NB image, our 3-σ detection limit is: $F_\lambda = 3.78 \times 10^{-19}$ erg s^{-1} cm^{-2} A^{-1}($m_{AB} = 23.48$, where $m_{AB} = -2.5 \times \log(F_\nu) - 48.6$), or a flux of $F_{\mathrm{H}\alpha} = 3.70 \times 10^{-17}$ erg s^{-1} cm^{-2}. For the BB image, our 3-σ detection limit is: $F_\lambda = 2.9 \times 10^{-20}$ erg s^{-1} cm^{-2} A^{-1}($m_{AB} = 26.22$), or a flux of $F_{BB} = 5.82 \times 10^{-17}$ erg s^{-1} cm^{-2}.

Our 3-σ detection limits for an extended source are given by the 3-σ rms per pixel scaled by the square root of the number of pixels in a $1''$ by $1''$ square. The resulting limits are 4.70×10^{-18} erg s^{-1} cm^{-2} A^{-1}($m_{AB} = 20.75$) in the NB image and 3.60×10^{-19} erg s^{-1} cm^{-2} A^{-1}($m_{AB} = 23.49$) in the BB image. These correspond

---

[2] see http://www.stsci.edu/cgi-bin/nicmos/ under *documents* and *handbook*. See also Figures 4.6 to 4.10 from the NICMOS Instrument Handbook v3.0 Chapter 4.

to surface brightnesses of 4.60×10^{-16} erg s^{-1} cm^{-2} arcsec^{-2} for the NB image and 7.17×10^{-16} erg s^{-1} cm^{-2} arcsec^{-2} for the BB image.

The total flux densities of the QSO in the NB and the BB are $(4.55 \pm 0.18) \times 10^{-17}$ erg s^{-1} cm^{-2} A^{-1}($m_{AB} = 18.3$) and $(5.09 \pm 0.17) \times 10^{-17}$ erg s^{-1} cm^{-2} A^{-1}($m_{AB} = 18.1$), respectively. These results are summarized in Table 3.2.

3.2.4 Profile subtraction

In order to reveal any faint object with small impact parameter, we subtracted the QSO PSF in the following ways:

To search for faint emission both close to the QSO line of sight and throughout the 11″ field, we first subtracted the BB QSO image (scaled to the peak value) from the NB QSO image. The central part, shown in Figure 3.2 (a), has faint residuals (negative and positive peaks well below 3σ) near the QSO position. The total residual flux measured in a resel centered on the QSO is about 1.5σ above the mean or 0.7% of the (unsubtracted) NB QSO flux in the same aperture.

To assess how much of those residuals might be due to differences in PSF between the F108N and F110M filters and how much to real Hα flux from a dwarf galaxy exactly superposed on the QSO position, we used the same procedure (i.e. shifting & centering) to subtract the BB PSF from the NB PSF of the reference star (GSC2044.810). We find residuals with a similar pattern (see Figure 3.2 (b)). There is clearly a peak near the center with a depression above and below (in the y-direction). The stellar PSF residual pattern on larger scales is not seen in the QSO PSF subtracted image, since the SNR is more than ten times lower. The central peak pixel in the residuals is about 1.5% the flux of the stellar PSF peak, while the magnitude of the deepest depression is about -5% of the stellar PSF peak. The total residual flux measured in a resel centered on the position of the star is 0.7% of the NB

stellar PSF flux in the same aperture, consistent with the QSO residuals. Therefore, we concluded the residuals are likely due to PSF differences between the two filters.

Because the stellar NB - BB method described above produced such significant residuals, we then tried subtracting a stellar PSF (scaled to the peak) directly from the QSO high SNR image. This was performed for both the NB and BB filters. We used both a theoretical PSF generated by the software Tiny Tim[3] and the reference star (GSC2044.810) PSF. Unfortunately, the reference star turns out to be double, i.e. after the subtraction, a PSF-like hole was seen in both NB and BB images offset by 0.17″ from the central PSF (see Figure 3.3 (a)). We corrected for this by subtracting the primary component of the star PSF from the secondary component, and then subtracted the result from the QSO PSF. The final result shows very little residual (see Figure 3.3 (b)). However, when using the theoretical PSF instead, the result tends to leave a ring-like structure. For this reason, we adopted the stellar PSF as the best approximation of the true PSF. The final PSF subtracted NB image is shown in Figure 3.3 (b) after a gaussian smoothing of 1.5 pixels.

Finally, in the NB PSF subtracted image, we looked for faint emission throughout the 11″ field both by eye and using the algorithm SExtractor (Bertin & Arnouts, 1996) with a 3-σ threshold (with a minimum of 16 pixels above threshold, i.e. the keyword MINAREA in SExtractor). Three candidates were found. Every candidate was followed up by examining individual images and it turned out that all candidates were artifacts left over from the reduction process, e.g. cosmic ray residuals smeared out over several pixels by the magnification process.

---

[3]adapted by Richard Hook for NICMOS. The program and informations are available online at http://www.stsci.edu/software/tinytim.

3.3 Results

To summarize, we looked for emission objects near the QSO and throughout the field in the following way: (i) We subtracted the BB image from the NB image. Apart from residuals due to PSF differences between the two filters, no emission was detected; (ii) We subtracted the star PSF from the QSO PSF in both NB and BB images. No emission was detected; (iii) We searched by eye for any faint emission throughout the 11″ field in both the NB PSF subtracted and the 'NB minus BB' images; (iv) In the NB PSF subtracted image, we also used the algorithm SExtractor (Bertin & Arnouts, 1996) to look for faint emission. All three candidates were cosmic rays residuals smeared out over several pixels. Therefore, we conclude that no emission objects were detected in either the NB PSF subtracted, or in the 'NB minus NB' image.

We can use the lack of detection in the BB image to constrain the presence of luminous galaxies. The BB 3-σ detection limit for a point source is $m_{AB} = 26.22$, which corresponds to a continuum luminosity of $L_R = 5.02 \times 10^{40} h^{-2}$ erg s^{-1} in the rest frame R–band or $L_R = 7.23 \times 10^{7} h^{-2}$ L$_\odot$.

In addition, our lack of detection in the NB constrains directly the SFR. For the NB filter, our 3-σ detection limit for a point source (i.e. unresolved) corresponds to an Hα luminosity of $3.20 \times 10^{40} h^{-2}$ erg s^{-1} at the redshift of the DLA. The 3-σ detection limit for an extended source corresponds to an Hα luminosity of $3.98 \times 10^{41} h^{-2}$ erg s^{-1}arcsec^{-2} at the redshift of the DLA (see Table 3.2). Using the Kennicutt (1983) conversion factor for a constant SFR and a modified Salpeter–like IMF with variable slope, i.e. $SFR = L_{\mathrm{H}\alpha}/1.12 \times 10^{41}$ erg s^{-1}, we derive a SFR of $< 0.28 h^{-2}$M$_\odot$ yr^{-1} for an unresolved source or < 0.15 M$_\odot$ yr^{-1} kpc^{-2} for an extended source. Assuming the absorber is a disk of radius 2 kpc (i.e. ~ 9 times the resolution element; see discussion below), this gives a SFR of < 1.87 M$_\odot$ yr^{-1}.

A more recent calibration by Kennicutt (1998) yields a similar conversion factor $SFR = L_{\mathrm{H}\alpha}/1.26\times10^{41}$ erg s^{-1}. Our upper limit is a conservative one given that the actual SNR in the magnified images used in the subtraction process is higher than the calibrated images and that the conversion factor is the smallest of current estimates. The SFR estimate is also dependent on the IMF. Using a Salpeter IMF with a higher mass cutoff (125 $M_\odot$), Alonso-Herrero et al. (1996) found $SFR = L_{\mathrm{H}\alpha}/3.1\times10^{41}$ erg s^{-1}, which is about 3 times larger than the Kennicutt (1983) result. This would *decrease* our upper limit by a factor of three, i.e. a $SFR < 0.6$ $M_\odot$ yr^{-1} for a 2 kpc-radius disk.

This estimate depends strongly on the assumed size of the object. However, from Steidel et al. (1997), there can not be any $L > 0.05L^*$ galaxy (typically 10 kpc in size) as close as $0.5''$ ($2h^{-1}$ kpc). In other words, anything larger than 4 kpc (in diameter) would have been seen in both Steidel et al. (1997) and in our images. A 2 kpc-radius object is consistent with the size of the DLA candidate ($z = 1.89$) of Kulkarni et al. (2000). Similarly, Le Brun et al. (1997; see section 2.5) detect compact objects with sizes of 1–3 kpc along DLA lines-of-sight. On the other hand, if the absorbing object is smaller, it could have a higher SFR. This would require that it be smaller than 1 resel or ~ 0.5 kpc, and that it be exactly aligned with the line of sight of the QSO.

3.4 Discussion

Previous constraints of the star formation rates of DLAs were presented in section 2.5.

The nature of the DLA towards 3C336 can be addressed by comparing directly our SFR upper limit to various types of galaxies in the local universe. (i) *LSB galaxies* have H I surface densities of ~ 5 $M_\odot$ pc^{-2} or $N_{\mathrm{H\,I}} \sim 6.5\times10^{20}$ cm^{-2} (van der Hulst et al., 1993), well above the DLA cutoff of 2×10^{20} cm^{-2} but below the value of 10^{21} cm^{-2} usually quoted as the threshold for star formation. The mean star formation rates,

derived from the Hα luminosities, in LSBs are typically ~ 0.1 M$_\odot$ yr^{-1} (van den Hoek et al., 2000). (ii) *Blue compact dwarf galaxies (BCDG)*: From CFHT observations, Petrosian et al. (1997) reported a SFR of 0.3–0.5 M$_\odot$ yr^{-1} kpc^{-2} for the two main Hα emitting regions of the BCDG, IZW 18 at a distance of $D = 10$ Mpc; (iii) *Typical spiral galaxies*: The typical SFR for a spiral galaxy is ~ 10 M$_\odot$ yr^{-1} (van den Hoek et al., 2000), spread over at least several kpc; (iv) *Starbursts* can have a much larger SFR, e.g. Arp 220 is forming stars at a rate of ~ 240M$_\odot$ yr^{-1}, derived from Lyα (Anantharamaiah et al., 2000). The SFR in the DLA at $z = 0.656$ toward 3C336 is far less extreme than in a starburst or a typical spiral, and somewhat less extreme than a BCDG. However, it is consistent with that of an LSB galaxy and it could be even lower, i.e. zero.

Another example of a region of neutral hydrogen that exceeds the DLA threshold, but does not have significant star formation, has been found through 21-cm observations of the giant H I cloud 1225+0146 (Giovanelli & Haynes, 1989); these data show that neutral gas structures with little or no star formation do exist. This suggests that factors beyond a simple column density threshold govern the formation of stars. This giant H I cloud is 200 kpc along its major axis and has two peaks of H I emission, with N(H I) $= 2 \times 10^{20}$ and 1×10^{20} cm^{-2}, separated by 100 kpc. In a deep optical search, only a faint $M_B = -15.5$ dwarf irregular, 5 kpc in extent, was discovered, corresponding to the largest peak in emission (McMahon et al., 1990). Furthermore, no galaxy has been detected near the 10^{20} cm^{-2} peak, which is almost enough neutral hydrogen to produce a DLA.

The absorption line properties of the DLA towards 3C336 corroborate the observed low SFR. This DLA is unusual in that it is a "C IV–deficient" Mg II absorber — rest frame equivalent width W(C IV) ~ 0.5Å (Steidel et al., 1997) —, i.e. W(C IV) is less than half the typical value for DLAs. Churchill et al. (1999) found a correlation between the strength of C IV absorption and the kinematic spread of the Mg II profile

in high resolution absorption profiles, with the exception of several C IV–deficient absorbers. They hypothesize that there is a relationship between the strength of C IV absorption (which is generally consistent with arising in a corona similar to that around the Milky Way disk) and the level of star formation activity in the disk. The kinematic spread of the Mg II profile is also thought to be related to star forming processes that either eject or are triggered by high velocity Mg II clouds. In this scenario, the C IV deficient absorbers would be characterized by a lower than average star formation rate, and in fact a few of them do have red colors (Churchill et al., 2000), rather than the blue colors of actively star-forming systems. This is consistent with the fact that the 3C336 DLA has both a small W(C IV) and a strong limit on the star formation rate in its vicinity.

There is at least some theoretical reason to expect low SFRs from DLAs. Mo, Mao, & White (1999) hypothesize that — at least at $z \sim 3$ — Lyman break galaxies (LBGs), which are selected partly by their strong star formation, and DLAs could be disjoint populations: If LBGs are the central galaxies of massive dark halos at $z \sim 3$, then they should be small objects with substantial star formation but low angular momentum, while DLAs should be biased towards objects with large angular momentum. In the hierarchical structure formation models of Maller et al. (2000), meanwhile, DLAs arise from the combined effects of massive central galaxies and a number of smaller satellite galaxies in virialized halos, rather than only the central galaxies, so predicted SFRs associated with DLAs are low. Another interesting result from Maller et al.'s models is that the impact parameter distribution has a longer tail at $z \sim 1$ (up to 150 kpc) than at redshift $z \sim 3$, which could then reconcile the observed galaxy 120 kpc away from the line-of-sight of 3C336 and the $z = 0.656$ absorber.

3.5 Conclusions

Nearly all space-based observations of low redshift DLAs have revealed star formation in some sort of galaxy (see section 2.5). The $z = 0.656$ system toward 3C336 is an exception. Our lack of detection of any source with SFR greater than $0.28h^{-2}$ $M_\odot$ yr^{-1}, or 0.15 $M_\odot$ yr^{-1} kpc^{-2}, demonstrates that DLAs with very little star formation can exist.

Our results and the one presented in section 2.2, argue against the standard, DLA/H I disk scenario (see section 1.2). Thus, we must address the question: "What is a DLA?". Possibilities not yet ruled out include knots of H I hundreds of kpc away from the main galaxy as seen locally (Hibbard & Yun, 1999); small compact dwarfs or LSBs; or photo-dissociation regions that may produce much of the H I currently observed in galaxy disks (Smith et al., 2000). In some cases, both a cold (hundreds of K) and a warm (thousands of K) neutral medium are found along the line of sight through a DLA, based upon analysis of 21 cm emission profiles (Lane, Briggs, & Smette, 2000), while in others the warm phase dominates.

Our limits on Hα emission from the DLA towards 3C336 set the tightest constraints yet on star formation in a compact absorber at intermediate redshift, and probe closer to the QSO line-of-sight than any previous study. Our non-detection adds to the mounting evidence that low redshift DLAs are made of galaxies of diverse morphologies, luminosities and surface brightnesses rather than a uniform population of luminous disk galaxies.

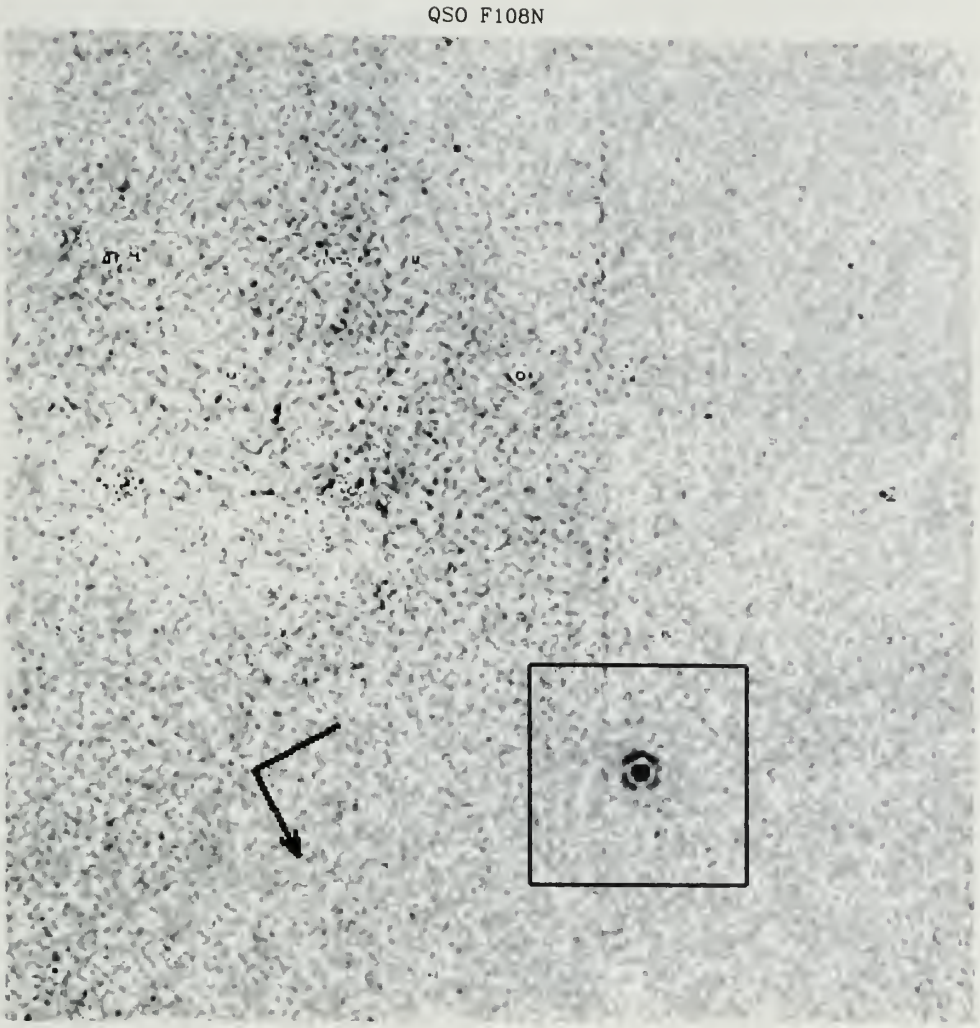

Figure 3.1 The calibrated NIC1 Narrow Band F108N image of the 11″x11″ field around QSO 3C336. The box is 2″x2″ and centered on the QSO. The arrow indicates North. The image is shown in reverse: positive flux is shown in dark.

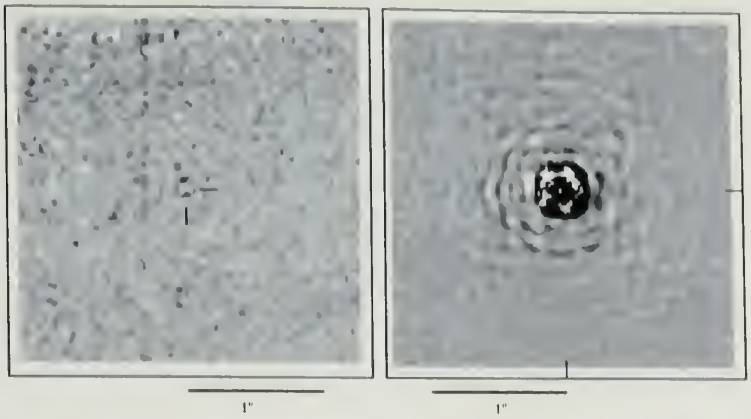

Figure 3.2 (a) Left: Narrow band NIC1 image of the QSO field after subtraction of broad band NIC1 image. Images were registered and scaled to the peak value prior to subtraction. Residual flux in 1 resel at the QSO position is only 1.5σ above the mean. The original position is shown by the tick marks. (b) Right: same as (a) for the reference star. Note the significant residual flux pattern, which is consistent with the QSO residuals, although at much higher signal-to-noise ratio. In both images, positive flux is shown in white.

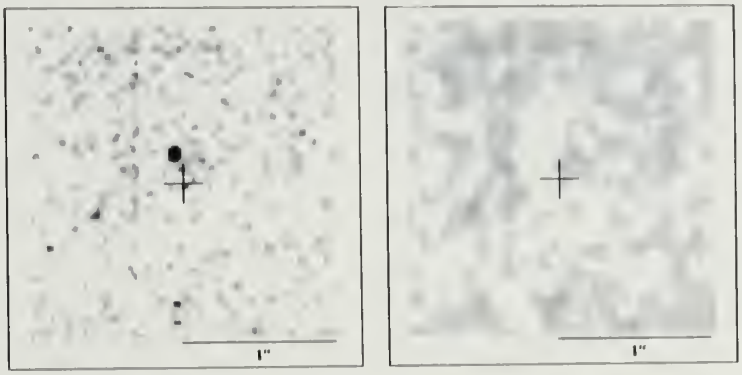

Figure 3.3 (a) Left: Close up of a 2″ x2″ region around the QSO after subtraction of the scaled star PSF. The cross marks the position of the QSO. The star used in the PSF subtraction turns out to be a close double, resulting in a "hole" 0.17″ from the QSO. (b) Right: Same as (a) after second PSF subtraction to remove the stellar companion. Image has been smoothed to 1.5 pixel resolution to enhance any extended emission. No residual flux is detected. (a) & (b) are the "high SNR" images and positive emission is in white.

Table 3.1. Exposure times for each filter.

Object	F108N (s)	F110M (s)
3C336	5×2050s	5×511s
GSC2044.810	5×303s	6×23s
GSC2044.810		5×14s

Table 3.2. Summary of the results in QSO 3C336 field.

Parameter	Value
QSO (z=0.927)	
R.A., decl. (J2000)	$16^h24^m39^s.13, +23°45'11.7''$
F_λ (1.08 μm) (erg s^{-1} cm^{-2} Å^{-1})	$(4.55 \pm 0.18) \times 10^{-17}$
$m_{AB,1.08\mu m}$ (mag)	18.28 ± 0.1
F_λ (1.10 μm) (erg s^{-1} cm^{-2} Å^{-1})	$(5.09 \pm 0.17) \times 10^{-17}$
$m_{AB,1.10\mu m}$ (mag)	18.11 ± 0.1
$3-\sigma$ Upper Limits on Continuum Emission at $z = 0.656$	
F_λ (1.10 μm) (erg s^{-1} cm^{-2} Å^{-1})	$< 2.92 \times 10^{-20}$
$m_{AB,1.10\mu m}$ (mag)	< 26.22
F_{BB} (Unresolved source) (erg s^{-1} cm^{-2})	$< 5.82 \times 10^{-17}$
L_R (Unresolved source) (h^{-2} erg s^{-1})	$< 5.02 \times 10^{40}$
$3-\sigma$ Upper Limits on Hα Emission from the Damped Lyα Cloud (z =0.656)	
F_λ (1.08 μm) (erg s^{-1} cm^{-2} Å^{-1})	$< 3.78 \times 10^{-19}$
$m_{AB,1.08\mu m}$ (mag)	< 23.48
$F_{H\alpha}$ (point source) (erg s^{-1} cm^{-2})	$< 3.70 \times 10^{-17}$
$L_{H\alpha}$(point source) (h^{-2} erg s^{-1})	$< 3.20 \times 10^{40}$
$\Rightarrow$ SFR (point source) (h^{-2} M$_\odot$ yr^{-1})	$< 0.28\ h^{-2}$
$\mu_{H\alpha}$ (3σ) (erg s^{-1} cm^{-2}arcsec^{-2}) [a]	$< 4.60 \times 10^{-16}$
$\Sigma_{H\alpha}$ (h^{-2} erg s^{-1} arcsec^{-2}) [b]	$< 3.98 \times 10^{41}$
$\Sigma_{H\alpha}$ (erg s^{-1}kpc^{-2}) [b]	$< 1.67 \times 10^{40}$

Table 3.2—Continued

Parameter	Value
QSO (z=0.927)	
$\Rightarrow$ Surface SFR ($M_\odot$ yr^{-1}kpc^{-2})	< 0.15
$\Rightarrow$ SFR (r=2kpc disk) ($M_\odot$ yr^{-1})	< 1.87

[a]Hα surface brightness.

[b]Hα surface luminosity.

CHAPTER 4

THE LBG/DLA CONNECTION I. DATA

The following three chapters will be devoted to our investigation of the clustering of galaxies around four DLAs at redshift $z \sim 3$. Chapter 4 describes the selection of the four fields, the observations, data calibration and the completeness of our sample. Chapter 5 describes our method to select galaxies in redshift slices. Chapter 6 presents the results of our clustering analysis.

Section 4.1 presents the four fields of study and what is already known about them. The following sections present our data and show the necessary steps between raw data and photometrically calibrated galaxy catalogs.

4.1 Field selection

Given that there are only about 300 DLAs known over the entire sky (Curran et al., 2002), and given the allocated telescope nights for our project, we optimized this precious telescope time by selecting fields with a DLA at $z \sim 3$ that were well visible throughout those nights and that had a low Galactic extinction. The redshift range $z \sim 3$ was particularly useful because we could use standard photometric filters to select LBGs efficiently as explained in chapter 5. Another important constraint in selecting our fields was that the QSO must be at a higher redshift than that of the DLA, i.e. $z_{\rm DLA} << z_{\rm QSO}$. This ensured that both the QSO and the DLA were physically unrelated. This was important since radio loud QSOs are known to lie in overdense regions (e.g. Ellingson, Yee, & Green, 1991).

Below, we describe the four DLA fields that we selected for this study.

4.1.1 The APM 082790+5255 field

This broad absorption line (BAL) quasi-stellar object (QSO) at redshift $z_{\rm QSO} = 3.91$ was discovered serendipitously during a survey of Galactic halo carbon stars by Irwin et al. (1998). Adaptive optics imaging (Ledoux et al., 1998b) with the Canada France Hawaii Telescope (CFHT) revealed two components (A&B) separated by 0.378″. Follow-up observations with HST/NICMOS (Ibata et al., 1999) confirmed the two components and revealed a third component. Spatially resolved HST/STIS spectra by Lewis et al. (2002) showed that the three images were lensed images of the QSO. The optical source was coincident with a strong IRAS source (0.6Jy at 60μm) and this QSO is sometimes referred to as 'IRAS 08279+5255'. With a flux of 75mJy at 850μm (Lewis et al., 1998), this QSO has a total luminosity (L) exceeding 10^{15} solar luminosities ($L_\odot$). Even taking into account its gravitational lens magnification of $10-100\times$, this QSO ranks among the most luminous objects in the Universe (e.g. Irwin et al., 1998; Lewis, Robb, & Ibata, 1999). The favored lensing model associate the lens with one or the two strong MgII absorbers at $z = 1.18$ and $z = 1.81$ (e.g. Ibata et al., 1999).

The DLA at $z_{\rm DLA} = 2.974$ was first reported by Petitjean et al. (2000) based on a high resolution spectrum (6 km s^{-1}) of this QSO obtained at the 10-meter Keck telescope with the High Resolution Echelle Spectrograph by Ellison et al. (1999a,b). Due to uncertainty in the continuum, the DLA's HI column density ($N_{\rm HI}$) was poorly constrained: $19.8 < \log N({\rm HI}) < 20.3$. The absorber could therefore technically be a 'sub-DLA'. However, the Lyα rest-frame equivalent width $W_{\rm rest}({\rm HI}) > 4.8$Å (Petitjean et al., 2000) and its strong MgII equivalent width $W_{\rm rest} > 0.6$Å (Kobayashi et al., 2002) supported its classification as a damped cloud.

Since the Lyα absorption line was black over 15Å, the cloud must cover the three lensed images, and Petitjean et al. (2000) estimated that the cloud must be larger than 200 h_{75}^{-1} pc (in a cosmology with no cosmological constant, i.e. $\Omega_M = 1$). Using

the geometry of the situation and the most current cosmological parameters (from the WMAP results of Bennett et al., 2003, presented in the introduction), we found a physical size of $218h^{-1}$ pc or $290h_{75}^{-1}$ pc. The most robust estimate of its size is from Tzanavaris & Carswell (2003) who found that the DLA must be $> 150h^{-1}$ pc (assuming the gravitational lens is at $z \sim 1$).

This DLA has a very low metallicity with iron abundance [Fe/H]=-2.31 (Petitjean et al., 2000). This is at the lowest range of values of DLAs (see Fig. 2.7). Zn, which is less subject to depletion on dust grains than iron (see § 2.7), is not detected [Zn/H]< -1.1. From Petitjean et al. (2000), the Fe abundance is, however, consistent with the O and C abundances: [C/H]=-2.35 and [O/H]=-2.26. This indicates that the amount of dust in this system must be very small. Indeed, as explained in section 2.7, if there was a large amount of dust, Fe would be more depleted than O, and C since Fe has has a higher condensation temperature. In that case, the abundances of Fe compared to O would differ significantly, but they do not.

This low metallicity is consistent with the results of Kobayashi et al. (2002), based on FeII and MgII doublet spectra (which lie in the near-IR at the redshift of the DLA $z = 2.974$). Kobayashi et al. (2002) found a metallicity [Fe/H] of -2.3. The metal properties of this DLA are summarized in Table 4.1.

In summary, this DLA is a very metal poor with little dust content, indicative of a 'younger' phase than other DLAs.

Table 4.1. Metal lines associated with the APM 08279+5255 DLA.

Specie	$\lambda_{\rm rest}$	W_o	W_r(A)	$\log N$ (cm^{-2})	[X/H]	Refs
C II	λ1334	...	...	14.5	-2.35[a]	1
O I	λ1302	...	...	...	-2.26[a]	1
Si II	λ1526	...	...	...	-2.10[a]	1
C IV	λ1550	...	...	13.1		3
Fe II	λ1606	...	...	...	-2.31[a]	1
Zn II	λ2062	...	...	...	< -1.1[a]	1
Fe II	λ2587	0.53	0.13	13.3	-2.3	2
Fe II	λ2600	0.97	0.24	13.3	-2.3	2
Mg I	λ2853	3.44	0.07	11.7	> -2.6	2
Mg II	λ2796	2.46	0.87	> 13.3	> -2.4	2
Mg II	λ2803	0.29	0.62	> 13.5	-4.2	2

References. — (a)$\log N_{\rm H\,I} = 20.3$ is assumed; (1) Petitjean et al. (2000); (2) Kobayashi et al. (2002); (3) Tzanavaris & Carswell (2003); (4) Ellison et al. (1999b) .

Table 4.2. Properties of the four DLAs and the QSO.

	APM 08279+525	PC 1233+4752	Q2342+3417	J0124+0044
R.A. (J2000)	$08^h31^m41.6^s$	$12^h35^m31.1^s$	$23^h44^m51.2^s$	$01^h24^m03.8^s$
Dec. (J2000)	52° 45′ 17″	47° 36′ 06″	34° 33′ 49″	00° 44′ 33″
l	165° 45′ 17″	130° 32′ 12″	107° 20′ 20″	139° 58′ 19″
b	36° 14′ 25″	69° 17′ 27″	−26° 19′ 38″	−61° 02′ 43″
A_U [a](mag) ..	0.20	0.08	0.32	0.13
E_{B-V}	0.04	0.02	0.06	0.03
		QSO Properties		
QSO	BAL		BAL	
z_{QSO}	3.81	4.447	3.057	3.840
m_R	15.2[1]	20.63[1]	18.6[1]	17.9[1]
2MASS	K=12	...	K=14.2	K=15.7
IRAS	$S_{60\mu m} = 0.51$Jy [2]	...	...	Y

Continued, next page.

Table 4.2—Continued

	APM 08279+525	PC 1233+4752	Q2342+3417	J0124+0044
Radio	$S_{20cm} = 1.3$mJy [3]	$S_{6cm} < 77\mu$Jy [4]	$S_{20cm} = 133$mJy [5]	$S_{20cm} = 0.11$mJy [6]
		DLA properties		
z_{DLA}	2.974	3.499[5]	2.908[5]	3.077[9]
H I W_r (Å) ..	> 4.8[7]	4.22[5]	8.18[5]	...
$\log N_{HI}$ (cm^{-2})	< 20.3[7]	20.9[5]	21.1[5]	20.1[9]
[Fe/H]	-2.31[7]	...	-1.62[8]	...
Z	< -1.1 [7]	...	-1.19[8]	...

[a]Galactic extinction from Schlegel et al. (1998), averaged over the field.

References. — (1) Veron-Cetty & Veron (2001); (2) Irwin et al. (1998) (3) McMahon et al. (2002) (4) Schneider, Schmidt, & Gunn (1991) (5) White, Kinney, & Becker (1993); (6) Carilli et al. (2001) (7) Petitjean et al. (2000); (8) Prochaska & Wolfe (2003) (9) Péroux (2003) .

4.1.2 The PC1233+4752 field

This QSO at $z_{\rm em} = 4.447$ was until the beginning of the nineties one of the highest redshift quasars known. Schneider, Schmidt, & Gunn (1991) first reported this DLA from their medium resolution spectroscopy survey of high redshift quasars. The column density is $\log N_{\rm H\,I} = 20.9$ (White, Kinney, & Becker, 1993).

4.1.3 The Q2342+3417 field

This QSO is a powerful radio source, it is 'radio bright'. The DLA was first reported by White, Kinney, & Becker (1993). It has a large observed equivalent width of $W_o(\mathrm{Ly}\,\alpha) = 32$Å, which is due to its high column density $\log N_{\rm H\,I} = 21.1$.

Since the QSO is a strong radio source, Carilli et al. (1996) attempted to detect H I (21cm) in absorption using the Westerbrook Synthesis Radio Telescope (WSRT), but failed to detect any H I absorption despite twelve hours of integration. A more recent attempt to detect H I absorption has also failed (Kanekar & Chengalur, 2003, and reference therein).

This field had the highest galactic extinction A_U since it had the lowest galactic latitude $b = -25\,\mathrm{deg}$. This field was included in our sample because its galactic extinction A_U was mild relative to other QSO fields that were available during run II.

4.1.4 The J0124+0044 field

This QSO was discovered more recently by the Sloan digital sky survey (SDSS) (Fan et al., 1999). The redshift of the absorber $z_{abs} = 3.08$ was published in Fan et al. (1999) and its column density was about 20.1 (Péroux , 2003).

The published information on the four QSOs and the four DLAs are listed in Table 4.2.

4.2 Optical imaging

> Astronomers take beautiful images of the heavens and turn them into ugly numbers.
>
> —author unknown

To identify galaxies at a redshift near that of the DLA, we needed to use multi-band imaging. At redshift $z \sim 3$, star forming galaxies are easily selected by the strong Lyman break (see § 2.1). At that redshift, the break occurs at an observed wavelength of about 3500Å, so we needed to sample at lower and higher wavelengths in order to select galaxies with such a break. This will constrain the redshift. Thus, a minimum of three filters is necessary. It turns out that the U and B filters sample the Lyman break. Here, we imaged each of our four fields in four filter bands, U, B, V & I.

The observations had to reach a depth of $m_I = 26$ in order to allow the selection of galaxies at redshifts $z \sim 3$, which are typically $m_I = 23 - 25$. A telescope with a mirror larger than 4 meters can achieve this type of depth with reasonable exposure times. In addition, since we were ultimately looking at the clustering of galaxies, a large field of view was important. This was provided by the camera, MOSAIC, used to make the observations.

Section 4.2.1 describes our observations and data reduction. Section 4.3 presents our photometric calibration. Section 4.4 shows our completeness and the optimization of our detection limit. Section 4.5 describes the galaxy detection algorithm. Section 4.6 describes how we applied the photometric calibration to our data. Finally, section 4.7 presents the first selection of one third of our catalog using color cuts. This was necessary in order to save on computing time in the next step. This next step is the subject of chapter 5, which describes the use of photometric redshift techniques. Photometric redshifts will enable us to perform the final selection of our galaxy catalog.

4.2.1 Observations

The observations, summarized in Table 4.3, were carried out with the MOSAIC camera at the 4-m telescope of the Kitt Peak National Observatory in two runs: on UT 2000 February 7 and 8 (run I), and on UT 2000 September 23-26 (run II). The five nights were photometric, although we reached a depth slightly less than our goals ($26m_{AB}$ in the I-band) because the seeing was larger than one arcsecond, not as good as expected (≤ 1 arcsecond).

In addition, we obtained near-IR $K_s(2.17\mu\mathrm{m})$ data of the two fields of run I using the IRIM imager at the 4-m telescope of the Kitt Peak National Observatory. We reached a depth of $K_s \sim 22$ (3σ). Unfortunately, in one field, the data was not useful because of strong fringing. Furthermore, the size of the IRIM detector is only $150''$, so it turned out that at the most one or two of our galaxies (selected in § 5.6) fell onto the detector. Most importantly, in chapter 5, we show that IR imaging was not necessary.

The wide field imager MOSAIC had eight 2k×4k thinned SITe coupled charge devices (CCDs). The thinning process made them more sensitive in the U, which was crucial for our study. With 0.258 arcsec per pixel, the camera had a field of view of 36 arcmin on a side, and covered a field of view as large as the full Moon. This is a very large field of view for CCDs in astronomy. The readout noise was $\sim 6e^-$ pixel^{-1}, the dark current was negligible ($\sim 5e^-$ hr^{-1}), and the average gain was $3e^-$ ADU^{-1}. Even though each CCD had been thinned for detecting U-band photons, the quantum efficiency still fell off rapidly blue-ward of 4000Å. This explains why the exposure time was significantly larger in the U-band.

We imaged each field using four broad-band filters, U (Stromgren), and BV& I (Harris set). The filter transmission curves are shown in Fig. 4.1 and their central wavelength and width are summarized in Table 4.4.

We used a standard dither pattern of five pointings to cover the small gaps (50pixels) between the CCDs. The total integration times for each field was typically 4hr(U), 1hr(B & V) and 2.5hr(I) and are shown in Table 4.3. The longest exposures were 5-15min per frame. The exposure time depended mainly on the level of the sky background, and on the faintest star that would saturate the detector. Typically 10 frames were necessary to ensure proper cosmic ray removal and good flat fielding. In total, there were about 7-15 frames per filter.

In addition, we observed several calibration fields, known as the Landolt fields (Landolt, 1992). Two to three, and five to six fields were observed for run I and run II, respectively. These observations were made through each filter, usually several times throughout the night. Normally, to achieve our photometric goals, three calibration fields observed during the run would be sufficient because each field contains a large number of calibration stars. Unfortunately, the three fields observed in run I did not contain enough stars, and in the I-band, all the calibration stars saturated the detector. We will discuss how we solved this problem in section 4.3.

Next, we show the reduction steps that we performed on the entire data sets. To measure fluxes accurately, it was necessary for the background to be constant across the entire field of view to $< 1\%$. This required accurate flat-fielding. In addition, since we will measure the colors of galaxies, the four filters needed to be registered to the same pixel coordinate grid.

4.2.2 Data reduction

To put the size of the data in perspective, the raw data amount to more than $\sim$ 100Gb. The following steps filled several large hard drives ($\sim$ 300Gb) and took more than a year to complete.

Table 4.3. Summary of the observations.

Field	U band	B band	V band	I band	UT Date of Obs.
	Total Exposure Time				
APM 08279+5255	3.75hr	35min	50min	2.08hr	Feb. 7, 8, 2000
PC 1233+4752	3.50hr	40min	50min	1.92hr	Feb. 7, 8, 2000
Q2342+3417	5.11hr	47min	67min	2.78hr	Sept. 23–26, 2001
J0124+0044	3.72hr	47min	52min	2.08hr	Sept. 23–26, 2001

Table 4.4. Filter characteristics [a]

Filter	$\lambda_{\rm eff}$ (A)	$\Delta\lambda$ (A)	$\rm conv_{AB}$ (mag)
U	3651.525	387.272	0.709
B	4425.672	726.075	-0.074
V	5438.597	684.722	0.019
I	8088.854	1071.439	0.457

Table 4.5. The effective wavelength ($\lambda_{\rm eff}$), the effective bandpass ($\Delta\lambda$), and the conversion factor to AB magnitudes ($m_{\rm AB} = m + \rm conv_{AB}$).

[a]The filter transmission curve has been convolved with the CCD response function (see Fig. 4.1).

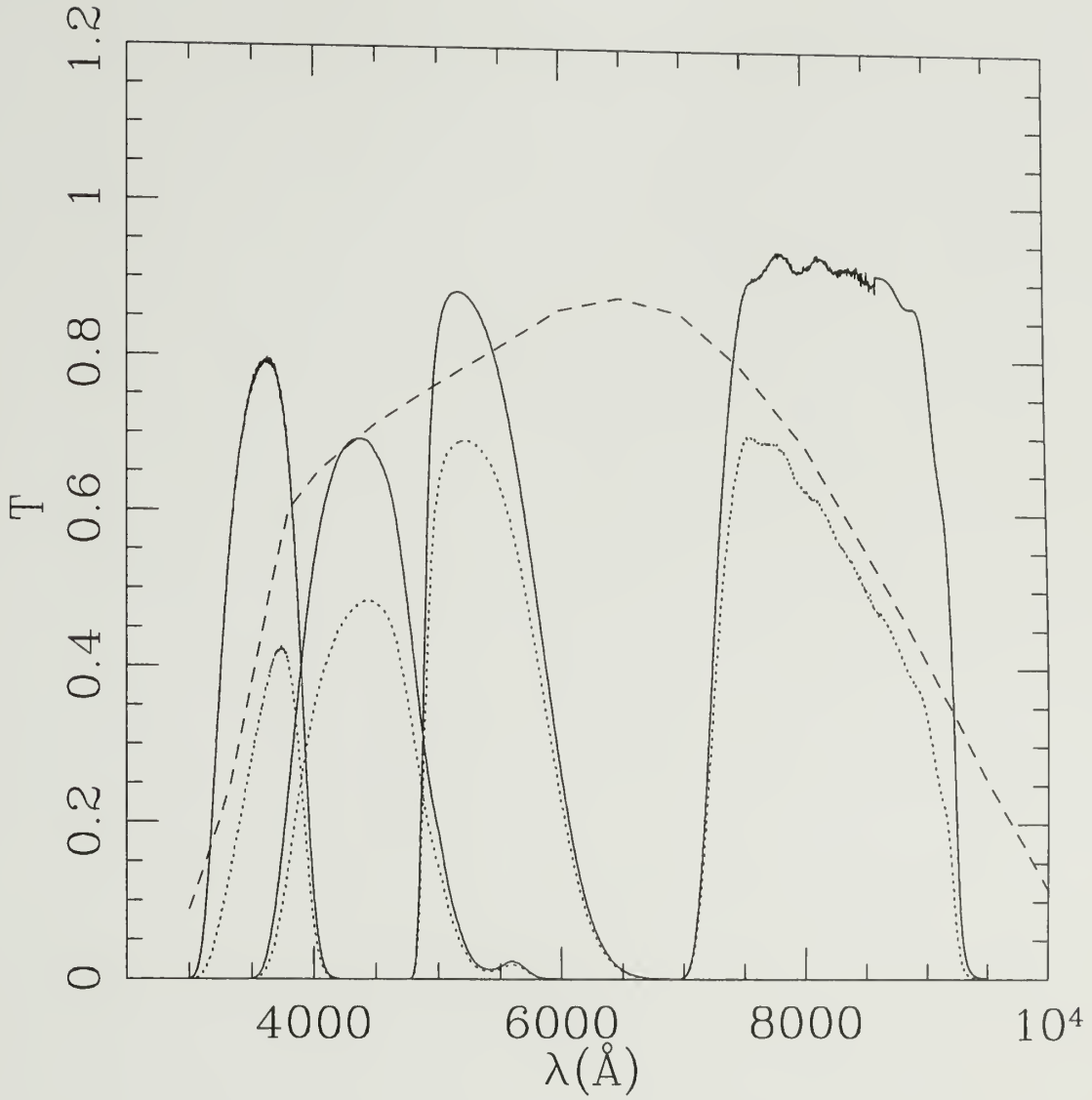

Figure 4.1 The solid lines show the transmission curves for our four filters U, B, V, and I. The dashed line shows the CCD response function. The dotted lines show the filter transmission convolved with the CCD response function.

The data were reduced using the package MSCRED (v4.1) within IRAF[1] (v2.11.3), following the data reduction guidelines of the NOAO Deep Wide-Field Survey (Jannuzi & Greer, 2000). This package was specifically designed for MOSAIC data. The reduction process contained more steps than typical optical observations. All the steps are detailed here.

We first performed the over-scan level subtraction. The over-scans are data that are added to the images at the time of read-out. This additional data contain a bias level for each line of data read. It reflects the temporal variation of the bias of each line of the CCD. Then, the images were trimmed for the over-scans. A small correction ($< 0.5\%$) was necessary due to cross talk between pairs of adjacent chips sharing the same electronics. Each CCD had a bias (DC) level that needed to be removed. For each night, we took three to ten zero second exposures, and averaged them to produce the bias frame that was applied to the rest of the data taken that night. The thinned MOSAIC chips required no dark correction since the dark current was only 5 e^- per hour.

In CCDs, each pixel has a different sensitivity to incoming photons. It is critical to correct for this in order to achieve precise photometry. To correct for pixel-to-pixel sensitivity variation, a process called flat-fielding, we took several (three per night) short exposures ($\sim$ 5s), for each filter, of a blank spot, inside the telescope dome, illuminated uniformly. The individual exposures, or dome-flats, were combined with a 3σ sigma-clip rejection to produce an 'average dome flat' that can be applied to the data.

However, we also had to overcome two non-traditional complications of the flat-fielding process: (i) due to a reflection in the optics of the camera, the MOSAIC instrument at the 4-m telescope suffered from a ghost image of the primary mirror,

[1]IRAF is distributed by the National Optical Astronomical Observatories, which are operated by AURA, Inc. under contract to the NSF.

known as a 'pupil' image, that needed to be removed, and (ii) dome flat-fields matched the night sky to no better than several percent[2], thus a sky-flat correction is needed to correct for large scale gradient residuals.

For the dome-flats, the pupil image was removed by dividing the dome-flats by an estimate of the pupil image pattern. The challenge was to determine the pupil image contribution in the presence of other flat field structures. Thus, to remove the pupil image from the averaged dome-flats, we had to first extract the pupil image from the dome-flat itself. This was done by fitting the image with an axially symmetric pattern of the pupil. The fitting was performed using the task MSCPUPIL setting the parameter 'type' set to 'ratio' as required by the MSCRED manual (Jammzi & Greer, 2000). We then flat-fielded all the science frames with the pupil-free dome-flats.

At this point, the pupil was still present in the data. It is more difficult to remove the pupil image from individual science images directly than from the dome-flats because (i) the pupil pattern is much fainter than in the dome-flats, and (ii) the pupil image is mixed with all the faint and bright objects present in the data.

A simple way to remove the pupil image from the data would be to first mask the objects and then to extract the pupil image from the average of the masked science data. However, we could not mask the objects using some threshold cutoff because the pupil was still present in the data, and the data themselves are poorly flat-fielded.

Thus, to remove the pupil image from the data and to make a sky-flat, we had to extract the pupil image from the science images themselves through the following iterative process:

---

[2]This is due to the fact that even if the telescope was being illuminated completely uniformly (by, say, the night sky) the CCD is unlikely to be illuminated uniformly, because of vignetting due to guider mirrors, non-uniformities in the filters (and dust on the filters). In other words, large-scale flat-fielding is necessary in addition to dome flat-fielding.

- (i) we created a sky-flat (v1) from the combined science images using the task SFLATCOMBINE with a median rejection and low and high thresholds of two and six respectively;

- (ii) we extracted the pupil image from the sky-flat (v1) using the task MSCPUPIL (parameter 'type' set to 'data') and removed the pupil image from the science images using the task RMPUPIL (parameter 'type' set to 'difference') in order to produce a 'first-pass' pupil-free data, following the guidelines of Jannuzi & Greer (2000);

- (iii) we created another sky-flat (v2) using the pupil-free data and applied it to the data by dividing the data by the normalized flat.

We found that low-level light from bright stars left large features in the sky-flat even if strong minmax rejection was used. To solve this problem, we created object masks by using a 2σ threshold on each of the eight individual CCDs of each image. In addition, we masked out large areas around the brightest objects. Steps (i) through (iii) were repeated using the masks and the final sky-flat (v4) was normalized and applied to all the frames. This improved procedure enabled us to have a uniform sky level ($< 1\%$) over the entire field.

The last effect that we had to take into account is fringing. Fringing, which is due to interference of the incoming light within the pixel height, occurs only at long wavelengths, i.e. in the I-band. We removed these fringes using the procedure described in (Jannuzi & Greer, 2000) before applying the final sky-flat.

Bad-pixel masks were constructed and cosmic ray removal was done using the task XZAP from the package DIMSUM and custom made routines.

4.2.3 Astrometry

The MOSAIC field of view was large enough (36arcmin by 36arcmin) that it revealed large optical distortions of several arcsecs across the field. Thus, we had to

de-project the eight CCDs onto a single image. To de-project the images, we had to derive accurate astrometric solutions. We ran the task MSCCMATCH interactively for each dithered exposure using the coordinates of several hundred calibration stars from the US-Naval Observatory (USNO) to derive the astrometric solution with an rms of $\leq 0.5''$. We then used the task MSCIMAGE[3] to match the World Coordinate Solution (WCS) to a common reference position for each filter. This mapped the eight CCDs onto a single image by rebinning the pixels to a tangent-plane projection, thus producing pixels of constant angular size.

To stack the individual flat-fielded, astrometrically calibrated images, we scaled each individual dithered image with MSCIMATCH using ~ 300 USNO stars common to each image. We then stacked the individual frames using a sigma-clipping rejection.

For the I-band of run II, large scale residuals ($\sim 1\%$) were still present in the background. To correct for this, we applied a median filtering to a block averaged image of the stacked frame and applied the normalized results to the image.

At this point, the stacked images were registered to a different astronomical coordinate in each filter band. It was crucial for the photometry to have identical pixel positions for each object in each band. The normal procedure was to use a common reference frame for all the filter bands during the registration process described above. The task MSCIMAGE allowed us to match the coordinate solution of the stacked image of all filter bands to a common reference position (here the U-band) after the fact. This ensured that each band had been de-projected to the same tangent-plane solution using the same position and orientation in the sky and that the de-projected pixel scale was the same across all images.

However, because of dithering and effects such as flexure of the telescope, differential refraction, and filter distortions, the relative positions of objects in the different

---

[3]The parameter *fluxconserve* has to be set to *no* because the science exposures have been flat-fielded with flats equally distorted to yield a constant sky per pixel.

bands were still not identical. To correct for this, we shifted and rotated slightly (~ 0.003 deg) each band image with respect to the reference band (U). To map this transformation, we measured the x & y positions of ~ 25 stellar objects that are well detected in all bands throughout the field of view using IMEXAM. We then used the tasks GEOMAP and GEOTRAN (preserving fluxes) which re-mapped each image onto the same pixel grid (translation & rotation). The rms of the transformation was typically 0.3 pixels.

To check the registration across filter bands, we compared the positional errors of an additional 1000 stellar sources with $20 < m_I < 22.5$ in the APM 08279+5255 field. Fig. 4.2a shows the relative magnitude difference Δm_U (measured using the I and the U band positions) as a function of magnitude m_U. At $m_U > 25$, the scatter in the position and therefore the magnitude difference is due to the limiting magnitude of our data. For the more reliable sources, i.e. the brightest 150 stellar objects with $m_U < 23.5$, Fig. 4.2b shows the relative magnitude difference Δm_U measured using the I and the U band positions as a function of position x. This shows that the slight residual offsets, shown in panel c) and d), do not create magnitude differences larger than 0.01 mag, a difference that is well below our photometric uncertainty quantified in section 4.6. Panels c) and d) of Fig. 4.2 show the relative astrometry difference Δx as a function of position (x or y) for the brightest 150 stellar objects with $m_U < 23.5$ and show that, throughout the field, the uncertainty in the relative astrometry is typically ~ 0.3 pixel, or less than $0.1''$.

The final stacked images are presented in Fig. 4.3 for the four fields. Each row of Fig. 4.3 corresponds to a particular field, and each column shows the the four filter band images. Each panel shows the full $35' \times 35'$ field of view. A color image of the central region $3.8' \times 3.8'$ of the APM 08279+5255 field is shown in Fig 4.4. The QSO is labelled and appears red for it is at a redshift of 3.9. A zoomed region of the field ($75'' \times 75''$) is shown in Fig. 4.5.

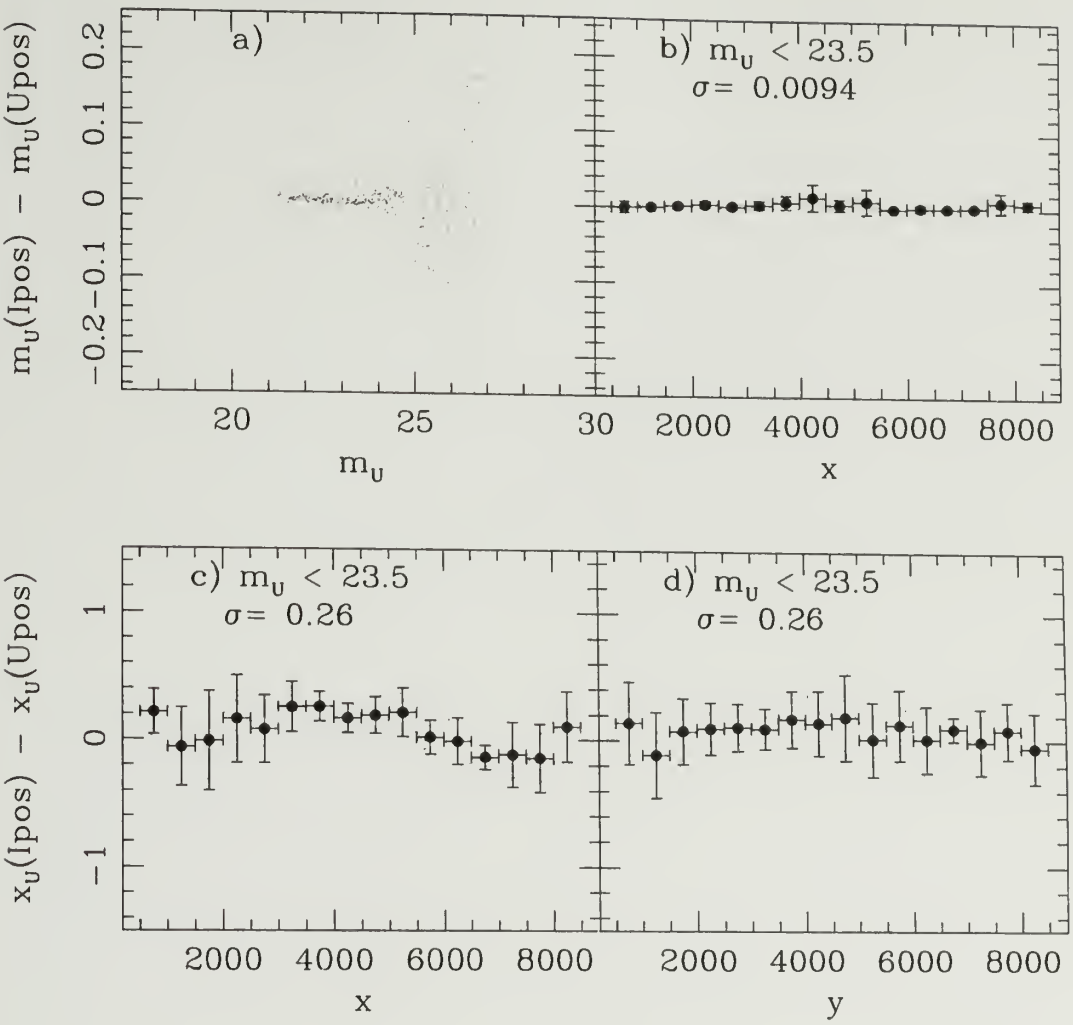

Figure 4.2 Relative astrometric accuracy. a) For 1,000 sources, the difference between the U-band magnitude measured using the x-y positions in the I-band ($m_U(I\text{pos})$) and the U-band magnitude measured using the x-y positions in the U-band ($m_U(U\text{pos})$) as a function of the mean m_U. b) The difference between the U-band magnitude measured using the x-y positions in the I-band ($m_U(I\text{pos})$) and the U-band magnitude measured using the x-y positions in the U-band ($m_U(U\text{pos})$) as a function of position for the brightest 150 sources with $m_U < 23.5$ The rms σ is shown. c) The relative difference in x positions for the brightest 150 sources with $m_U < 23.5$ as a function of mean x position. A systematic offset would produce a tilted line. No systematic variation is seen across the 8 CCDs. d) Same as c) but as a function of y positions. Again no systematic offset is seen. The rms of our relative astrometric solution between bands is 0.26 pixel ($\sim 0.1''$). All x and y positions are in pixel units.

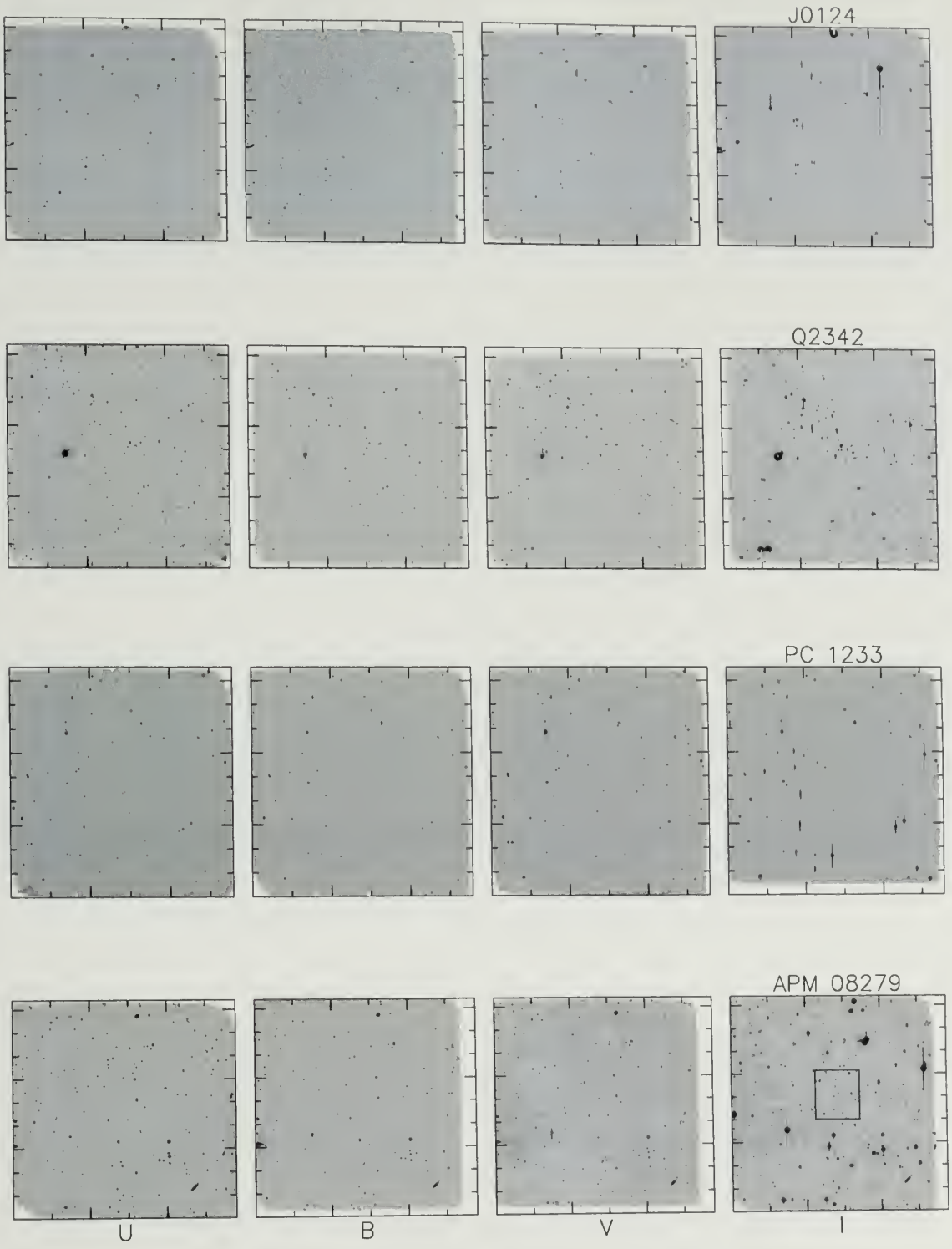

Figure 4.3 For each of the four fields, the four bands (U, B, V, and I) are shown. The rows correspond to a particular field. The columns correspond to the four filter bands. Each panel shows the entire 0.5 deg $\times$0.5 deg field of view. North is left, East is down. For the APM 08279+5255 field, the box shows the central region shown in Fig. 4.4.

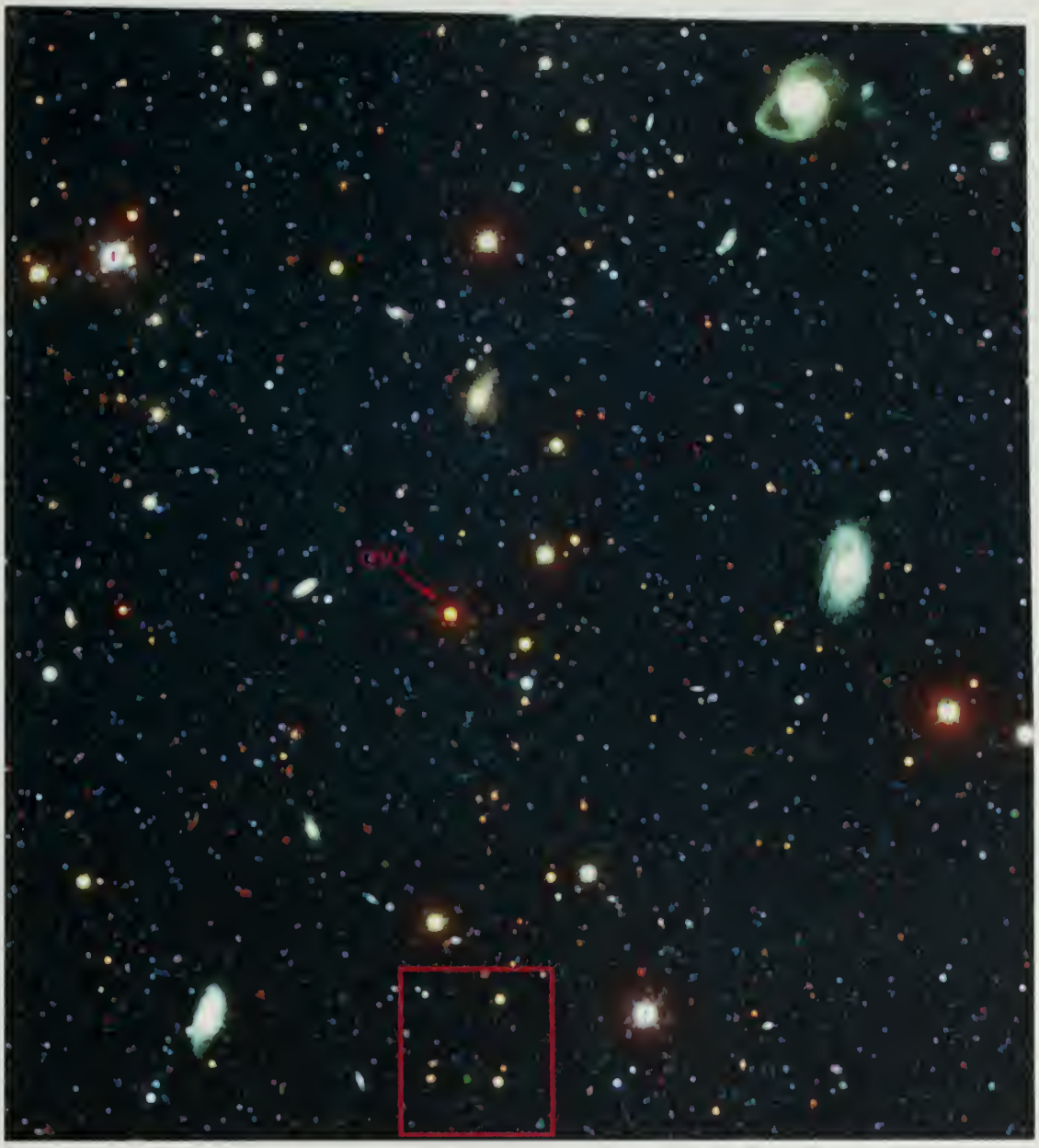

Figure 4.4 BVI-color image of the $8.1' \times 8.1'$ central region of the field APM 08279+5255. North is left; East is down. The quasar is shown and appears red because of its redshift of $z = 3.91$. The red box shows the zoomed region shown in Fig. 4.5

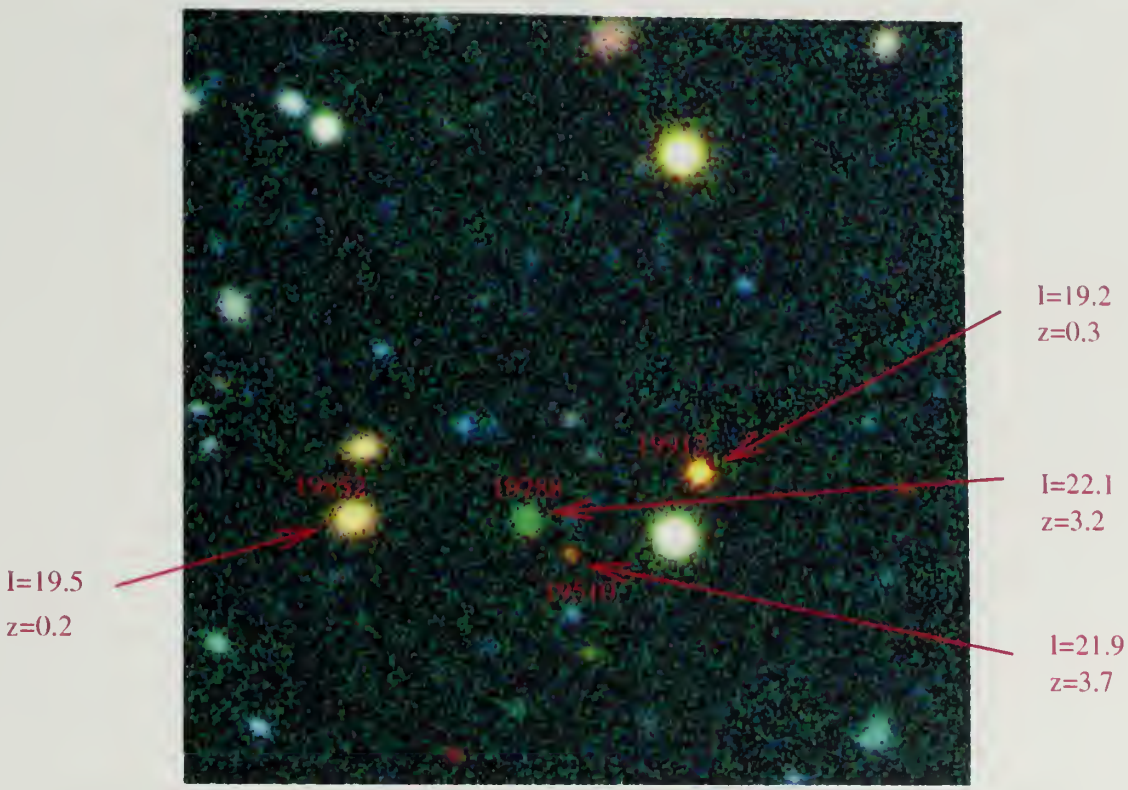

Figure 4.5 UBI color image of the inset shown in Fig. 4.4, 1.3′on a side. A few objects are labelled by their ID number. Their I magnitude and their photometric redshift are shown. Note that most of our LBGs are much fainter than these objects, since they have $I \simeq 24$. Galaxies so faint are not visible on such a picture.

4.3 Photometric calibration

To calibrate our data, we used the standard stars imaged during both observing runs. The star fluxes were obtained from Landolt (1992). The frames of these standard stars were reduced as described in the previous sections including the deprojection to a common tangent-plane projection.

The field of view of MOSAIC was so large that it was possible to observe up to forty standard stars at once. Unfortunately, not all the Landolt fields are rich in stars: for run I only, the number of stars per calibration field turned out to be too low ($0-5$ per filter band) to perform accurate photometry. In addition, all the standard stars were saturated in the I-band.

On the other hand, for run II, we observed ~ 150 standard stars (Landolt, 1992) through each filter, which enabled us to perform accurate photometry on both run I and run II. To tie the photometric zero-point (ZP) of run I to that of run II, we observed a common field (APM 08279+5255) during both runs.

The photometric solution (ZP, α, β) is the fit to the following equations for each filter band:

$$\begin{aligned} m_U^{L92} &= m_U^{obs} + ZP_U + \alpha_U \cdot X_U + \beta_U (U-V)_{L92} \\ m_B^{L92} &= m_B^{obs} + ZP_B + \alpha_B \cdot X_B + \beta_B (B-V)_{L92} \\ m_V^{L92} &= m_V^{obs} + ZP_V + \alpha_V \cdot X_V + \beta_V (B-V)_{L92} \\ m_I^{L92} &= m_I^{obs} + ZP_I + \alpha_I \cdot X_I + \beta_U (V-I)_{L92} \end{aligned} \tag{4.1}$$

where $m^{obs} = -2.5 \log \frac{C}{t_{exp}}$ is the observed magnitude from the total number of counts C during an effective exposure time t_{exp}, ZP is the photometric zero-point which quantifies the total sensitivity, $\alpha \cdot X$ measures the atmospheric extinction as a function of airmass X, and β in the color term accounts for variations in the effective wavelength sampled by the filter for stars with very different colors. We neglected

the differential atmospheric extinction as a function of color, $\gamma X(B-V)$. The observed fluxes m^{obs} of the standard stars were measured by fitting the point spread function (PSF) and computing the integrated flux. Landolt (1992) used an aperture $7''$ in radius. However, we found that the integrated flux was similar to the aperture used by Landolt (1992) at the 0.015mag level, which was well within our error bars. If this induced a slight systematic offset in our total magnitudes, it would not affect the colors and hence our results.

For run II, the parameters in Eqs. 4.1 were well constrained, since we observed a large number of calibration stars spanning a large range of colors and air-masses. The Eqs. 4.1 are linear in their parameters, ZP, α, β, and thus can be solved at once by a single matrix inversion. We performed the matrix inversion using a Single Value Decomposition (SVD) routine. Figs. 4.6-4.9 shows the results. Fig. 4.6 shows the calibrated magnitude with the fitted parameters to Eq. 4.1 as a function of published magnitude m^{L92} for the U-band. Figs. 4.7, 4.8 and 4.9 show the same for the B, V and I-band respectively. Table 4.6 summarizes the results.

For run I, we used the short exposure (30sec) of the APM 08279+5255 field taken during run II to tie the zero-point of run I to that of run II. The difficulty here was to find enough stars detectable in the short exposure that did not saturate in the deep exposure (900sec) of run II, with the additional constraint that this condition must be met in all four filters in order to use the color term in Eqs. 4.1. Typically a dozen stars met both criteria and the zeropoints are shown in Figs. 4.10– 4.13 using α and β from run II.

For the U, B and V filters of run I only, we were able to find an independent photometric solution to Eqs. 4.1 from the few original calibration stars of run I. However, the photometric solution was very different from the telescope's typical values and was poorly constrained. As a consistency check, we plotted the stars that were used to tie the two runs using solely the calibration from run I as red filled circles

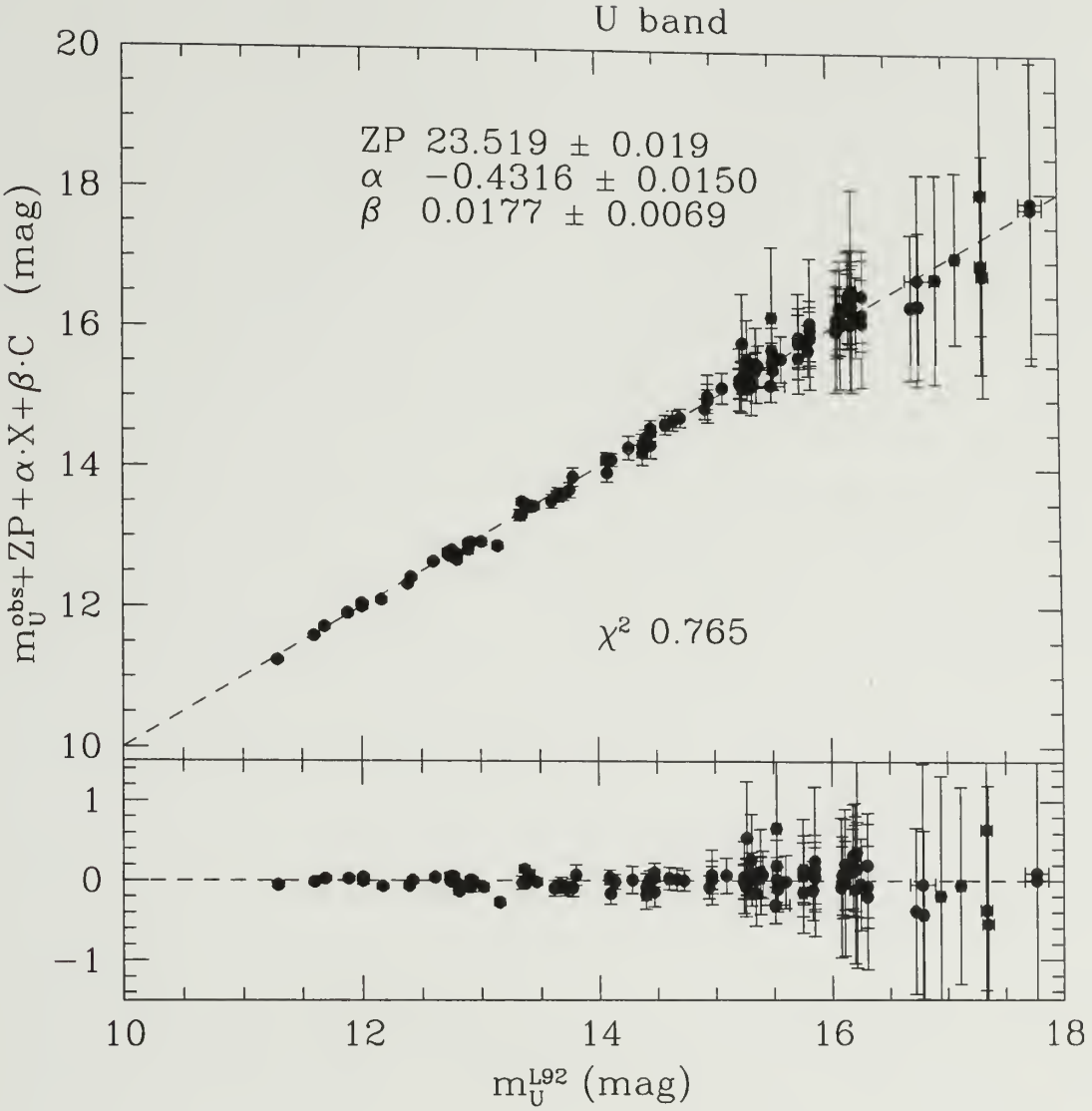

Figure 4.6 U-band photometric calibration of run II. The calibrated magnitude is shown as a function of published magnitude m^{L92} for each standard star (Landolt 1992) used to find the photometric solution (ZP, α, β). The bottom panel shows the residuals.

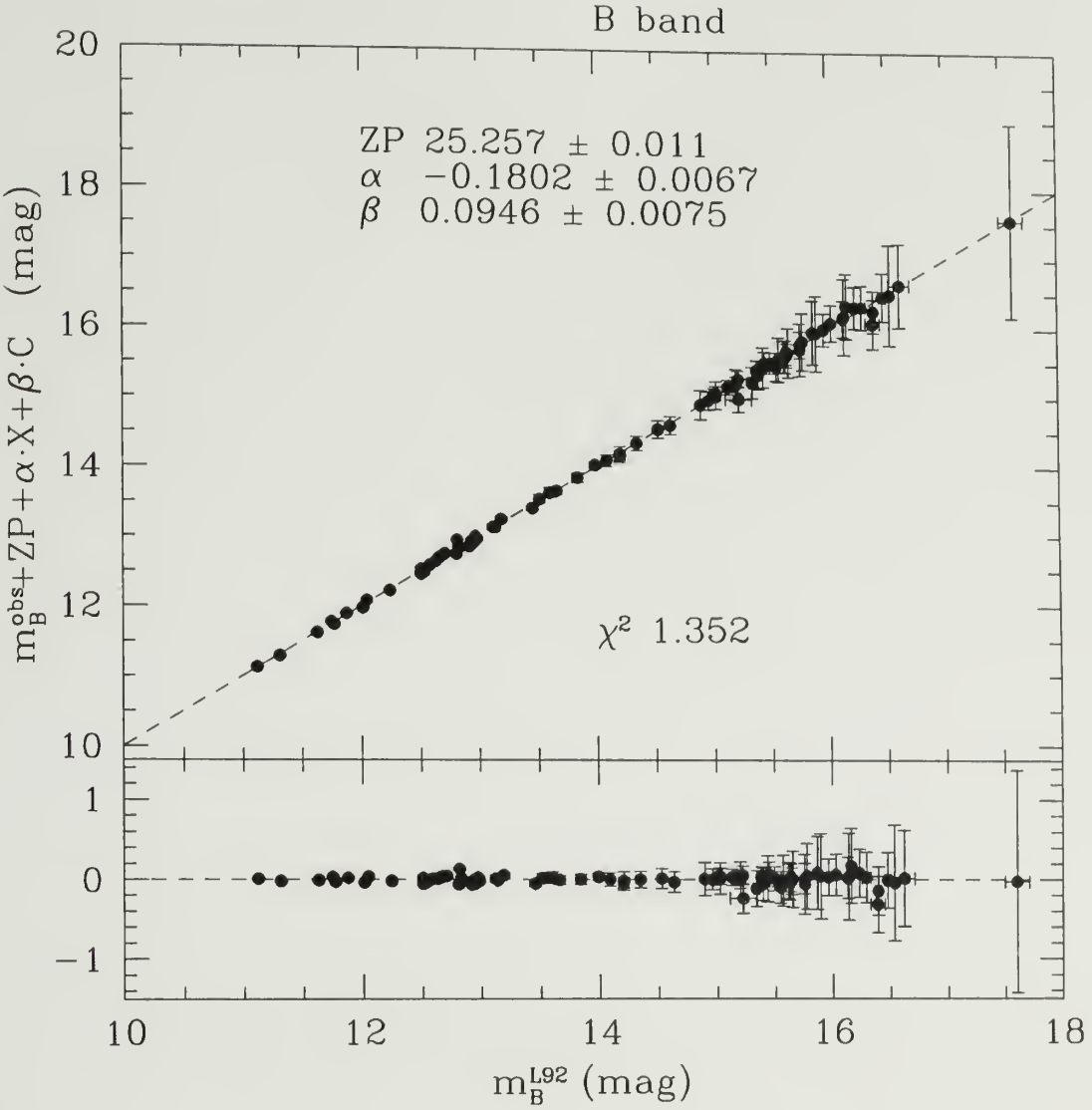

Figure 4.7 B-band photometric calibration of run II. The calibrated magnitude is shown as a function of published magnitude m^{L92} for each standard star (Landolt 1992) used to find the photometric solution (ZP, α, β). The bottom panel shows the residuals. The value of the reduced χ^2 is shown.

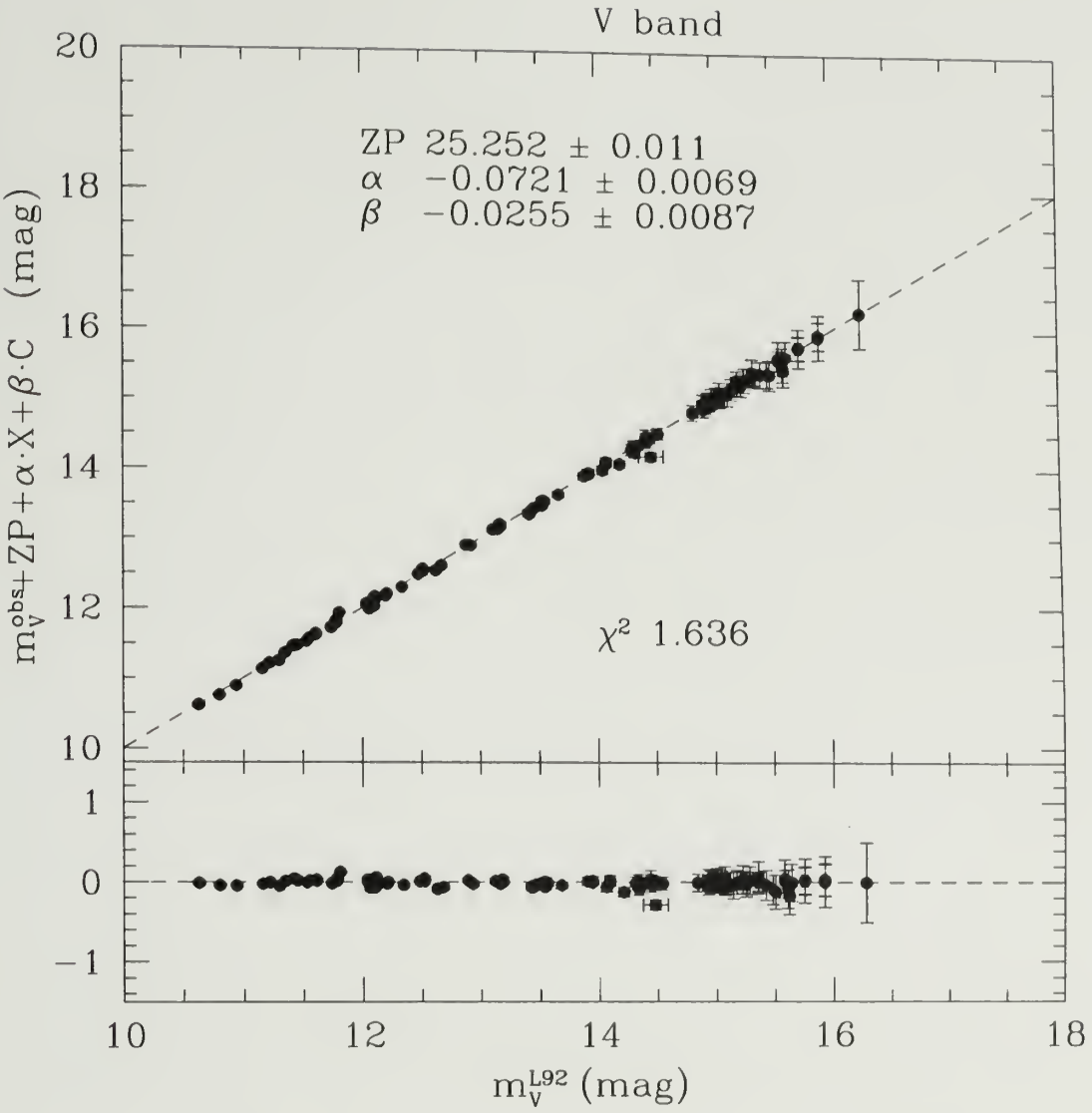

Figure 4.8 V-band photometric calibration of rm II. The calibrated magnitude is shown as a function of published magnitude m^{L92} for each standard star (Landolt 1992) used to find the photometric solution (ZP, α, β). The bottom panel shows the residuals. The value of the reduced χ^2 is shown.

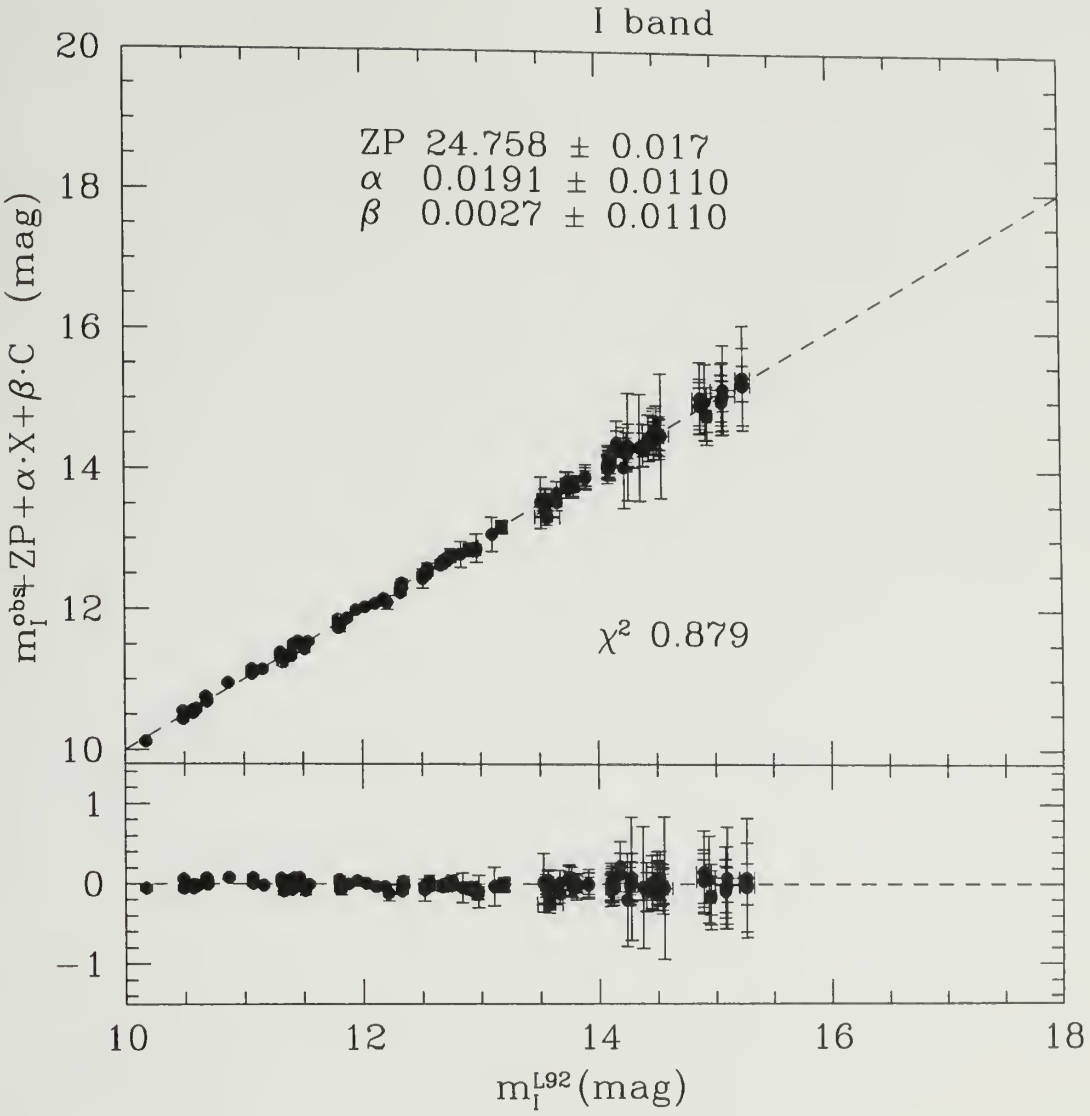

Figure 4.9 I-band photometric calibration of run II. The calibrated magnitude is shown as a function of published magnitude m^{L92} for each standard star (Landolt 1992) used to find the photometric solution (ZP, α, β). The bottom panel shows the residuals. The value of the reduced χ^2 is shown.

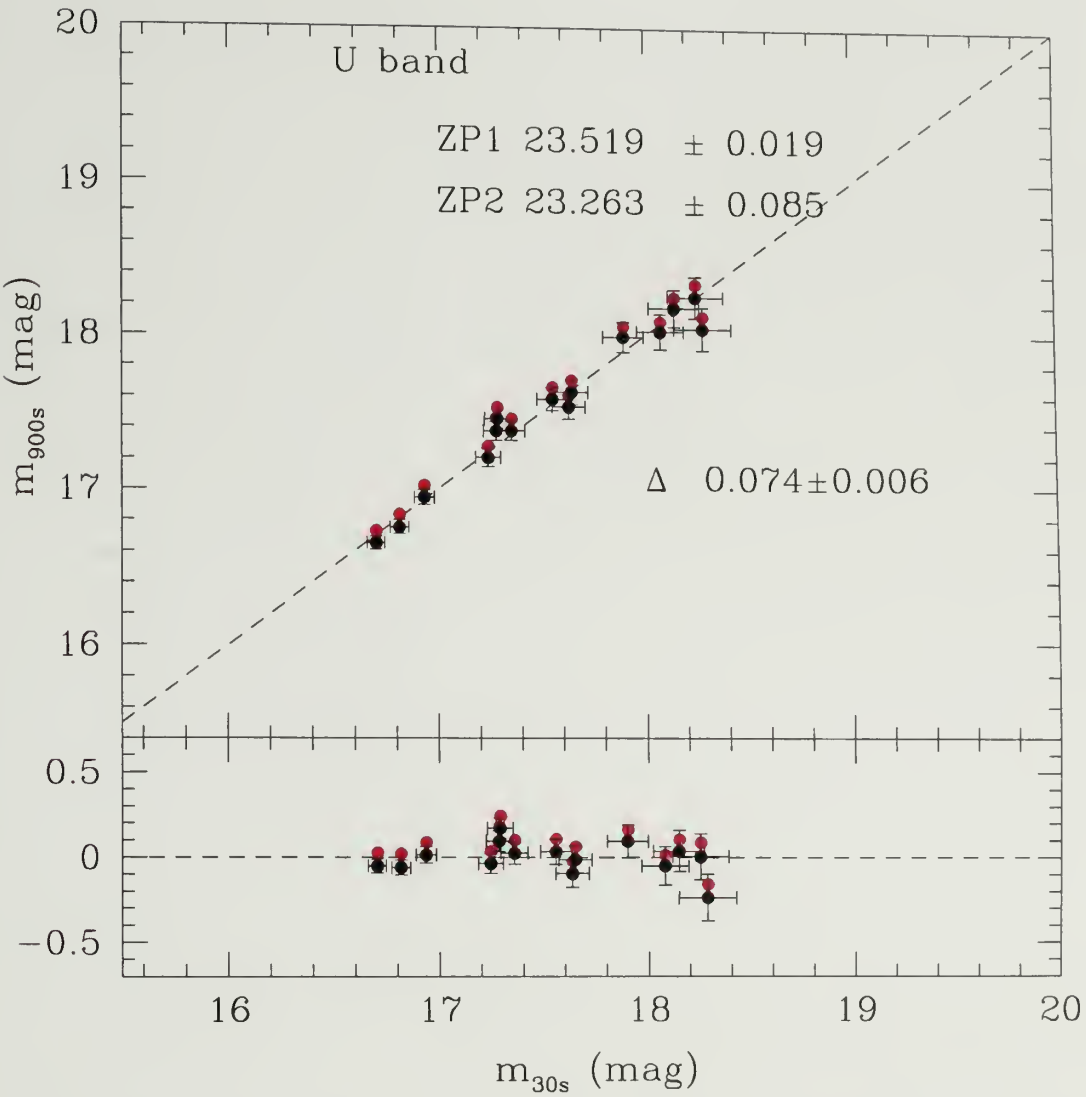

Figure 4.10 U-band calibration of run I (m_{900s}) against run II (m_{30s}). The red points (with no error bars) show the magnitude obtained using the poor self-calibration of run I. The mean offset Δ between the two methods is shown.

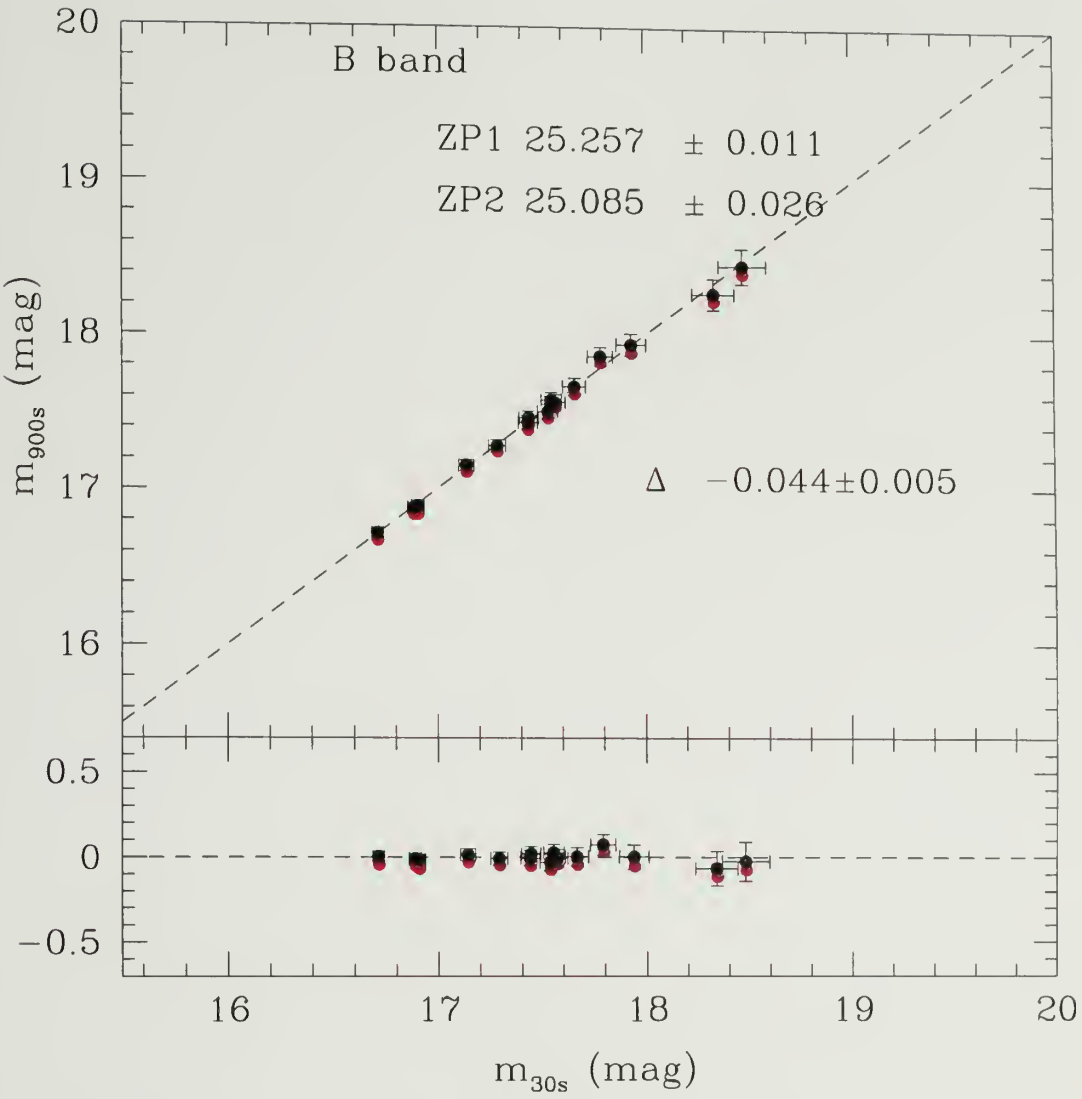

Figure 4.11 B-band calibration of run I against run II. The red points (with no error bars) show the magnitude obtained using the poor self-calibration of run I. The mean offset Δ between the two methods is shown.

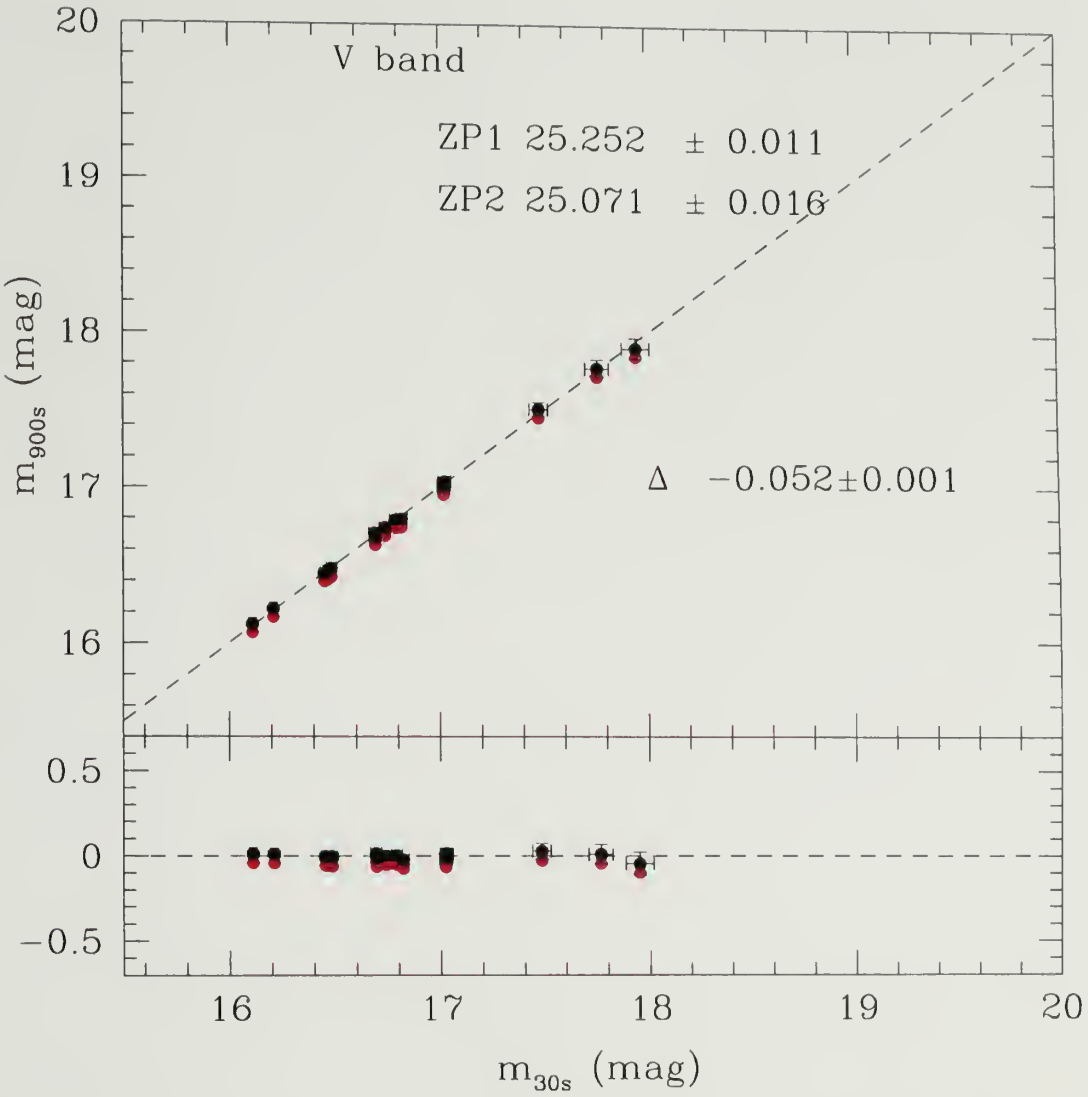

Figure 4.12 V-band calibration of run I against run II. The red points (with no error bars) show the magnitude obtained using the poor self-calibration of run I. The mean offset Δ between the two methods is shown.

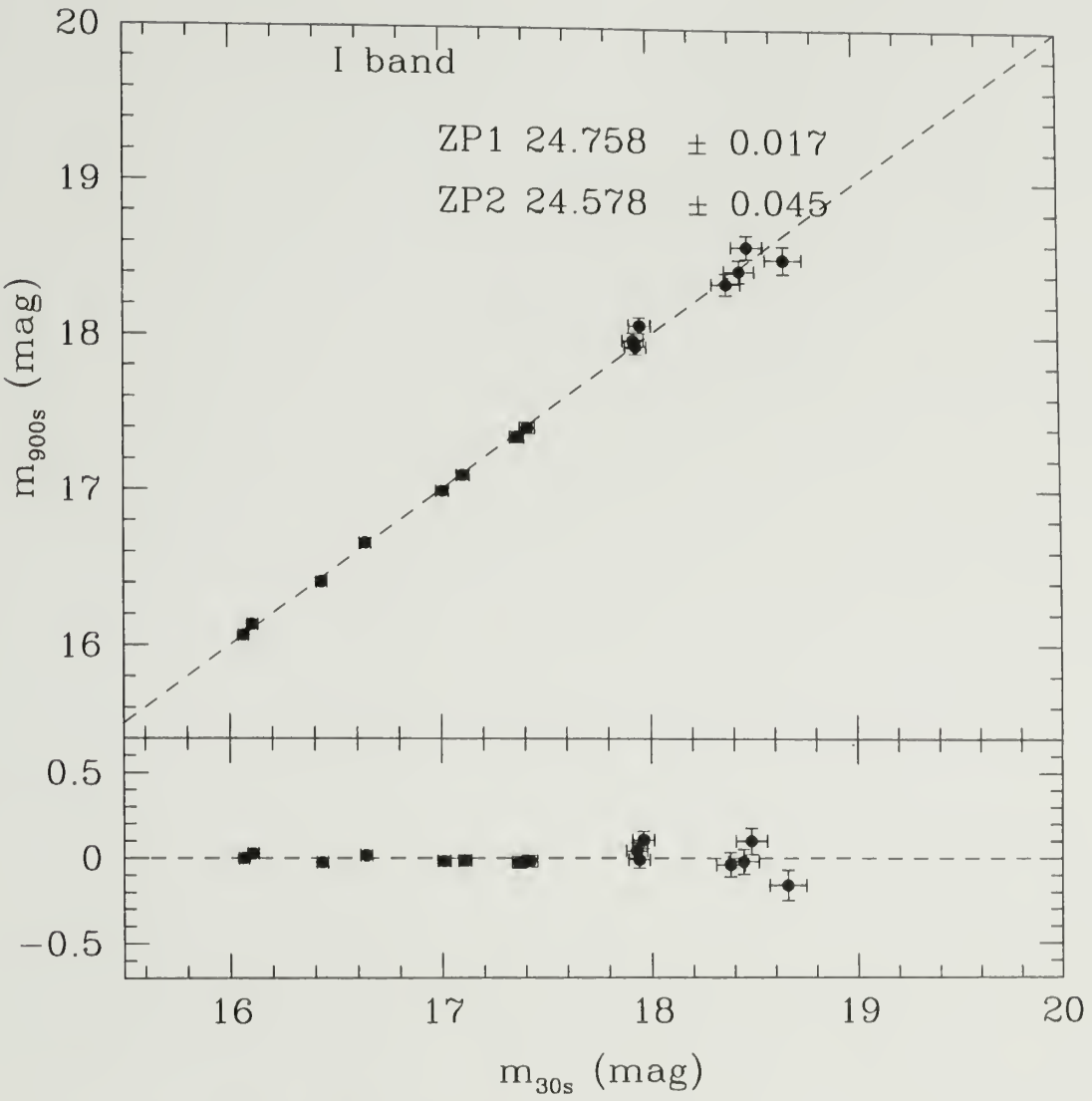

Figure 4.13 I-band calibration of run I against run II.

Table 4.6. Photometric solution of Eqs.4.1.

Filter	Run I [a] ZP	Run II ZP	α	β
U	23.26 (0.08)	23.52 (0.02)	-0.421 (0.02)	0.018 (0.007)
B	25.085 (0.03)	25.26 (0.02)	-0.180 (0.01)	0.095 (0.008)
V	25.07 (0.02)	25.25 (0.02)	-0.072 (0.01)	-0.025 (0.009)
I	24.58 (0.05)	24.76 (0.02)	0.020 (0.01)	0.003 (0.011)

[a]For run I, we assumed the airmass coefficient α and the color term β to be the same as for run II.

in Figs. 4.10 – 4.12. There was a small systematic and unexplained difference between the two runs indicated by the mean offsets Δ. However, because of the paucity of standard stars in run I, and for consistency, we used α and β from run II in the remaining analysis, summarized in Table 4.6.

The sky background level sets our flux limit. We, usually, achieved a flux limit of 25.5mag (3σ) in the I-band. This flux limit was measured in an aperture of diameter equal to twice the full-width at half maximum (FWHM), or 2×FWHM. The sky surface brightnesses in the other bands are summarized in Table 4.7.

4.4 Completeness

It was important to quantify how complete our sample of detected galaxies was at a given magnitude. The completeness is defined by the fraction of objects detectable at a given magnitude.

To estimate our completeness and to optimize our detection algorithm, we added fake stellar objects with known total fluxes in the I-band image of our four fields. Each faked object had a PSF that matched the image PSF.

We optimized the algorithm (described in the next section) used to extract objects in our data to maximize our completeness, i.e. to reach the faintest flux levels (3σ)

Table 4.7. Depth of the observations

Fields	Filter	Exp. /Frames (sec./#)	Airmass X[a] (min-max)	FWHM (pix/")	$SB_{lim}(1\sigma)$ [b] (mag/m_{AB})	$SB_{lim}(5\sigma)$ [b] (mag/m_{AB})	$m_{lim}(3\sigma)$ [c] (mag/m_{AB})	Completeness 50% [c] (mag/m_{AB})
APM 08279+5255	U	13500/15	1.07-1.17	4.32/1.1	27.78/28.49	26.03/26.74	25.85/26.56	
	B	2100/7	1.14-1.27	4.37/1.1	28.44/28.36	26.69/26.61	26.49/26.42	
	V	3000/10	1.21-1.37	4.76/1.2	28.21/28.23	26.46/26.48	26.17/26.19	
	I	7590/20	1.11-1.62	4.35/1.1	27.15/27.61	25.40/25.86	25.21/25.66	24.40/24.87
PC 1233+4752	U	12600/14	1.04-1.16	4.10/1.05	27.82/28.52	26.07/26.78	25.94/26.65	
	B	2400/8	1.05-1.07	3.94/1.0	28.59/28.51	26.84/26.76	26.75/26.68	
	V	7590/10	1.07-1.11	3.59/0.9	28.18/28.20	26.43/26.45	26.45/26.47	
	I	6900/15	1.07-1.45	4.14/1.1	27.19/27.64	25.44/25.90	25.30/25.76	24.48/24.95
Q2342+3417	U	18400/21	1.00-1.16	5.05/1.3	27.91/28.62	26.17/26.88	25.81/26.52	
	B	2350/7	1.23-1.47	5.09/1.3	27.93/27.86	26.18/26.11	25.82/25.74	
	V	3350/10	1.01-1.11	5.91/1.5	28.38/28.40	26.63/26.65	26.11/26.13	
	I	10000/25	1.12-1.72	4.85/1.25	28.08/28.54	26.34/26.79	26.03/26.48	24.32/24.79
J0124+0044	U	13400/16	1.19-1.51	5.93/1.5	27.88/28.59	26.13/26.84	25.60/26.31	
	B	2800/7	1.24-1.42	5.76/1.5	28.69/28.61	26.94/26.87	26.44/26.37	
	V	3100/7	1.46-1.90	5.47/1.4	28.33/28.35	26.58/26.60	26.14/26.16	
	I	7500/20	1.21-1.80	4.28/1.1	27.67/28.13	25.93/26.38	25.75/26.21	24.48/24.95

[a]The airmass is $\frac{1}{\cos\zeta}$ where ζ is the zenith angle of the telescope.

[b]surface brightness per square arcsecond.

[c]measured inside a 2× FWHM diameter aperture.

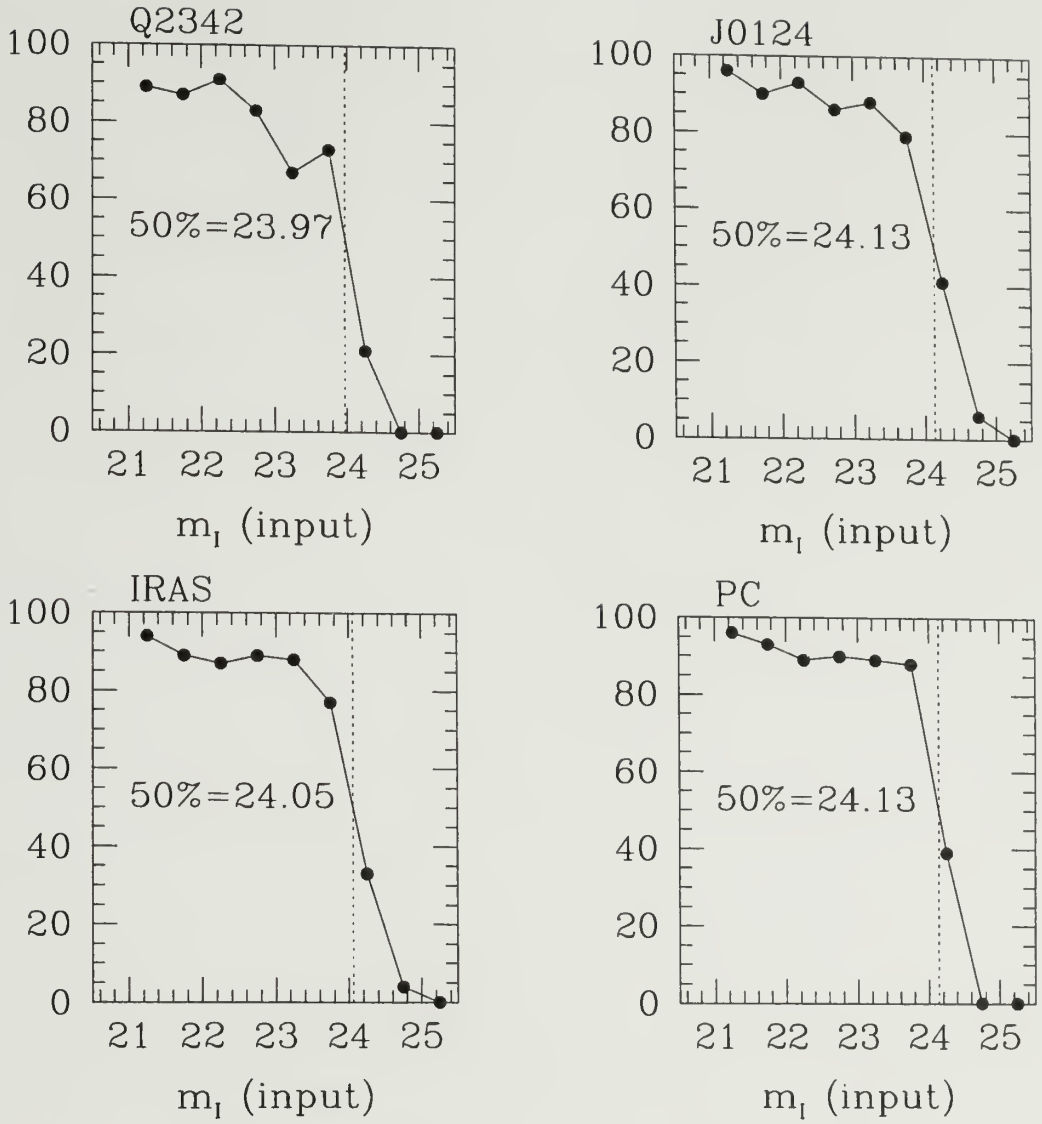

Figure 4.14 For each of the four fields, the fraction of faked sources recovered (in percent), i.e. our completeness is shown as a function of input magnitude. Our 50% completeness level is ~24 (total magnitudes) which corresponds to $m \simeq 24.35$ in a 2×FWHM aperture magnitude.

without being contaminated with spurious detections. Usually, spurious detections clustered on streaks from saturated stars or ghost images of such stars from the previous readouts. Since our ultimate goal was to look at fluctuations in the number of galaxies across the images (subject to small number statistics), the reliability of the catalog was kept high. Upon visual inspection, we found that 1-2 percent of the sources in our final catalog were not real.

Fig. 4.14 shows our completeness as a function of input magnitude using the now optimized detection algorithm. Fig. 4.14 also shows that we recovered $> 90\%$ of the faked objects at magnitudes brighter than ~ 23. For each of the four fields, we reached a 50% completeness level of $m_I(\mathrm{input}) \simeq 24.0$ shown by the dotted lines. Since fluxes were measured in a fixed aperture of 2×FWHM (see section 4.6), our 50% completeness level corresponds to $m_I \simeq 24.35$ or $I_{\mathrm{AB}} \simeq 24.85$, which is approximately $R_{\mathrm{AB}} \sim 25$ (Steidel & Hamilton, 1993). This is about 0.5 magnitude brighter than the survey of Steidel (2000).

Is this limit sufficient to detect galaxies at $z \sim 3$?

Since our redshift of interest was $z \sim 3$, it was important to answer this question. The standard check was given by the comparison between our magnitude limit against the typical luminosity of a galaxy at $z \sim 3$, referred to as L^*. L^* is found from the knee of the luminosity function. From the luminosity function of Steidel et al. (1999), a typical galaxy at $z \sim 3$ has a magnitude $m^*_{\mathcal{R}} = 24.5$. Therefore, our completeness limit ($I_{\mathrm{AB}} = 24.8$ or $\mathcal{R} \simeq 25$) corresponds to L^* at redshift $z \sim 3$, and we have reached a depth sufficient to ensure that we sample well the bright end of the luminosity function at redshift $z = 3$.

In terms of luminosity, this magnitude limit, $I_{\mathrm{AB}} = 24.8$, corresponds to $M_{g,\mathrm{AB}} = -21.1 - 5\log h$. In performing this conversion, we used Eq. C.15 in Appendix C with a distance modulus $DM(z = 3) = 46.25$, and converted our I-band limiting magnitude to the g filter, using the correction $K_{gI}(z) = m_I^{obs} - m_g^{rest}$, known as

the K-correction (Eq. C.6). In this case, the K-correction $K_{gI}(z = 3)$ was -0.30 assuming a spectral type of an Irr galaxy (Coleman, Wu, & Weedman, 1980) which best represents star-forming galaxies at redshift $z \sim 3$.

As a side note, we compared our completeness limit to the typical luminosity of galaxies today at $z = 0$, keeping in mind that galaxies at $z \sim 3$ are much more luminous. Recent results of the Sloan Digital Sky Survey (SDSS) luminosity function showed that L^* is (in absolute magnitude): $M^*_{g,\mathrm{AB}} - 5\log h$=-20.04 (Blanton et al., 2001). Thus, our completeness limit, $M_{g,\mathrm{AB}} = -21.1 - 5\log h$, corresponded to $2.5 \times L^*(z = 0)$.

4.5 Source extraction

At this point, we knew the limiting magnitude of our images and we could create our source catalogs by applying the detection algorithm that had just been optimized to the real data.

All sources were detected in the I-band using the software of Bertin & Arnouts (1996) SExtractor (v2.1.6). We used a background mesh size of 64 pixels. The local background estimation was made within an annulus 24-pixel wide. The images were convolved with a 2 pixel FWHM gaussian kernel before source detection. The detection threshold was set to 1.5σ with a minimum area of 5 pixels. This detection threshold corresponds to a 3σ source detection against the background noise.

Bad-pixel masks were used as flag images. SExtractor is able to perform deblending of close objects. The number of deblending sub-thresholds (DEBLEND_THRESH) was set to 32 pixels, and through experimentation, the minimum contrast parameter (DEBLEND_MINCONT) was set to 0.0001. The other parameters are listed in Appendix D. Our catalog contains a total of approximately 40,000 objects per field, 30,000 of which have $I > 22.5$.

4.6 Photometry

Now that we have described the data, the calibration, our completeness and our detection algorithm, the last reduction step was to apply the photometric calibration (§ 4.3) to the data in order to produce our catalog of galaxies.

To be able to apply the photometric solution, we had to figure out a reliable way to measure fluxes. This is a frequently encountered problem in optical astronomy. There are a number of ways to measure magnitudes in the literature. Is it best to use fixed aperture, Kron or Petrosian magnitudes? It often comes down to choosing between a magnitude that is the least biased, or that has the least uncertainty. In our case, because we were after small, star-like, and faint objects, we needed a magnitude that has a small uncertainty.

From our sample of fake stars (see section 4.4), we tested the following magnitude measurements: the 'BEST_MAG' magnitude from SExtractor[4] and magnitudes measured with three different fixed apertures. 1.67×, 2×, and 2.5×FWHM. For each of these magnitudes, Fig. 4.15 (top left panel and bottom panels) shows the difference between the observed and input magnitude (filled dot) as a function of input magnitude for the 800 faked stars. The average magnitude m(obs) for each input magnitude m(input) bin is shown as the large red filled circles with error bars that show the rms dispersion. The average magnitude error for each input magnitude m(input) bin is shown attached to the filled square, offset for clarity. From Fig. 4.15, it is clear that the 'BEST_MAG' magnitude is unbiased. It recovers the input flux, although its uncertainty is the largest among these four magnitudes. On the other hand, aperture magnitudes, which are biased, recover more of the input flux as the aperture increases—as expected—but, their uncertainty increases with larger apertures as more of the sky is included in the aperture measurement.

---

[4]it is a choice between a Kron-like elliptical aperture and an isophotal magnitudes; see http://www-int.stsci.edu/ holwerda/se.html for details.

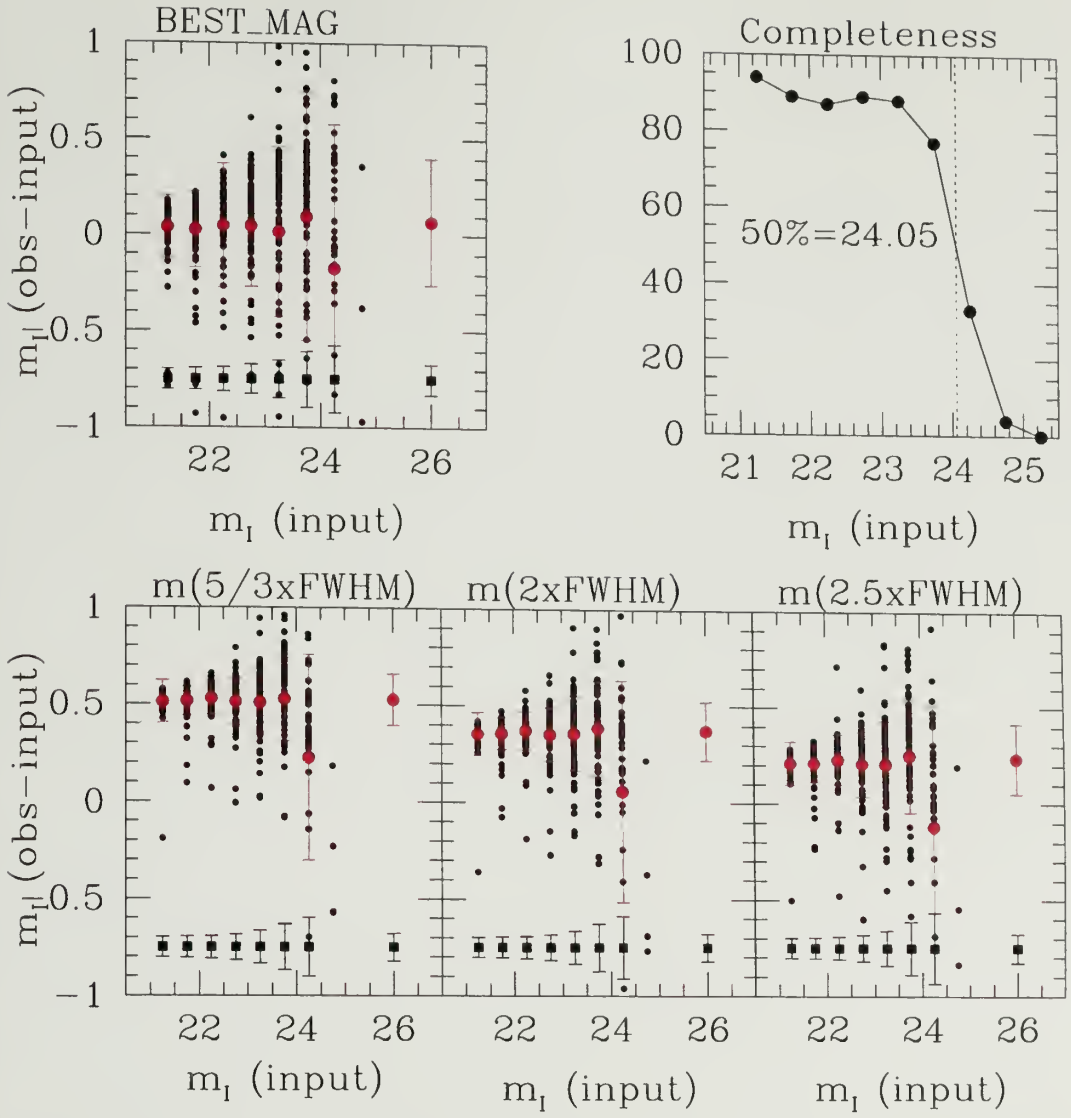

Figure 4.15 Bottom: from Monte Carlo simulations in the APM 08279+5255 field, the difference between the aperture magnitude m(obs) and the total flux m(input) is shown for three different aperture sizes (1.67×, 2× and 2.5×FHWM) as a function of input magnitude for 800 faked stars (black circles). The red circles with 1σ error bars show the weighted means, and the black squares show the mean errors with the rms as error bars. The data at $m = 26$ show the weighted mean for the 800 stars. As expected, as the aperture increases, the observed flux is closer to the total flux. Top left: Same as bottom panels, for the 'BEST_MAG' magnitude from SExtractor which shows that this is a good estimate of the total flux, however it is noisier as the errors are larger. Top right: fraction of detected stars (completeness) as a function of input magnitude. The 50% completeness level is shown.

For faint galaxy surveys, it is preferable to use small apertures which miss a known fraction of the flux rather than to try to recover the total flux, because all galaxies have approximately the same size at these faint fluxes. In addition, aperture magnitudes have the advantage of smaller uncertainty. We elected to use 2×FWHM apertures, which are biased by about 0.35 mag. Indeed, from Fig. 4.15 (middle panel), the total magnitude is found to be $m_{\rm tot} = m_I(2 \times FWHM) - 0.35$.

To take the variation of the seeing in the different filter bands into account, the color of object i between two bands, $b1$ and $b2$, is $(m_{b2} - m_{b1})_i = m_i(2\times\mathrm{FWHM}_{b2}) - m_i(2\times\mathrm{FWHM}_{b1})$ [5]. Normally, obtaining magnitudes in two different apertures would be performed by running the algorithm software SExtractor with different apertures in 'dual mode' where the detection is performed on only one image. Unfortunately, a bug in the software in 'dual mode' produced catalogs where some sources were missing or out of sequence. This problem occurred only when the two apertures were significantly different, and its origin is not known. To solve this problem, we performed the photometry in all four bands using the IRAF package PHOT at the positions of each source of the SExtractor catalog.

To apply the photometric solution (Eq. 4.1) to the observed counts, we iterated the color correction three times, and the solution converged very rapidly. For, sources that were not detected in one band (i.e. flux$< 1\sigma$), the magnitude is set to the 1σ flux limit measured in a 2×FWHM diameter aperture and no color term was computed in Eq. 4.1.

Finally, for each object at position x, y on the sky, we subtracted the Galactic extinction by adopting $E(B-V)$ values taken from Schlegel et al. (1998) assuming a $R_V = 3.1$ extinction curve, i.e.

---

[5]This procedure is valid for star-like objects, and it has been shown to be a good approximation for faint galaxies (Smail et al., 1995).

$$m^o(x,y) \;=\; m_X^{obs} - A_X(x,y) = m_X^{obs} - \frac{A_X}{E_{B-V}}\, E_{B-V}(x,y) \qquad (4.2)$$

where, $\frac{A_X}{E_{B-V}}$ is 4.968, 4.325, 3.240, 1.962 for U, B, V, and I, respectively.

4.7 Towards selecting high redshift galaxies

At this point, we have calibrated the flux measurements of all of our 40,000 sources in each field. To trim this large sample to a cleaner and more manageable sample for the photometric redshift technique presented in the next chapter, we applied the following selection.

We rejected spurious detections, i.e. objects with a size less than 0.85% of the FWHM for stellar objects. We also removed objects that were near the edge of the field, that is, only objects meeting the condition

$$500 < x < 8450, \quad 600 < y < 8300 \qquad (4.3)$$

were kept, where x and y are the pixel coordinate.

As mentioned in section 2.1, star-forming galaxies at $z \sim 3$ are characterized by their break at the Lyman limit. Since at $z \sim 3$, this break occurs in the optical, between the U and the B filters, Lyman break galaxies will have very red $U - B$ colors, $U - B > 1.5$. On the other hand, their $B - I$ color will be blue ($B - I \sim 0$) since their intrinsic spectrum is blue due to their strong star formation. Thus, we can select LBGs at $z \sim 3$ using simple color cuts.

Since, we will estimate the redshift of a particular galaxy using photometric redshift techniques, we selected a large subset (one third) of our catalog using colors cuts that included not only all the $z \sim 3$ galaxies, but also a significant fraction of

$z \sim 0 - 3$ galaxies. This ensured that the color cuts did not introduce a sharp drop in our $z \sim 3$ selection function. The color cuts were:

$$(U - B) > -0.5 \quad . \quad (B - I) < 4.2,$$
$$(U - B) > -0.5 + 0.55 \cdot [(B - I) - 2.5] \qquad (4.4)$$

or in AB magnitude,

$$(U - B)_{\rm AB} > 0.3 \quad . \quad (B - I)_{\rm AB} < 3.7,$$
$$(U - B)_{\rm AB} > 0.3 + 0.55 \cdot [(B - I)_{\rm AB} - 2]. \qquad (4.5)$$

with a minimum brightness of $m_I = 24.35$, corresponding to our 50% completeness level. Fig. 4.16 shows the $x - y$ positions of approximately 10,000 objects per field that meet this color criteria. Fig. 4.16 also shows the mask regions that we placed around bright stars. This ensured that no spurious detection due to streaks or other defects were included in our catalogs. Fig. 4.16 shows a uniform sensitivity across the entire field of view.

Fig. 4.17 shows the color-color $(U - B), (B - I)$ diagram for the APM 08279 field. The color cuts used are shown with the dot-dashed line. Open squares are our LBG candidates in the redshift slice centered on the DLA ($z_{\rm DLA} = 2.974$) selected using photometric redshifts. Objects with U and B upper limits (filled triangles) have their $U - B$ color set to an arbitrary constant of five. A sample of stellar objects with $18.5 < I_{\rm AB} < 20.5$ was selected from their FWHM being less than 1.25× the FWHM of the PSF with the additional constrain that their ellipticity was less than 1.25. Stars show the locus of these stellar objects.

This color cut (Eq. 4.5) removed most of the low-redshift objects at $z \leq 1$. We, then, used photometric redshift techniques to select galaxies at a redshift near that of the DLA $z \sim z_{\rm DLA}$ which is described in the next chapter.

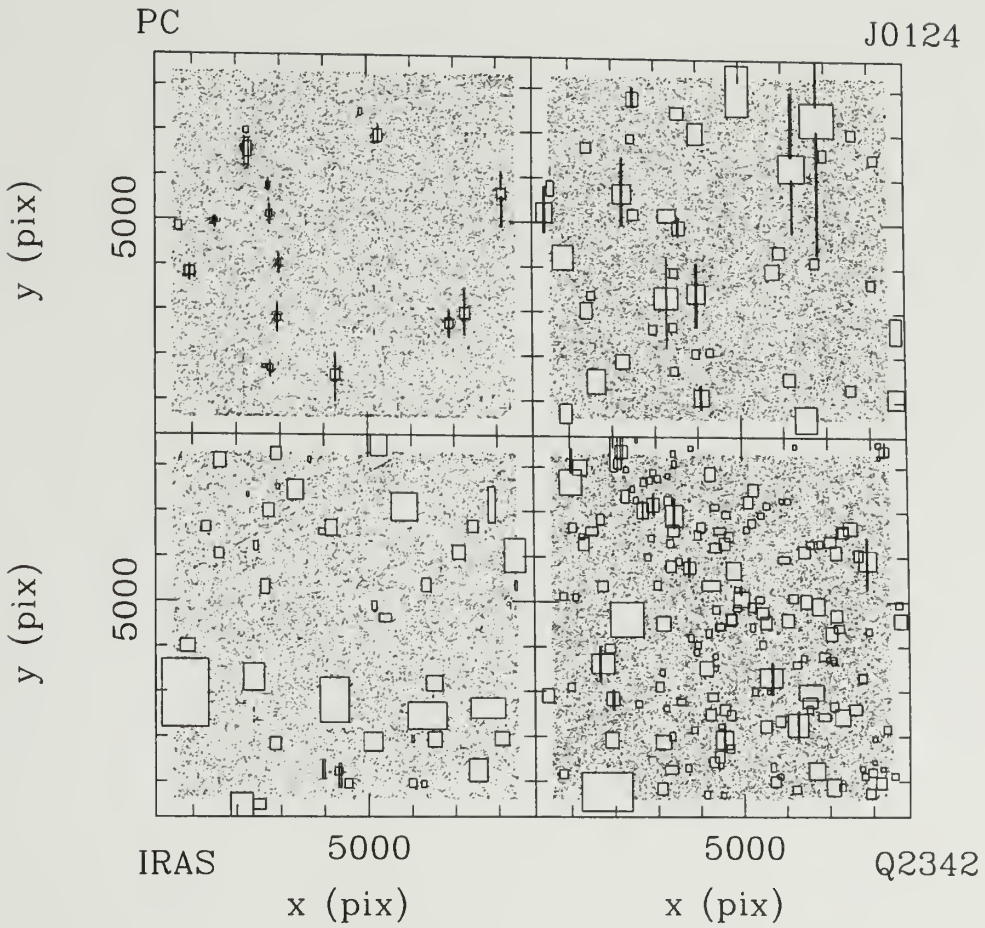

Figure 4.16 x–y position of all the 40,000 objects detected in the four fields, which shows the uniform sensitivity across the entire field of view. The rectangles show the masked areas due to bright stars that saturated the detector. Compare with Fig. 4.3.

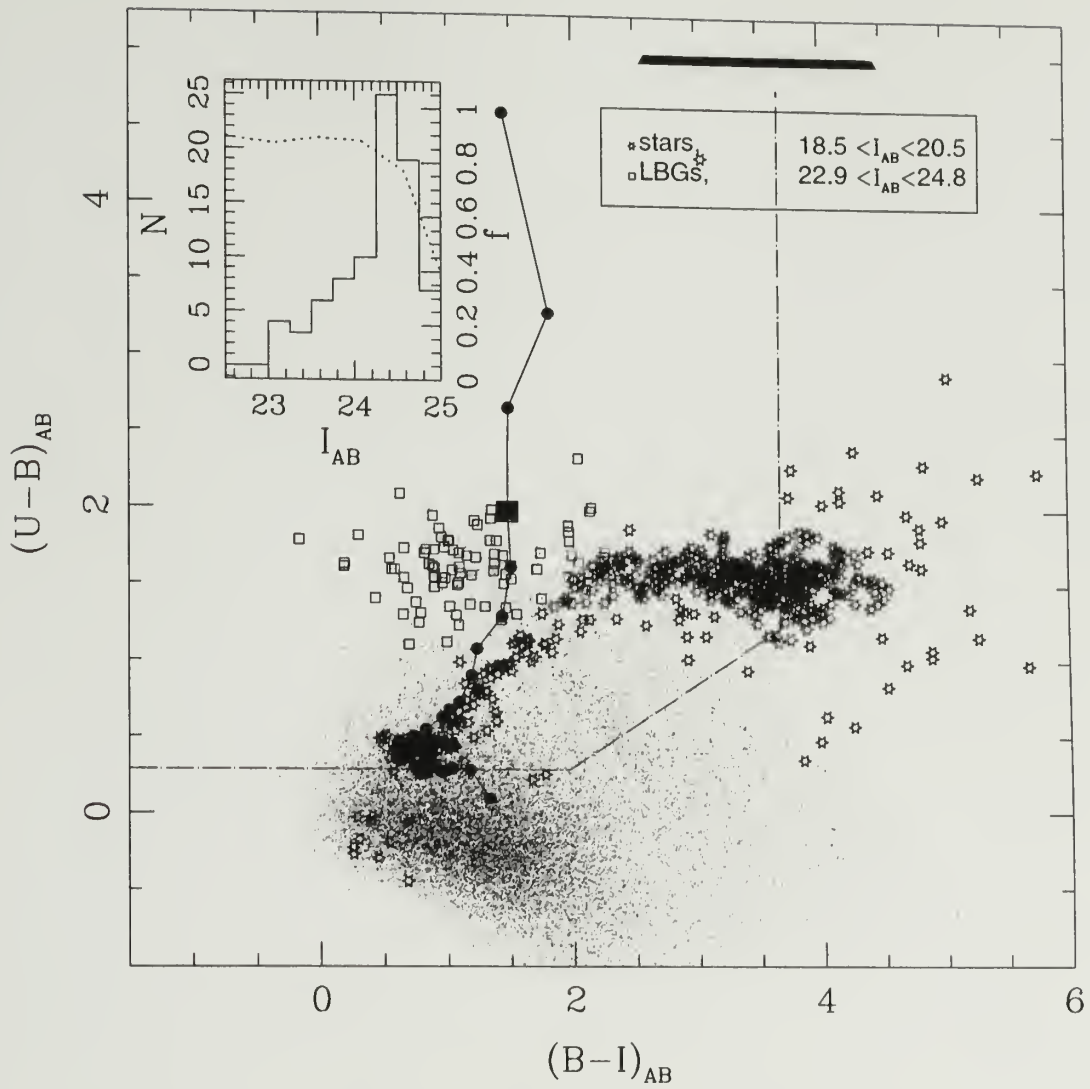

Figure 4.17 Color-color $(U-B)_{AB}$, $(B-I)_{AB}$ diagram for the field APM 087289+5255. Each dot represents one of our $\sim 30,000$ objects in our image with $23.0 < I_{AB} < 24.80$. Open squares are our LBG candidates in the redshift slice centered on the DLA ($z_{abs} = 2.974$), i.e. with $P_{\rm DLA} > 0.5$ (see text). Objects with U and B upper limits (filled triangles) have their $U - B$ set to an arbitrary constant (= 5), and the B magnitude limit is used in $B - I$. Stars show the locus of stellar objects with $18.5 < I_{AB} < 20.5$. The evolutionary track of an Irr SED (averaged over different dust content, from $A_v = 0$ to $A_v = 1.2$) is shown in redshift steps of 0.1 for illustrative purposes; the $z = 3$ mark is shown with the large filled square. We pre-selected objects within the color cut shown with the dot-dashed lines. The inset shows their number counts (N) as a function of magnitude (I_{AB}). The dotted line shows our completeness (f).

4.8 Stellar contamination

We did not perform any star-galaxy separation since we were searching for faint star-like objects. However, at $I < 22.5$, we were able to select point sources, those with FWHM of the source $< 1.15 \times$FWHM of seeing and with a morphological classification CLASS_STAR > 0.98 from SExtractor. This enabled us to determine empirically the stellar locus on the color-color diagram (Fig. 4.17) and to extrapolate the number counts of stars that fall in sample of LBG candidates at $2.75 < z < 3.25$.

In other words, to estimate our stellar contamination, we used our star counts up to $m_I \leq 23.5$ and extrapolated them to our magnitude limit, assuming the counts to be constant at $m_I > 23.5$. This gives an upper limit to our stellar contamination. We found the stellar contamination to be $< 7\%$, $< 15\%$, $< 20\%$, $< 13\%$, for the APM 08279+5255, the PC 1233+4752, the Q2342+3417, and the J0124+0044 field, respectively. It is not surprising that the Q2342+3417 has the highest contamination, since it has the lowest galactic latitude $b = -25\,\mathrm{deg}$ and the highest galactic extinction. For this reason, we suspect this field to be highly contaminated by stellar objects.

CHAPTER 5

THE LBG/DLA CONNECTION II. PHOTOMETRIC REDSHIFTS

At this point, we had a photometrically calibrated catalog of approximately 10,000 sources for each field. From this catalog, about 4,000 were at a redshift $2.5 < z < 3.5$. To study the clustering of galaxies that are near the DLAs, we needed to narrow the selection further to the hundred or so galaxies that are close in redshift to the DLA, i.e. within $< 100h^{-1}$ Mpc, This chapter is dedicated to the method used for this selection. Section 5.1 presents the photometric redshift technique. Sections 5.2 and 5.3 describe several optimizations of the technique. Section 5.4 shows the improvement in using Bayesian statistics. Finally, we show our selection of LBGs in section 5.6 and discuss the sample properties.

5.1 A description of the technique

Even with the recent advance of multi-object spectrographs on a 10-m class telescope, it is nearly impossible to obtain spectroscopic redshifts ($\delta z \leq 0.001$) of each of our 4,000 LBG candidates. Thus, it is often mandatory for faint galaxy surveys such as this one to rely on 'photometric' redshift techniques that have been developed and well refined.

There are two approaches to photometric redshift estimations, the empirical training set method and the spectral energy distribution (SED) fitting. The training set method (e.g. Connolly et al., 1995) starts with a multicolor galaxy sample that has been spectroscopically identified, from which an empirical relationship among color

C, magnitude m and redshift z is determined using a multi-parametric fit from the relation $z(C, m)$.

The SED fitting method (Lanzetta, Yahil, & Fernández-Soto, 1996; Sawicki, Lin, & Yee, 1997) is based on a set of templates (empirical or synthetic). These templates are redshifted, corrected for intergalactic and internal extinction, and then are compared with the observed galaxy colors to determine which redshift z is the best fit. The two methods are comparable in their performance at $z \leq 1$. The training set method is not always feasible at high-redshifts because (1) there are degeneracies in the colors of galaxies at very different redshifts and (2) this method cannot extrapolate to redshifts beyond the highest redshifts in the training set.

Photometric redshift codes using the SED fitting method were rapidly developed for the Hubble deep field (HDF) by Sawicki, Lin, & Yee (1997), Lanzetta, Yahil, & Fernández-Soto (1996) and Fernández-Soto et al. (2001) with an accuracy of typically $\Delta z \sim 0.06(1+z)$. The SED fitting method works best when there is a strong feature in the SED, such as the 4000Åbreak, or the 912Å, Lyman break. In other words, this method fits the shape of the SED, which in turn provides an estimate of the redshift $z_{\rm phot}$ via a χ^2 minimization.

In our case, we used the publicly available code 'Hyperz' (Bolzonella, Miralles & Pelló, 2000), which took a set of spectra or templates provided by the user. The spectra were corrected for intergalactic extinction due to the Lyman forest and for internal extinction due to dust in galaxies. The effect of the Lyman forest is discussed in § 5.2. For the internal extinction due to dust in galaxies, we used the Calzetti extinction curve (Calzetti et al., 2000) and varied extinction values A_V from 0 to 1.2. The spectra were then convolved with the filter response curves (Fig. 4.1) to obtain the expected fluxes $F_{{\rm temp},i}$. The χ^2 calculation considered all redshifts, spectral types, and extinction values. The minimization found the most likely redshift and the corresponding 'best' template via

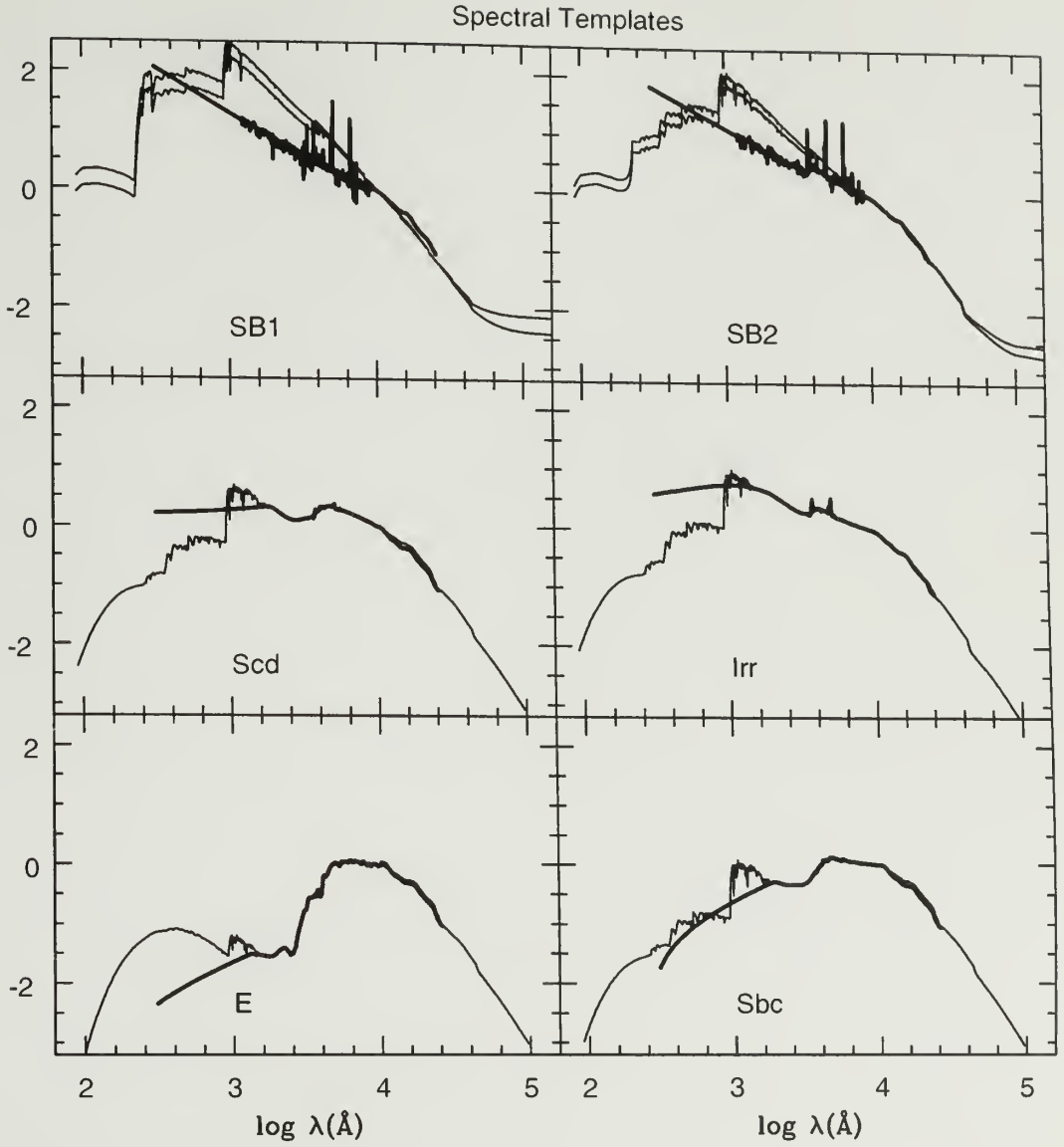

Figure 5.1 The four bottom panels show the empirical SEDs of four galaxy types (E, Sbc, Scd, & Irr) from Coleman, Wu, & Weedman (1980). The thick curves show the six templates used in the HDF analysis of Fernández-Soto et al. (2001), including the two starburst SEDs (SB1, SB2) (top panels). The thin curves show the SEDs that were used in this work. For the CWW SEDs, note the significantly different extension in the UV using theoretical models from Bruzual & Charlot (1993). We experimented with synthetic starburst SEDs from Leitherer et al. (1999); the ones with solar metallicity and 1/10th solar metallicity are shown on the top left and right panels, respectively, each with two different ages: 50 Myr and 100 Myr.

Table 5.1. Synthetic templates parameters

Parameters for the extension of CWW SEDs in the UV from BC93[a]			
SED type	SFR	age	IMF
E	exp ($\tau = 1$Gyr)	12.7Gyr	Miller-Scalo (0.1-100M$_\odot$)
Sbc	exp ($\tau = 5$Gyr)	12.7Gyr	Miller-Scalo (0.1-100M$_\odot$)
Scd	exp ($\tau = 15$Gyr)	12.7Gyr	Miller-Scalo (0.1-100M$_\odot$)
Irr	const	0.1Gyr	Schmidt ($n = 1$)
Parameters for the starburst SED of Starburst99[b]			
1.2	const	50Myr	Salpeter (1-100M$_\odot$)
2.3	const	100Myr	Salpeter (1-100M$_\odot$)
4.5	const	500Myr	Salpeter (1-100M$_\odot$)
6.7	const	800Myr	Salpeter (1-100M$_\odot$)

[a]Bruzual & Charlot (1993).

[b]Each for two different metallicities, solar and 1/10th solar.

$$\chi^2(z) = \sum_{i=1}^{N_{\rm filters}} \left[\frac{F_{{\rm obs},i} - b \times F_{{\rm temp},i}(z)}{\sigma_i} \right]^2 \qquad (5.1)$$

where $F_{{\rm obs},i}$, $F_{{\rm temp},i}$ and σ_i are the observed and template fluxes and their uncertainty in the filter i, respectively, and b is just a scaling factor. Fig. 5.2 shows an example of the fit to the observed photometry. This galaxy is best fitted with the Sbc galaxy type from the Coleman, Wu, & Weedman (1980) templates at a redshift $z_{\rm phot} \sim 2.94$ with an extinction $A_V = 0.2$. The inset to the left of Fig. 5.2 shows the likelihood distribution as a function of redshift. The redshift is clearly uniquely constrained. In some instances, like the one shown on Fig. 5.3, there is a degeneracy in the observed colors, the redshift likelihood distribution is broader and double peaked. We will discuss how we approached this problem in section 5.4.

It should be noted that this method is not an attempt to find the best fit to the photometry, but rather the best fit to the redshift. A bright galaxy with small photometric errors might not be well fitted by any template, its χ^2 may be much

larger than 1, but its most likely redshift can still be reliable. Therefore, large or small values of $\chi^2_{\rm min}$ are not necessarily meaningful (as discussed in Fernández-Soto et al., 2001).

5.2 Towards better photometric redshifts

We updated the code by using the more recent prescription of Massarotti et al. (2001) for the intergalactic extinction. The extinction due to the intergalactic medium is usually given by the optical depth $\tau_{\rm eff}$. This effect comes from the cumulative absorption of photons in the intergalactic medium due to line blanketing of the Lyα forest. The effective optical depth $\tau_{\rm eff}$, defined from:

$$F_{rmobs}(\lambda, z) = F_{\rm int}(\lambda) \exp(-\tau_{\rm eff}(z)),$$

describes the cumulative absorption of photons, where $F_{\rm obs}$ is the observed flux, and $F_{\rm int}$ is the intrinsic flux.

Hyperz uses the prescription of Madau (1995), whereas Massarotti et al. (2001) has shown that using

$$\tau_{\rm eff}(z) = A\left(\frac{\lambda_{\rm rest}(1+z)}{1216\text{Å}}\right)^{1+\gamma} \text{ with } \begin{cases} A = 0.0036, \gamma = 2.46 & \text{at } z > 4 \\ A = 0.00759, \gamma = 1.35 & \text{at } z < 3 \end{cases}, \quad (5.2)$$

with a linear interpolation between the two regimes ($3 \leq z \leq 4$), i.e. $A = 0.00759 - 0.00399(z-3)$ and $\gamma = 1.35 + 1.11(z-3)$, the photometric redshift estimates are improved. The resulting Lyα flux decrement $D_A(z)$, defined as

$$D_A(z) = 1 - \frac{1}{\Delta\lambda}\int_{1050\text{Å}(1+z)}^{1170\text{Å}(1+z)} \exp(-\tau_{\rm eff})\mathrm{d}\lambda, \quad (5.3)$$

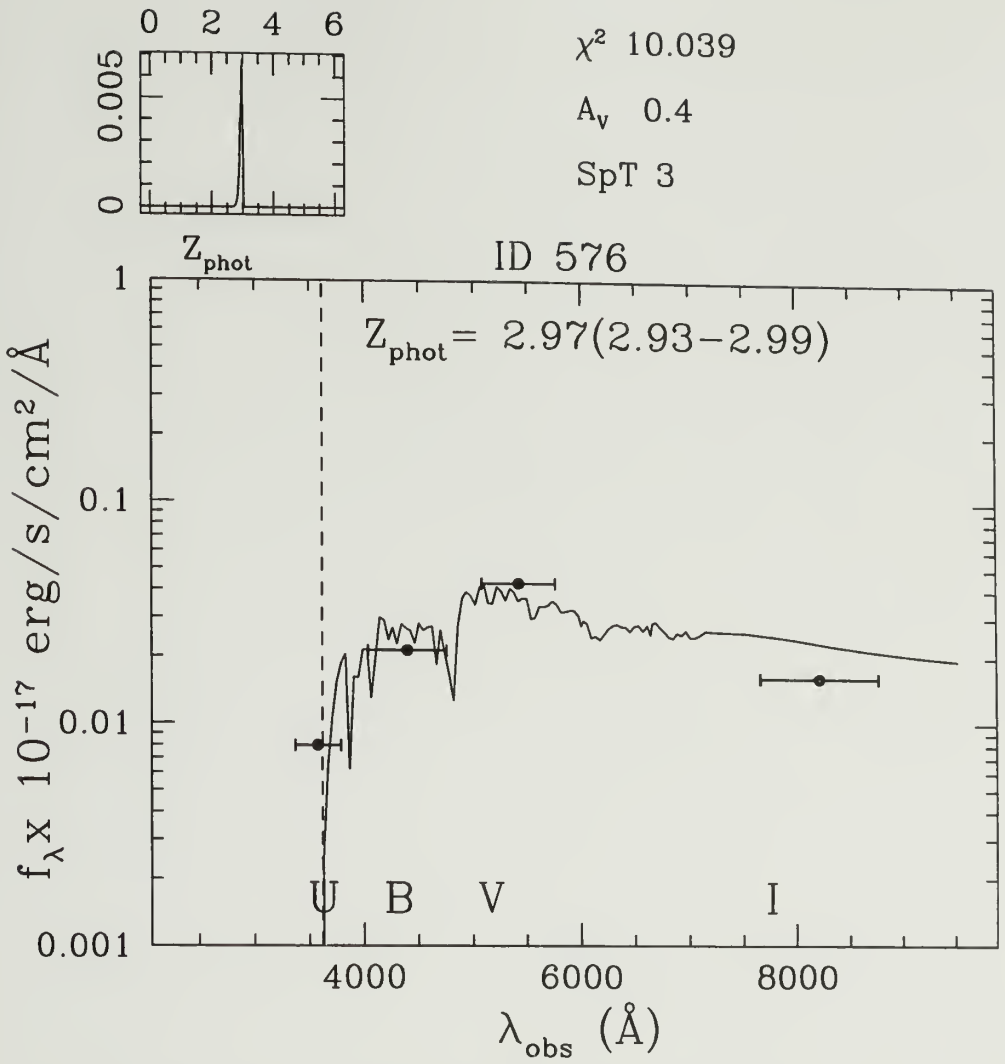

Figure 5.2 An example of a solution to the template fitting using Eq. 5.1 of a galaxy at $z_{\rm phot} \sim 2.94$ (the 1 sigma range is shown in parentheses). The observed flux for the four filter bands (UBV&I) are shown as the filled circles. Error bars in our flux measurement are shown but are smaller than the filled circles. Error bars in wavelength shows the width of the filters. The vertical dashed line shows the rest-frame Lyman break at 912A that enters the U band at this redshift. This break in the energy distribution is responsible for the redshift constraint. The inset on the top left shows the likelihood distribution as a function of photometric redshift. It is very narrow, and the redshift is well constrained. The top right shows the χ^2, the fitted extinction A_V, and the best fit template (SpT=3 corresponds to Scd).

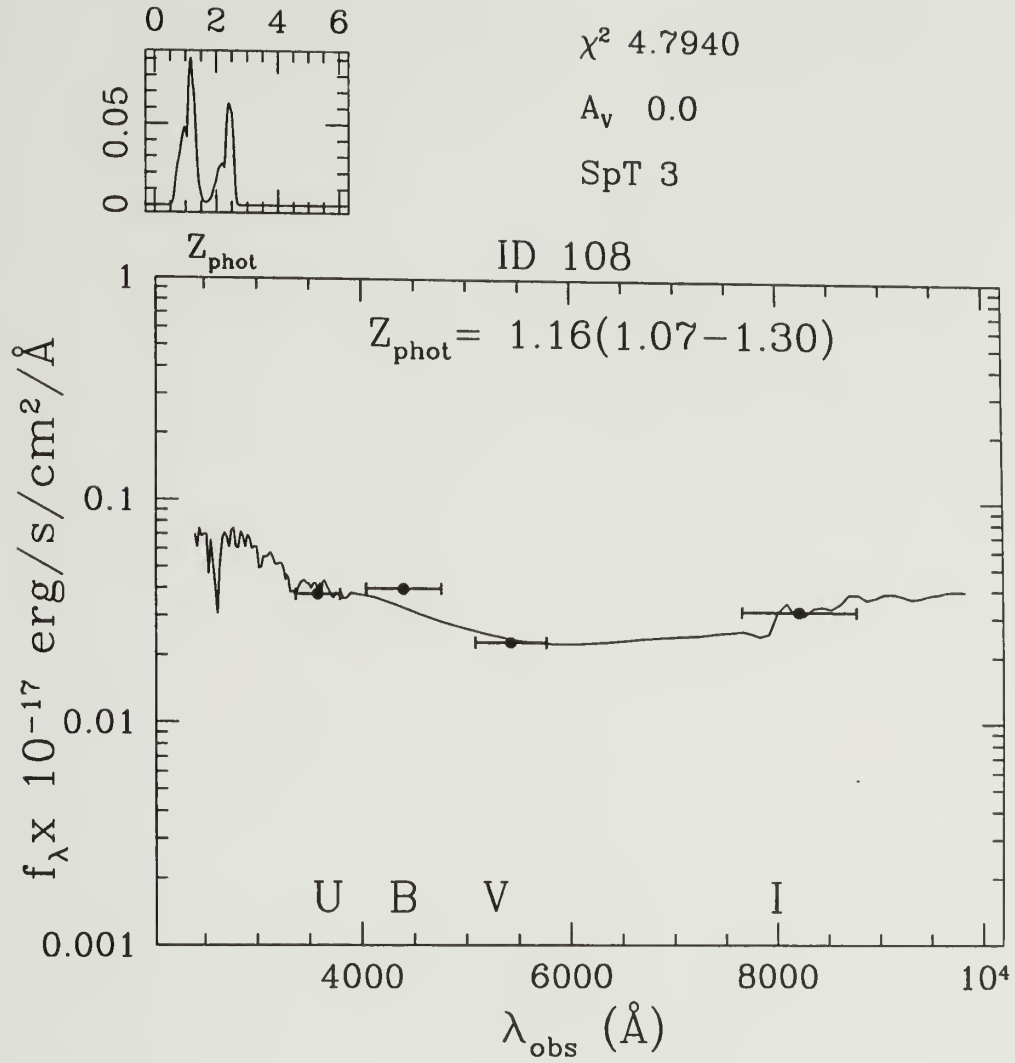

Figure 5.3 Same as Fig. 5.2, this time for a galaxy whose redshift likelihood is degenerate. The redshift is not as well constrained as that in Fig. 5.2. Section 5.4 presents a solution to this type of degeneracy.

with $\tau_{\rm eff}$ from Eq. 5.2 fits better the observed absorption and is shown in Fig. 5.4. Hyperz takes into account line blanketing for other Lyβ and higher order lines, which we have not modified.

5.3 Towards the right set of templates: optimization on the HDF

Two important questions need to be answered regarding the photometric redshift technique:

- What is the best set of templates?
- Do we need near infra-red (IR) data?
- How reliable are they?

To answer these questions, we used our photometric redshift technique on the Hubble deep field North (HDF-N), where there was a large sample of spectroscopically confirmed galaxies up to redshifts $z \leq 6$ and to magnitudes $R = 24$ (from the compilation of Cohen et al., 2000). Fernández-Soto et al. (2001) noted that this is the cleanest, deepest, most complete list of spectroscopic redshifts for the HDF available to date. The sample consisted of ∼150 galaxies and 18 galaxies at $2.75 < z < 4.5$ and enabled us to perform a few checks between spectroscopic and photometric redshifts.

What are the best templates?

Since photometric redshift techniques rely on template fitting, one has to use templates that describe the SED of real galaxies in a large redshift range. There are usually three ways to tackle this problem: i) one can use empirical templates from a set of observed galaxies (such as Coleman, Wu, & Weedman, 1980); ii) one can use synthetic templates from galaxy models such as the ones from Bruzual & Charlot (1993) or from Starburst99 (Leitherer et al., 1999); or iii) one can use a combination of the two approaches.

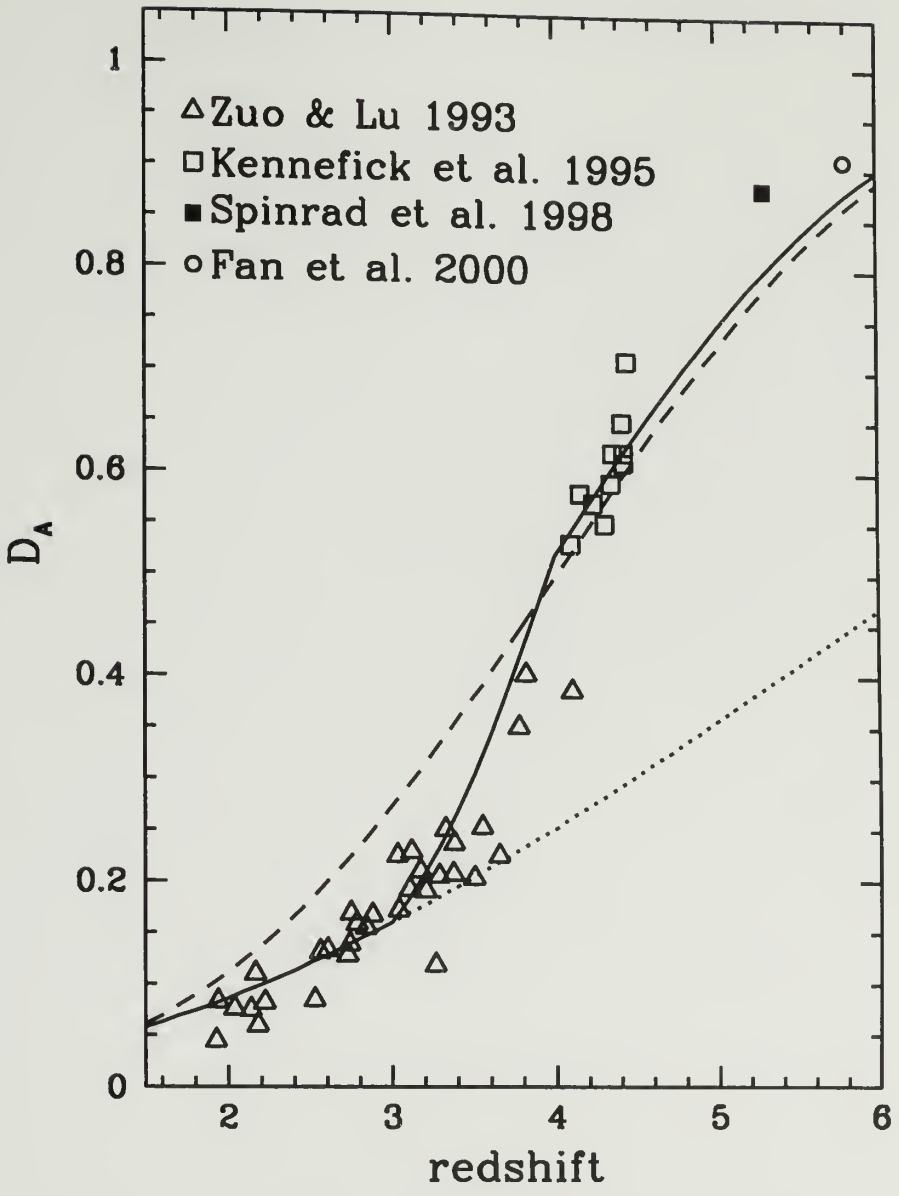

Figure 5.4 The flux decrement $D_A(z)$ due to the accumulated absorption of the Lyα forest. The dashed lines are obtained by the prescription of Madau (1995), while the dotted lines are those by Scott, Bechtold, & Dobrzycki (2000). The solid line is the model adopted by Massarotti et al. (2001). References for the data points are indicated. (Reproduced from Massarotti et al., 2001) .

The empirical approach is naturally favored since the templates are observed. Indeed, the four templates of Coleman, Wu, & Weedman (1980) (CWW) are based on an average of observed SED of a sample of four classes of nearby galaxies – Ellipticals (E), Spirals (Sbc, Scd), and Irregulars (Irr). They are shown in Fig. 5.1. Because those SEDs were observed in the local universe from ground-based telescopes, these four templates cover only the range 1400-10000Å. Thus, one needs to extend these templates at wavelengths $\lambda < 1400$Å since, at redshift $z > 2$, the observed galaxy colors are sampling the rest-frame UV. For instance, Fernández-Soto et al. (2001) extended them using a power law following Kinney et al. (1993), shown in Fig. 5.1 by the thick lines. A power law is a crude approximation for the region where we expect a strong Lyman break at 912Å. Instead, we used the CWW templates that were extended by Bolzonella, Miralles & Pelló (2000) using the synthetic models of Bruzual & Charlot (1993) with parameters (SFR and ages) that matched the observed spectra at $z = 0$. The parameters are listed in Table 5.1.

We experimented with several SED combinations and tested the following set of templates:

A: 4 CWW templates extended in the UV using Bruzual & Charlot (1993) models,

B: Same as A, with 2 empirical starburst (SB) templates from Kinney et al. (1996),

C: Same as A, with 2 synthetic SB templates from Starburst99 (Leitherer et al., 1999),

D: Same as A, with 4 synthetic SB templates from Starburst99 (shown in Fig. 5.1 on the top two panels),

E: 4 CWW templates extended in the UV using Kinney et al. (1993) and with 2 empirical starburst (SB) templates from Kinney et al. (1996) (This is the set used by Fernández-Soto et al. (2001) which is shown in Fig. 5.1 by the thick lines),

Table 5.2. Template sets list.

Template set ID	4 CWW	+UV	additional SED	Other name
A	yes	BC93		'1'
B	yes	BC93	plus 2 SB (K96)	'10'
C	yes	BC93	plus '3,4' from Table 5.1	'11'
D	yes	BC93	plus '1-4' from Table 5.1	'7'
E	yes	K93	plus 2 SB (K96)	'8'
F	no	n.a.	15 from BC93	'4'
G	no	n.a.	7 SB from L99	'5'

References. — CWW: Coleman, Wu, & Weedman (1980); BC93: Bruzual & Charlot (1993); K93: Kinney et al. (1993); K96: Kinney et al. (1996); L99: Leitherer et al. (1999)

F: 15 templates from the models Bruzual & Charlot (1993) with varying ages,

G: 7 SB templates from Starburst99.

The parameters for each template set are listed in Table 5.2.

Figs. 5.5, 5.6 and 5.7 show the results of our tests in the HDF for template sets A, F and G respectively. The left panels show our photometric redshifts $z_{\rm phot}$ vs. spectroscopic redshifts $z_{\rm spec}$ using the $UBVI$ photometry (bottom), adding the near-IR K-band, $UBVI + K$ (middle) and using the three IR JHK-bands (top). The right panels show the distribution of the residuals $\frac{z_{\rm spec}-z_{\rm phot}}{1+z_{\rm zspec}}$.

Notice that the four CWW templates (Fig. 5.5) describe well the SEDs of galaxies at high redshifts regardless of the use of IR photometry. On the other hand, the purely synthetic templates (Fig. 5.6) underestimate the redshifts at $z_{\rm spec} > 1.5$ when using only the four optical bands. This is due to degeneracies in the SEDs between age and metallicity: a young galaxy has an SED similar to an older one with a low metallicity. The situation was no better for the pure 'starburst' templates (Fig. 5.7). In this case, the templates describe well galaxies with redshifts $z_{\rm spec} > 2$, but assigned

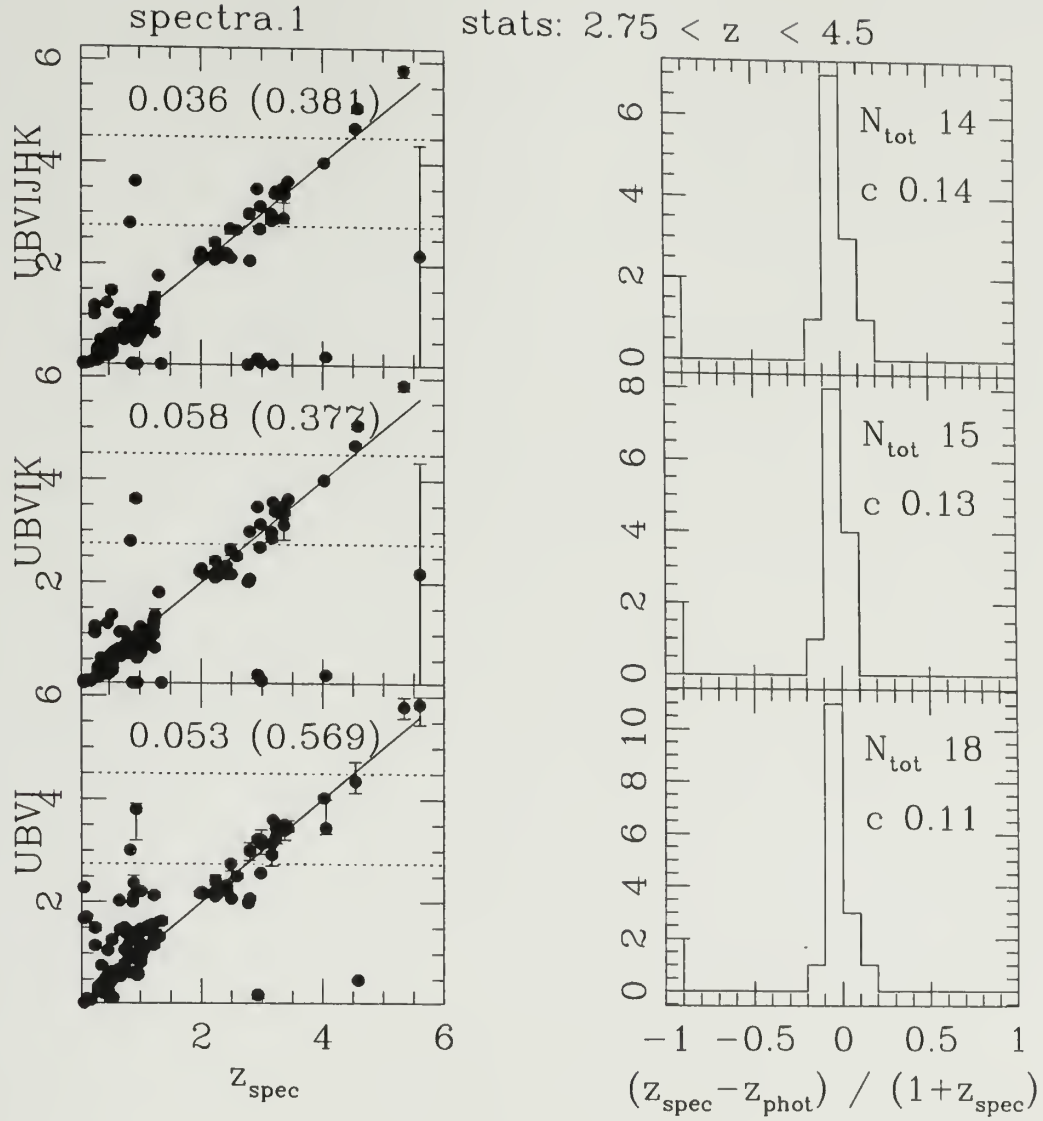

Figure 5.5 Photometric redshift test on the HDF in the redshift range $2.75 < z < 4.5$ based on template set A. Left: photometric redshift ($z_{\rm phot}$) as a function of spectroscopic redshift ($z_{\rm spec}$) using the photometry from the $UBVI$ bands (bottom), the $UBVI + K$ bands (middle) and the $UBVI + JHK$ bands (top). The rms of the residual distribution (rejecting outliers) is shown in each panel. Right: the residual distribution $(z_{\rm spec} - z_{\rm phot})/(1 + z_{\rm spec})$. $N_{\rm tot}$ is the number of objects that fall within $2.75 < z < 4.5$. The numbers preceded with a c indicate the contamination fraction of low redshift object. From the left panels, one can conclude that adding the IR photometry does not improve greatly the photometric redshifts at $z > 2.75$, while adding IR is a clear improvement at lower redshifts $z < 1.5$ (see text).

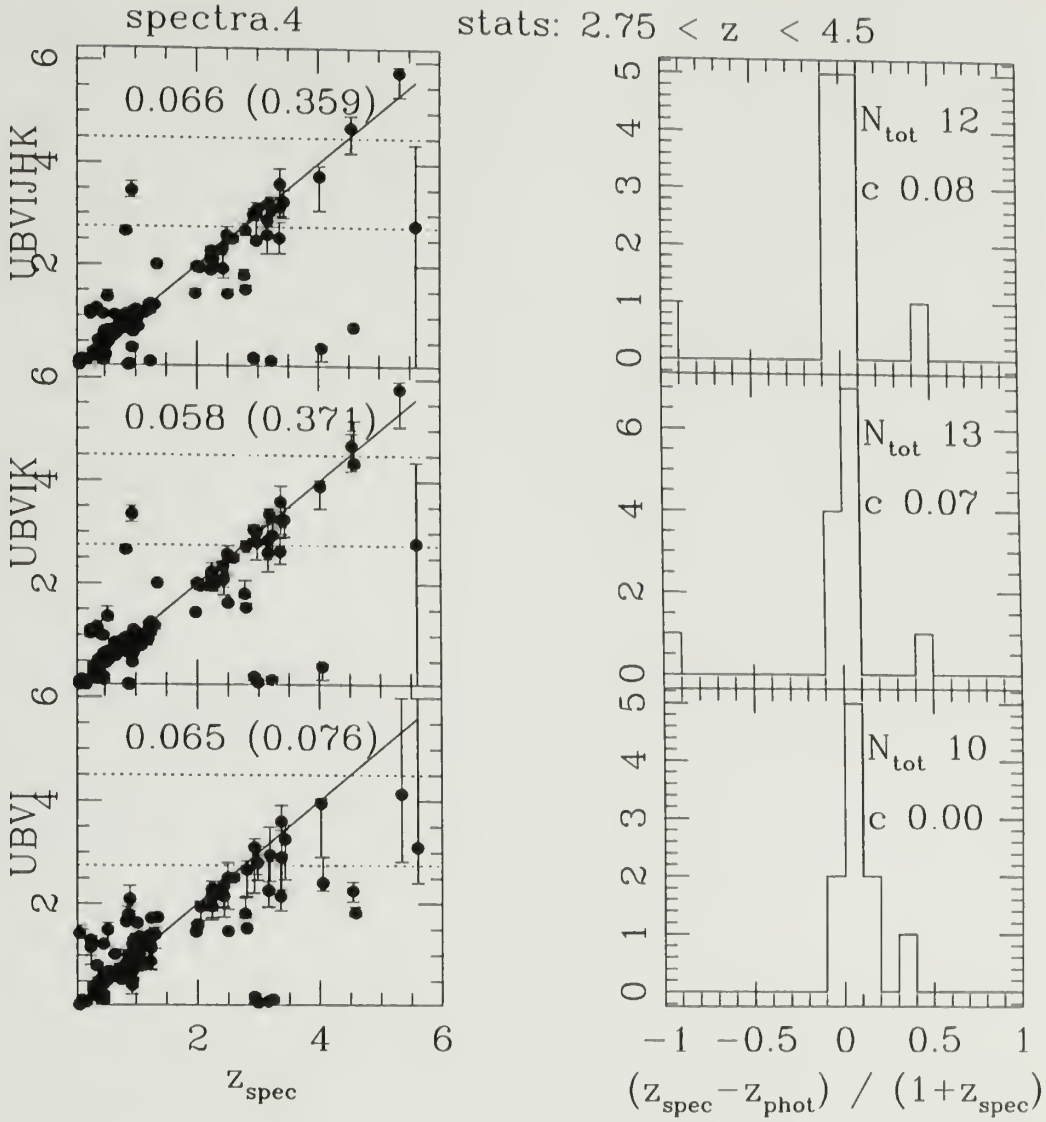

Figure 5.6 Same as Fig. 5.5 for the template set F based on SEDs from the Bruzual & Charlot (1993) synthetic models and with evolution (see text). Because of the degeneracies in these templates when using the four optical bands, photometric redshifts are consistently underestimated at $2 < z_{\rm spec} < 4$ (lower left panel). For this reason, we rejected this set of templates in the rest of our analysis.

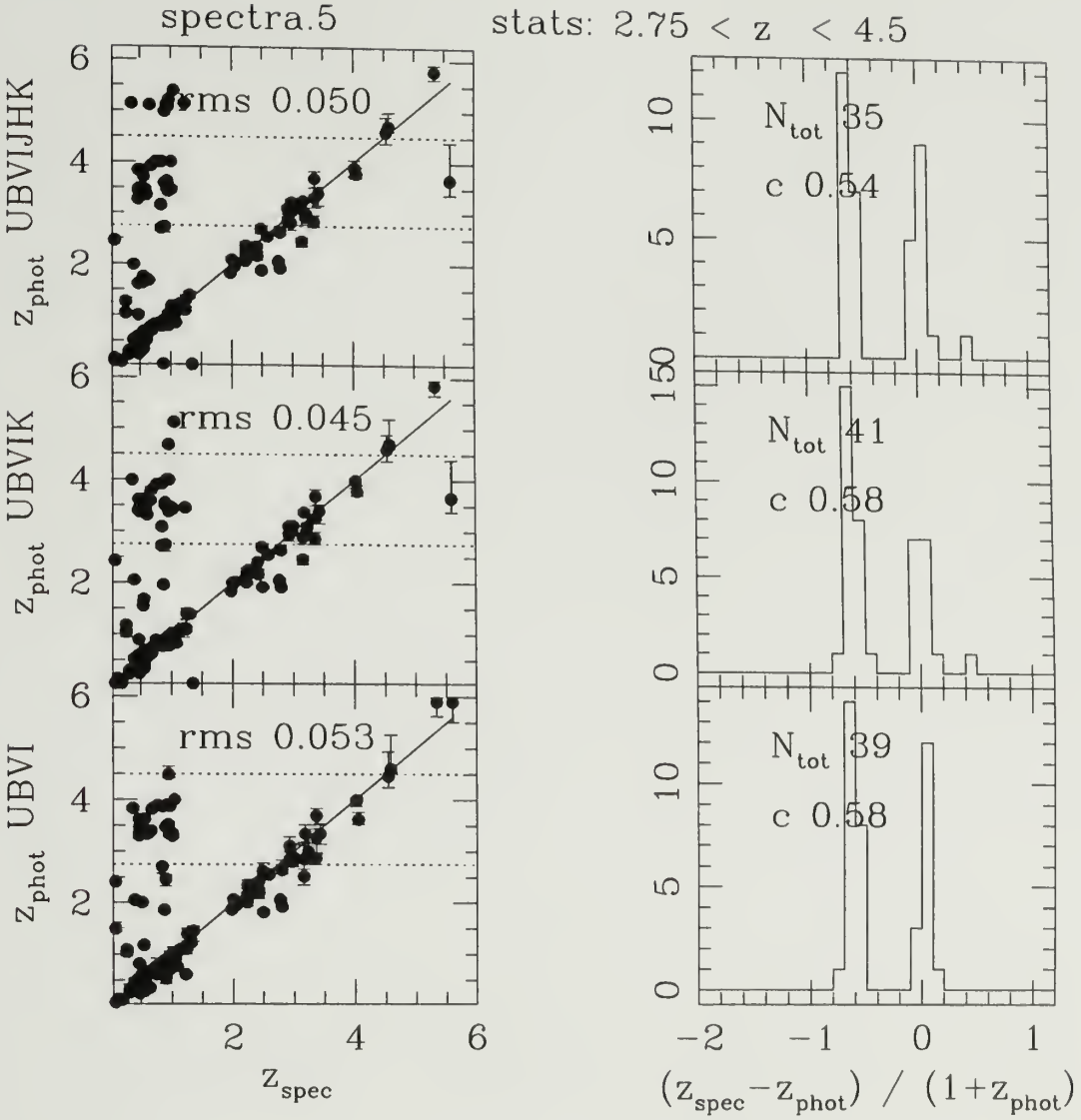

Figure 5.7 Same as Fig. 5.5 for the template set G based on synthetic templates from Starburst99. One sees that these templates work well at high-redshift, while they produce a large number of outliers at low redshifts, that contaminated the high-z sample. For this reason, we rejected this set of templates in the rest of our analysis.

a faulty high photometric redshift to a large fraction of low redshift $z_{\rm spec} < 1$ galaxies. Thus, pure starburst templates did not reproduce the SEDs of galaxies from $z = 0$ to $z = 6$. Therefore, we rejected template sets F and G.

To figure out the performance of these various template sets, we performed the same analysis, i.e. computed the mean and the rms of the residuals in different redshift ranges for each of our template sets (A, B, C, D & E). Because, ultimately, we will have only photometric redshifts in our data, we computed the mean and the rms of the residuals for galaxies in a given photometric redshift $z_{\rm phot}$ range (as opposed to $z_{\rm spec}$ range). The results are shown in Fig. 5.8 and 5.9 for two redshift ranges, at $0 < z_{\rm phot} < 2.75$ and at $2.75 < z_{\rm phot} < 4.5$ respectively. Each figure shows the mean of $\Delta_z/(1+z)$ (bottom panel) and the rms of $\Delta_z/(1+z)$ (top panel) for the five template sets. The filled circles are for when we used only the four optical bands ($UBVI$), the green squares are for when we used the K band in addition to the four optical bands, and the red triangles are for when we used the three IR bands JHK in addition to the optical bands.

Are photometric redshifts biased?

In Figs. 5.8–5.9 (bottom panels), we see that there is no systematic offset, at either low redshifts or high redshifts. The technique does not over- or under-estimate the redshift.

The best template set at high redshifts

At high redshifts, which is our redshift range of interest, the black points in Fig. 5.9 show the best template set. Indeed, ignoring the results with IR photometry, the sets B, C and E have the worst rms, and a close look at Fig. 5.9 shows that the set with the smallest rms is the template set 'A'. This set is the 4 CWW template set shown in Fig. 5.10). We therefore can conclude from Fig. 5.5 that these four templates are describing accurately the colors of high redshift galaxies.

Do we need infrared photometry?

Another important aspect of photometric redshift techniques is the following: Since the rest-frame UV of high redshift galaxies is observed in the optical, is it important to have rest-frame optical photometry, i.e. IR photometry?

The answer to this question is obtained by comparing the top panels of Fig. 5.8 and Fig.5.9. At low redshifts, we see from Fig. 5.8 that the rms of the residuals is decreased if one includes the IR data in the photometric redshift estimate. IR adds more information and breaks some of the degeneracies in the redshift estimates.

Fig. 5.9 shows that, at high redshifts, the situation is very different. The rms of the residuals in the redshift range $2.75 < z_{\rm phot} < 3.5$ does not improve so significantly if one includes IR in the photometric redshift estimate.

Therefore, we conclude that IR is not needed at $z > 2.75$. This is an important result since we do not have IR data covering our fields because IR detectors are still much smaller than optical. With the current technology, it would be a daunting task to seek to obtain IR of sufficient depth on one square degree, the area covered by our data.

5.4 The Bayesian approach

The SED fitting method is solely based on the χ^2 fit between the observed colors and the templates. In the case where more than one combination of redshift and morphological type have similar colors, the goodness of the fit and the likelihood would have more than one extremum. However, there is additional information that we have discarded in the maximum likelihood method, which could potentially help to resolve such ambiguities. For instance, a bright galaxy is more likely to be nearby than a faint one. More distant (fainter) ones are more likely to have a certain spectral type than others. In other words, it may be known from previous experience that

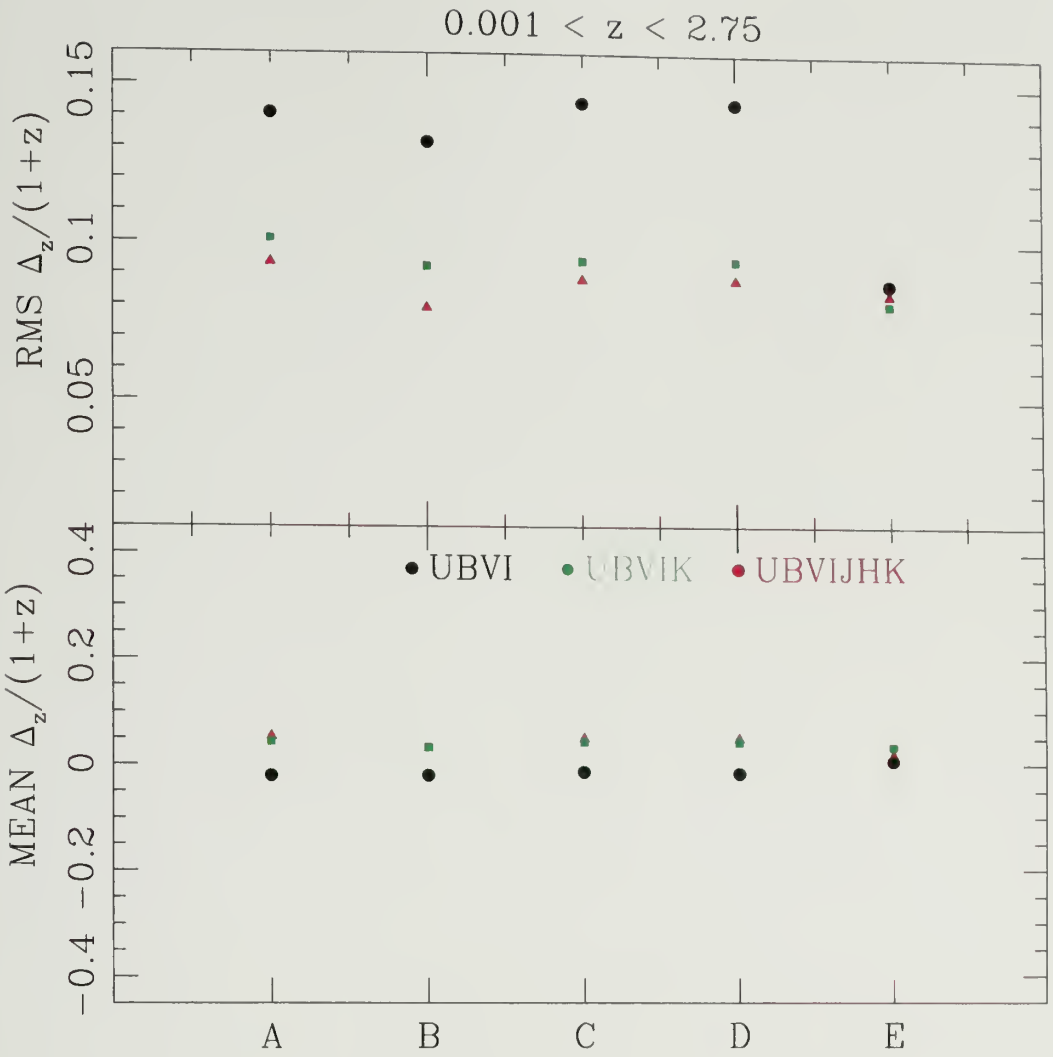

Figure 5.8 Top: rms of the photometric redshift residuals for our five template sets in the redshift range $0 < z_{\rm phot} < 2.75$. The black filled circles show the results when using the $UBVI$ bands, the green filled squares when using $UBVI+K$ bands and the red filled triangles when using $UBVI+JHK$ bands. Adding the IR bands increases the amount of information in the SED fitting, and the rms of the residuals decreases. The bottom panel shows the mean of the residuals for the five template sets and shows that there is no systematic bias.

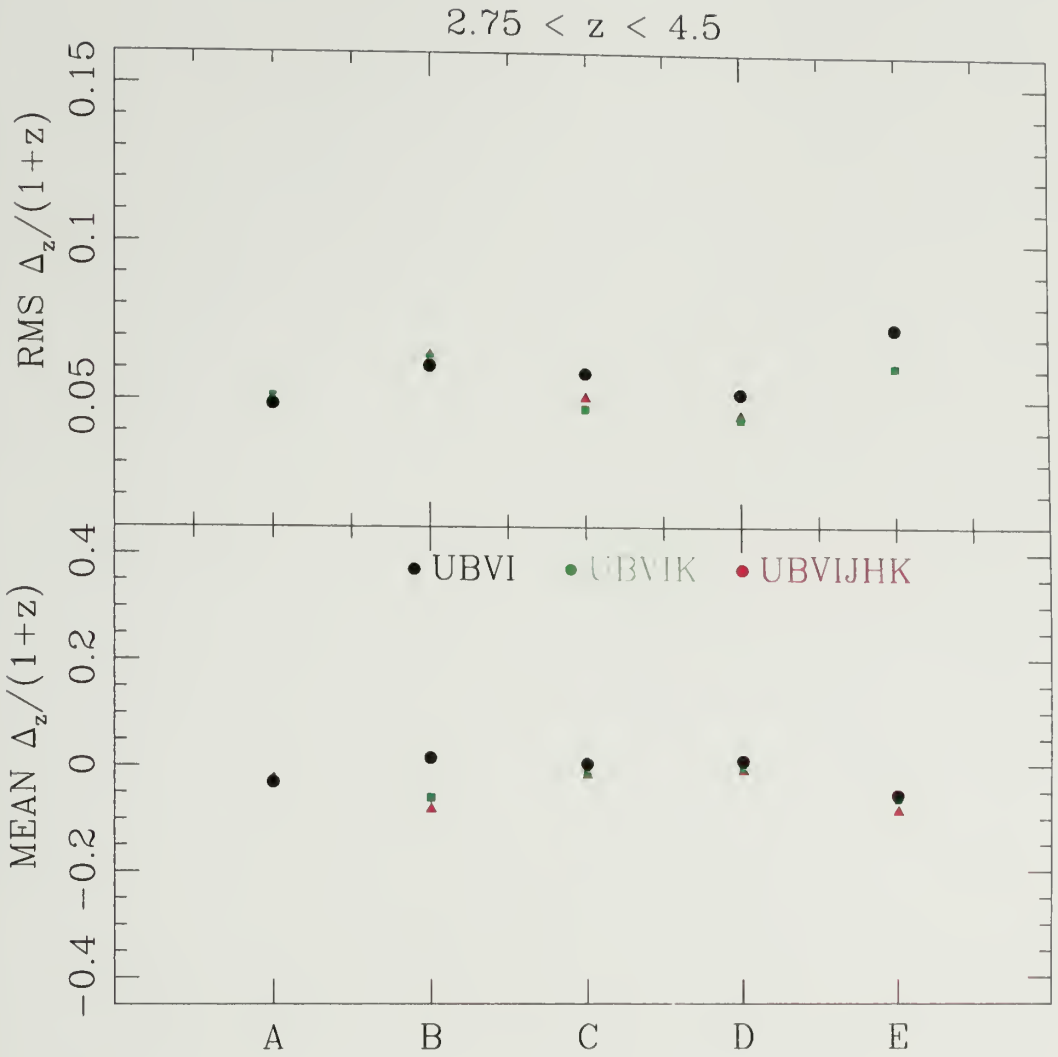

Figure 5.9 Same as 5.8 in the redshift range $2.75 < z_{\rm phot} < 4.5$.

one of the possible redshift/spectral type combinations is much more likely than any other, given the galaxy magnitude, angular size, shape, etc.

In that case, Bayesian probability states that the best option would be the one more likely *a priori*. The Bayesian probability theory allows one to include this additional information in a rigorous and consistent way. This method was developed by Benítez (2000) and we summarize it here.

From the above discussion, our interest is now the probability $p(z|C, m_o)$ of a galaxy to be at a redshift z, given its colors C *and* its magnitude m_o. From Bayes' theorem, this probability can be split into:

$$p(z|C, m_o) \propto p(C|z)p(z|m_o) \tag{5.4}$$

where the first factor, $p(C|z) \equiv \mathcal{L}(z)$, is simply the likelihood of observing the colors C if the galaxy is at redshift z, which is given by $\mathcal{L}_T(z) \equiv P(\chi^2(z)) \propto \exp(-\chi^2/2)$ of Eq. 5.1, and the second factor, $p(z|m_o)$, is the prior probability of being at redshift z given its magnitude m_o.

A galaxy belongs, however, to a specific morphological class, which is represented by a template in a particular set of spectra (e.g. { E, Sbc, Scd, Irr, ...}). Note that this set is exclusive: a galaxy cannot belong to two types at the same time. Under these assumptions, the probability $p(z|C, m_o)$ can be expanded into a sum of basis functions formed by $p(z, T|C, m_o)$, which is the joint probability of the galaxy redshift being z and the galaxy type being T. Therefore, Eq. 5.4 becomes

$$p(z, T|C, m_o) \quad \propto \quad p(C|z, T)p(z, T|m_o) \tag{5.5}$$

and the sum of all these basis functions gives the total probability:

$$p(z|C, m_o) \quad = \quad \sum_T p(z, T|C, m_o) \propto \sum_T p(C|z, T)p(z, T|m_o) \tag{5.6}$$

where $p(C|z,T)$ is the probability of observing the colors C if the galaxy is at redshift z and has a type T and is the likelihood $\mathcal{L}_T(z)$. This is given by $\mathcal{L}_T(z) \propto \exp(-\chi^2_T/2)$, χ^2_T is Eq. 5.1 using solely one template type. The prior $p(z,T|m_o)$ is expanded using the product rule:

$$p(z,T|m_o) = p(z|m_o,T)p(T|m_o) \tag{5.7}$$

where $p(T|m_o)$ is the type fraction as a function of magnitude and $p(z|T,m_o)$ is the redshift distribution for galaxies of a given spectral type and magnitude.

The 'photometric' redshift $z_{\rm phot}$ can be taken from the maximum, the mode or the mean of the probability distribution $p(z|C,m_o)$. We elected to use the maximum z_m of the likelihood distribution $p(z,T|m_o)$ because it is independent of the shape of the likelihood distribution.

Following Benítez (2000), one can compute the probability of each galaxy being an outlier in the $z_{\rm spec} - z_{\rm phot}$ diagram from the likelihood distribution $p(z|C,m_o)$. Indeed, the contribution to the likelihood from redshifts outside the typical rms for photometric redshifts Δz, is given by

$$\begin{aligned} P(\text{outlier}) &= 1 - P_{\Delta z} \\ &= 1 - \int_{z_b-\Delta z}^{z_b+\Delta z} P(z), \end{aligned} \tag{5.8}$$

where $P(z) \equiv p(z|C,m_o)$, and $\Delta z = 0.2 \times (1+z_m)$. In other words, $P_{\Delta z} \equiv P(|z - z_m| < 0.2(1+z_m))$ gives the 'goodness' of the photometric redshift. The factor 0.2 is somewhat arbitrary. Since, the rms of photometric redshifts is typically $\sim 0.06(1+z)$, this 0.2 factor corresponds to approximately 3σ.

Fig. 5.11 shows an example of using the priors on the galaxy shown in Fig. 5.3. This galaxy redshift likelihood distribution has two maxima, its photometric redshift is degenerate. The four left panels of Fig. 5.11 show the two factors that come into

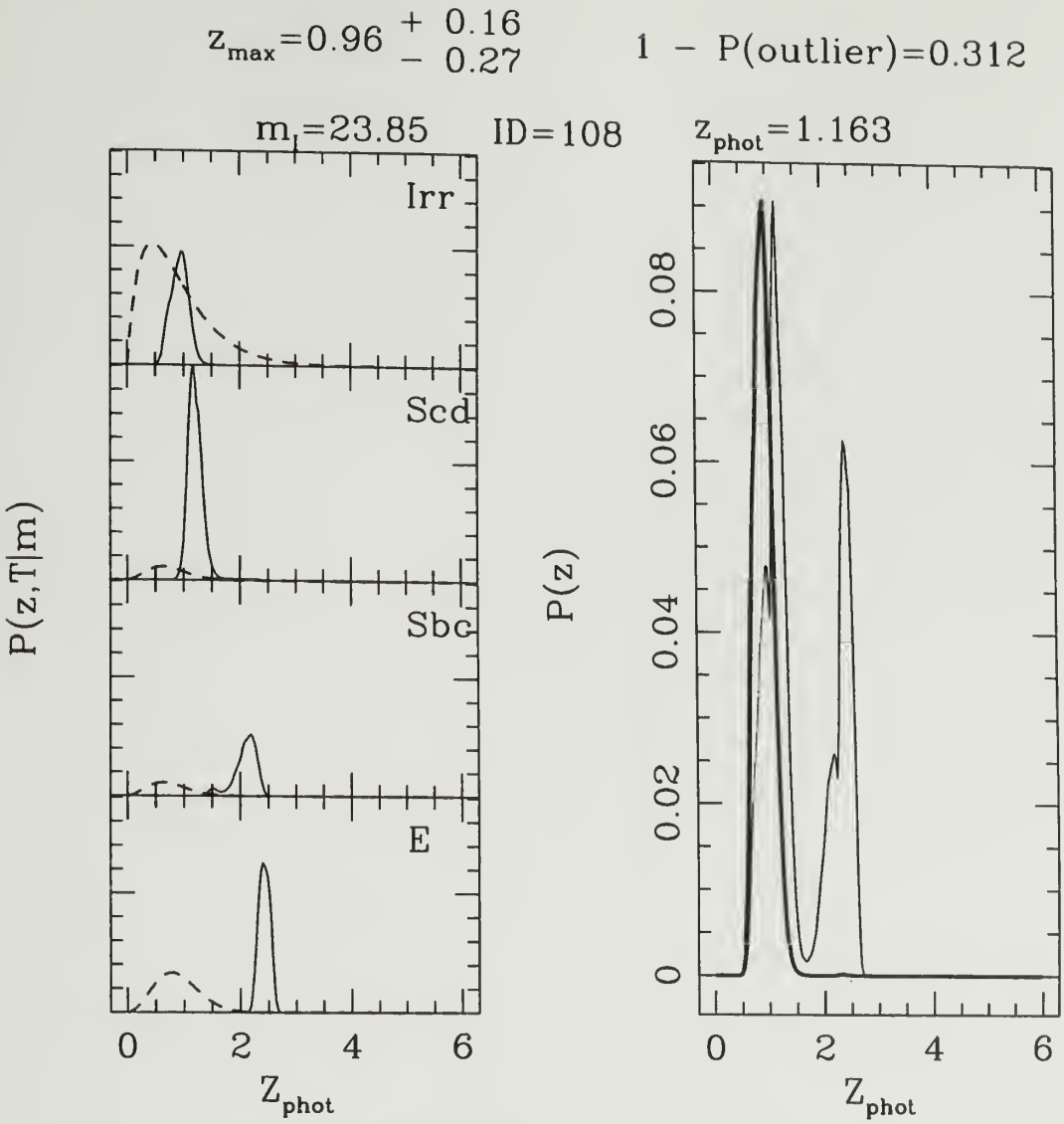

Figure 5.11 Example of the effect of using the priors on the galaxy shown in Fig. 5.3. Left: For each of the four CWW templates, the four panels show the two factors of Eq. 5.6. The dashed curves show the prior probability $p(z, T|m_o)$ given the galaxy magnitude $m_I = 23.85$. The continuous curves show the likelihood distribution $p(C|z, T)$ for each of the four CWW templates. Right: The thin curve shows the likelihood distribution one would find without using the priors (as in Fig. 5.3). The photometric redshift $z_{\rm phot}$ is shown. Multiplying the two curves on each left panel as in Eq. 5.6, and then summing them over the four types T, produces the thick curve shown in the right panel. The degeneracy has been broken. The photometric redshift is obtained from the max of thick curve $z_{\rm max}$ and is shown on the top left with the 1σ range. The 'goodness' of the redshift distribution is given by $1 - P$(outlier) from Eq. 5.8, and is low in this case because the likelihood distribution is wide.

Eq. 5.6. The dashed curves show the prior probability $p(z, T|m_o)$ given the galaxy magnitude $m_I = 23.85$ using the calibration presented in § 5.4.1. The continuous curves show the likelihood distribution $p(C|z, T)$ for each of the four CWW templates. The thin curve on the right panel shows the likelihood distribution one would find using only the color and not the priors (as in Fig. 5.3). This is given by $p(z) = \sum_T p(z|C, T)$, which is the sum over the four likelihood distributions shown on the left. Using the priors, each of the continuous curves on the left panels was multiplied by the dashed curves of that panel as in Eq. 5.6, and then summed over the four types T. This resulted in the thick curve on the right panel. The degeneracy was broken.

Since the best photometric redshift $z_{\rm phot}$ occurs from a linear combination of the templates, we computed the averaged extinction $A_V(z)$ and the average SED type $T(z)$ as follows:

$$A_V(z) = \sum_T \frac{\mathcal{L}_T(z)p(z, T|m_o)}{P(z)} A_V(z, T) \tag{5.9}$$

$$T(z) = \sum_T \frac{\mathcal{L}_T(z)p(z, T|m_o)}{P(z)} T \tag{5.10}$$

using the Bayesian likelihood distribution $P(z) = \sum_T \mathcal{L}_T(z)p(z, T|m_o)$ (Eq. 5.6) where $\mathcal{L}_T(z)$ is the likelihood of observing the observed colors C at redshift z for an SED T given by the χ^2 minimization.

We had to modify slightly the output of 'Hyperz' in order to couple it with the likelihood distribution for each type $\mathcal{L}_T(z)$ with our implementation of the Bayesian approach. Next, we present the parametric form for Eq. 5.7.

5.4.1 Parameterization

Benítez (2000) has calibrated the priors (Eq. 5.7) on a large sample of galaxies with spectroscopic redshifts from the Canada France redshift survey (CFRS) and from the Hubble deep field North (HDF-N). The HDF-N data set is formed by 737 galaxies

with $20 < I < 27$, including 130 objects with spectroscopic redshifts (Cohen et al., 2000; Fernández-Soto et al., 2001). Benítez (2000) augmented the calibration sample by the CFRS catalog (Lilly et al., 1995) at bright magnitudes with 591 galaxies with $20 < I < 22.5$ that were spectrally classified using their $V-I$ colors. This ensures the presence of enough galaxies at all magnitude ranges for a meaningful determination of the parameters in Eq. 5.7.

Benítez (2000) used the following parameterization of Eq. 5.7 for the type fraction

$$p(T|m_o) \quad = \quad f_T \exp[-k_T(m_o - 20)] \tag{5.11}$$

and for the prior distributions

$$p(z|m_o, T) \quad \propto \quad z^{\alpha_T} \exp\left[-\left(\frac{z}{z_{mT(m_o)}}\right)^{\alpha_T}\right] \tag{5.12}$$

where $z_{mT}(m_o)$ is the median redshift, and $T = 1$ for ellipticals, $T = 2$ for spirals (Sbc & Scd), and $T = 3$ for the irregulars (Irr, & SB), with the condition $\sum_T p(T|m_o) = 1$. The spectral type fractions f_T at $m_o = 20$ are E 35%, spirals 50%, and Irr 15%. The median $z_{mT}(m_o)$ is chosen to have a simple, linear dependence on magnitude,

$$z_{mT}(m_o) = z_{oT} + k_{mT}(m_o - 20). \tag{5.13}$$

Thus, there are 11 free parameters ($\{\alpha_T, z_{oT}, k_{mT}, k_T\}$) that have been determined from the CFRS+HDF sample. Their values are shown on Table 5.3. This parametric form is justified by numerous galaxy surveys (e.g Maller et al., 2003, and reference therein) which have shown that

$$\frac{\mathrm{d}N}{\mathrm{d}z}(z) \propto z^2 \exp\left[-\left(\frac{z}{z_m}\right)^2\right], \tag{5.14}$$

where $z_m = a + b(m_o - c)$.

Table 5.3. Parameters for the prior probabilities from Benítez (2000).

SED type	α_T	z_{oT}	k_{mT}	k_T
E/S0	2.46 ± 0.22	0.431 ± 0.030	0.091 ± 0.017	0.147 ± 0.013
Sbc, Scd	1.81 ± 0.10	0.390 ± 0.024	0.0636 ± 0.009	0.450 ± 0.036
Irr	0.91 ± 0.05	0.063 ± 0.013	0.123 ± 0.012	

To close this subsection, we note that we used out I-band magnitudes for m_o.

5.5 Photometric redshift summary

Fig. 5.12 compares the photometric redshifts $z_{\rm phot}$ with the spectroscopic redshifts $z_{\rm spec}$ for the galaxies in the HDF. The left panel shows the results of Fernández-Soto et al. (2001), who used the optical and the IR photometry. The middle panels shows the result when using only the four optical bands with no priors. The right panel shows the result when using the four optical bands with the Bayesian approach. One sees that the priors allow to break the degeneracies at $z_{\rm phot} \sim 2$.

To summarize our analysis of photometric redshift in the HDF,

- we used a better description of the extinction due to the Lyα forest,
- the best set of templates were the four CWW templates. We used a better extension of these SED than in Fernández-Soto et al. (2001),
- IR photometry was not necessary, and
- we included the Bayesian approach of Benítez (2000) that uses the galaxy magnitude as prior information.

Next, we applied our photometric redshift technique to the $\sim$ 10,000 objects (per field) presented in section 4.7 and from this catalog, we selected LBGs in redshift slices, as described below.

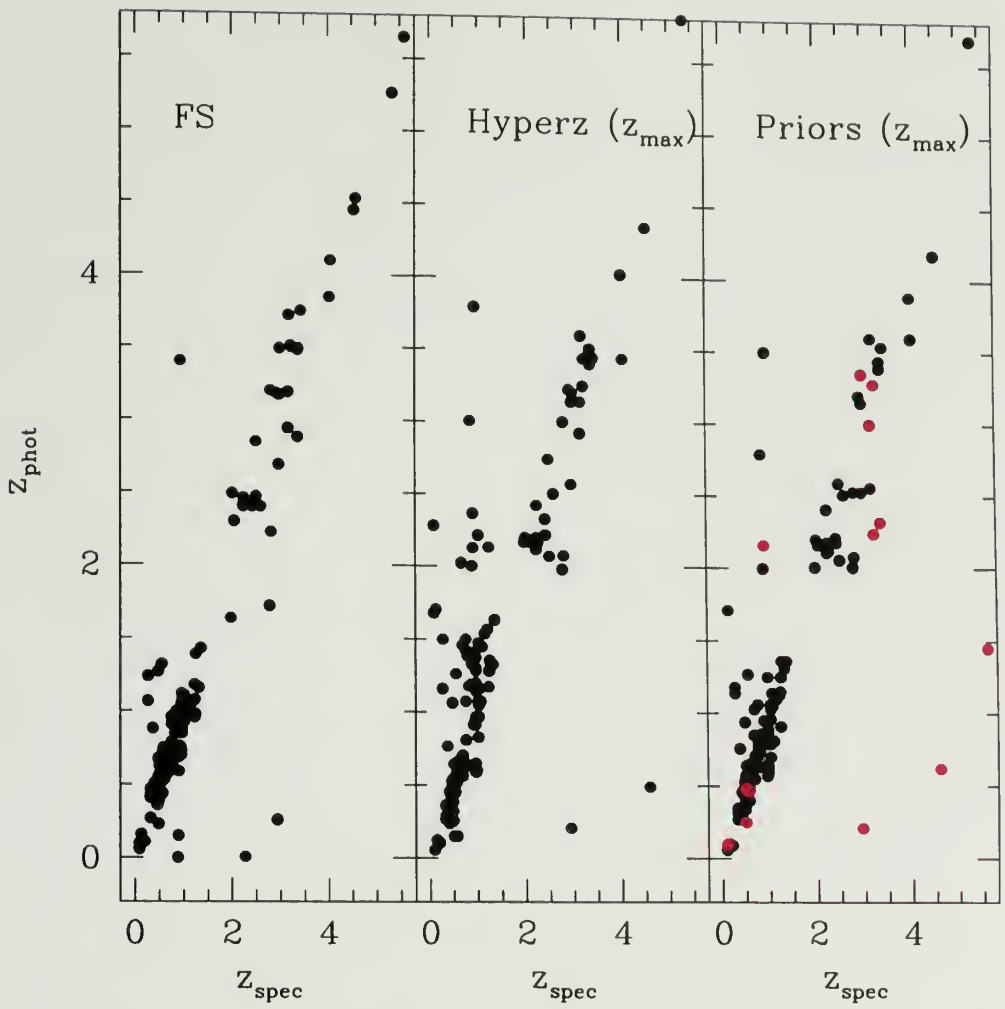

Figure 5.12 Left: reproduced from Fernández-Soto et al. (2001), photometric redshifts $z_{\rm phot}$ as a function of spectroscopic redshifts $z_{\rm spec}$ for the galaxies in the HDF, using the optical and IR photometry. Middle: photometric redshifts $z_{\rm phot}$ as a function of spectroscopic redshifts $z_{\rm spec}$ for the same galaxies using 'Hyperz' with no priors and no IR photometry. Right: photometric redshifts $z_{\rm phot}$ as a function of spectroscopic redshifts $z_{\rm spec}$ our implementation of the Bayesian approach of Benítez (2000). The priors allow to break the degenerancies at $z_{\rm phot} \sim 2$. The red points show the galaxies with a high probability of being an outlier (P(outlier) > 0.87).

5.6 Redshift distributions

Fig. 5.13 shows the redshift distribution of the four fields. The dotted histogram shows the photometric redshift distribution using the template set A with no priors. The continuous histogram shows the photometric redshift distribution using the prior probabilities presented in section 5.4. Fig. 5.13 shows that using the priors has the effect of removing galaxies with $z_{\rm phot} \simeq 2$ that are likely at lower redshifts. As expected, the distribution of galaxies at $z \sim 3$ is not affected much. This is due to the fact that we are sensitive to the shape of the SED, which has a strong break at that redshift.

Since we have the photometric redshift $z_{\rm phot}$ for all of our galaxies, we can use their apparent magnitude m_I and their distances $D(z)$ to determine the absolute magnitude M_I of our objects. For galaxies at $z \sim 3$, this corresponds to computing their luminosity at the rest-wavelength $\lambda \sim 8000$Å from their observed flux at the rest-wavelength $\lambda \sim 2000$Å. The extrapolation is performed by the K-correction, which is presented in Appendix C (see Eq. C.6). At redshift $z \sim 3$, the K-correction is, however, small: $K_{II} \sim 0.2$ as shown in Fig. C.1. Since the best-fitted SED was a combination of the spectral types T (see Eq. 5.10), we computed the K-correction for each type T separately, and took the weighted average, where the weights were given by the factors in Eq. 5.10 evaluated at $z_{\rm phot}$.

Then, we computed the absolute magnitude $M_I(AB)$ for each galaxy using Eq. C.15 and the K-corrections. Fig. 5.14 shows the absolute magnitude M_I as a function of z, known as the 'Hubble diagram'. Each dot represent one galaxy in our fields. The two continuous lines show our magnitude cuts and were computed using an Irr SED. This shows that the galaxies at redshift $z \sim 3$ are, as expected, between the two lines and near our completeness limit. Since galaxies at redshift $z \sim 3$ do not have absolute magnitudes that are outside the continuous lines, this is a consistency check of the photometric redshift technique.

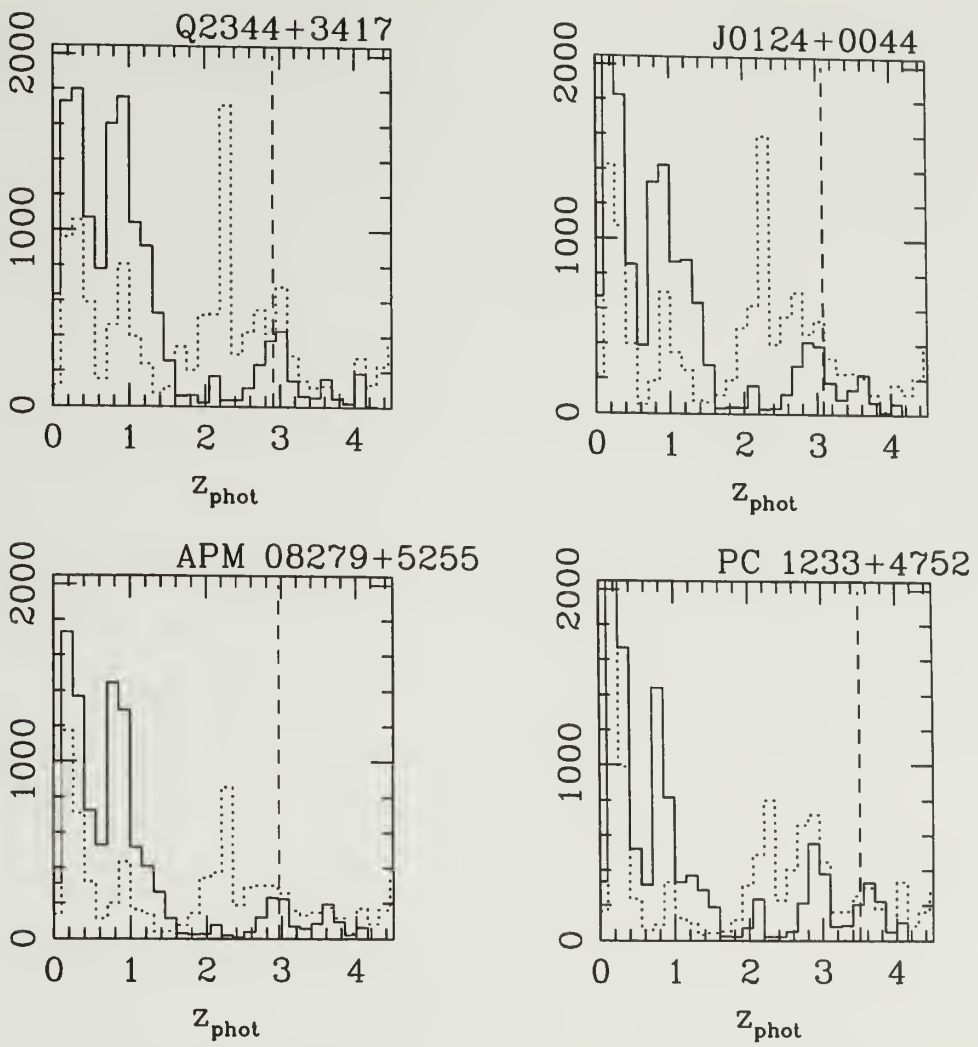

Figure 5.13 Redshift distribution for the four fields. In each panel, the vertical dashed line shows the redshift of the DLA $z_{\rm DLA}$. The dotted histogram shows the photometric redshift distribution using no priors and the template set A. The continuous histogram shows the photometric redshift distribution using the prior probabilities presented in section 5.4. Using the priors has the effect of eliminating the large number of galaxies that have been assigned $z_{\rm phot} \simeq 2$ wrongly, but does not affect the distribution at $z \sim 3$ significantly.

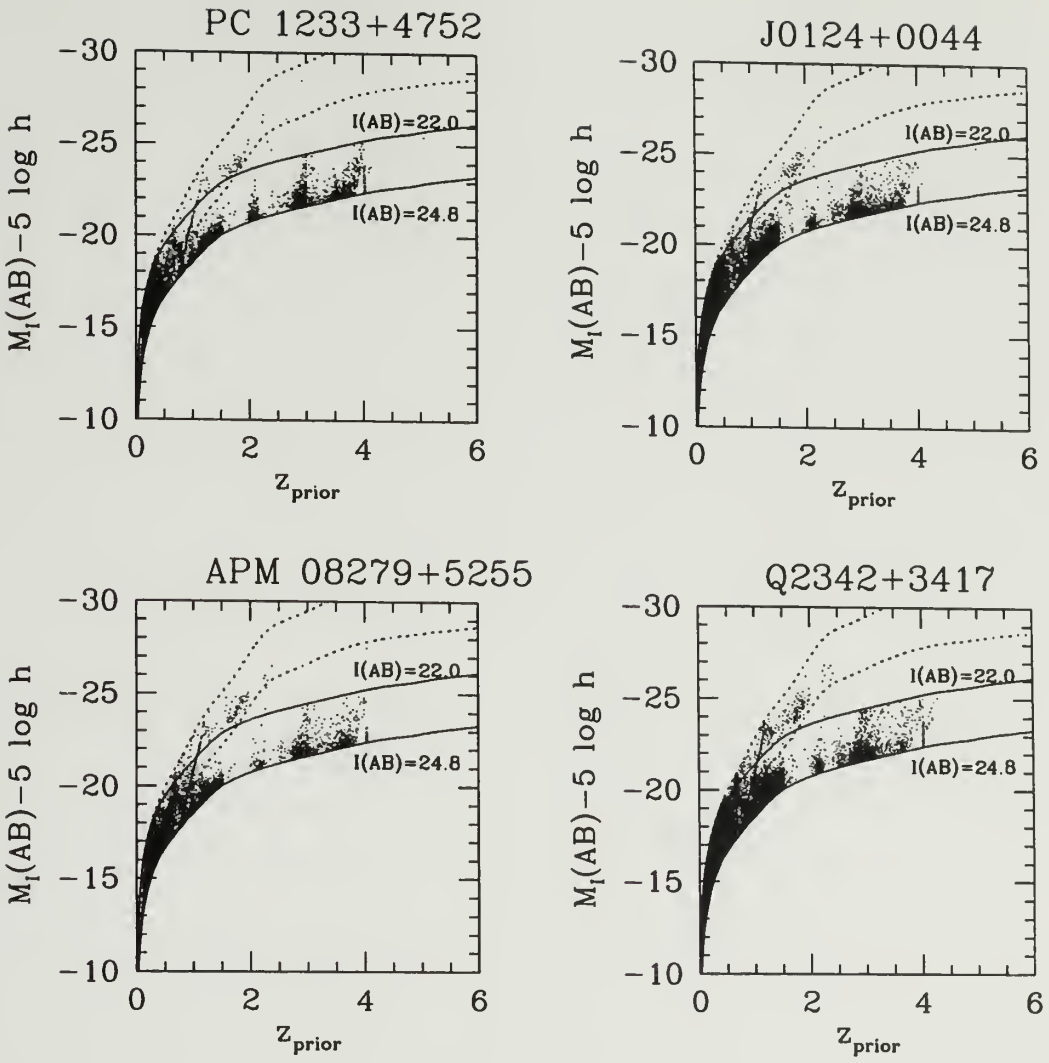

Figure 5.14 The 'Hubble diagram' for the four fields. Each dot represents one galaxy in our fields. The rest-frame absolute magnitude M_I was computed using Eq. C.15 with the K-correction $K_{II}(z)$ from Eq. C.6. The continuous lines show our magnitude selection $22 < I_{\rm AB} < 24.8$ and are computed using an Irr SED. At $z \sim 3$, galaxies are between the two continuous lines and near our completeness limit as expected, which gives us more confidence in the photometric redshifts. The dotted lines show our magnitude selection for an E/S0 SED. Clearly, points that are outside the range allowed by the continuous lines are best-fitted by the E/S0 type, which has a strong break at 4000Å (see Fig. 5.1) and thus a large K-correction.

Points that are outside the range allowed by the continuous lines are between the dotted lines, which represent our magnitude range using a E/S0 SED, and thus are best fitted by the E/S0 type. This SED has a strong break at 4000Å (see Fig. 5.1) that creates a large K-correction of four magnitudes. Since, they are too luminous for their apparent magnitude, the fitted SEDs are likely to be wrong, but their photometric redshift is still reliable. One would need IR photometry to constrain the SED at those wavelengths.

5.6.1 The selection of LBG candidates near the DLAs

Selecting galaxies in slices

From the subsample presented in section 4.7, we selected LBG candidates that have a high probability to be at the redshift of the DLA from their redshift likelihood distributions. For instance, the probability that a galaxy lies in a redshift range $[z_o - W_z/2; z_o + W_z/2]$ is given by:

$$P(z_o \pm W_z/2) = \int_{z_o - W_z/2}^{z_o + W_z/2} P(z)\mathrm{d}z. \tag{5.15}$$

where W_z is the redshift width of the slice. For each galaxy, we computed the probability to be in the following redshift slices:

$$P_+ \equiv P([z_{\mathrm{DLA}} + 0.15] \pm 0.075), \tag{5.16}$$

$$P_{\mathrm{DLA}} \equiv P(z_{\mathrm{DLA}} \pm 0.075), \tag{5.17}$$

$$P_- \equiv P([z_{\mathrm{DLA}} - 0.15] \pm 0.075). \tag{5.18}$$

For the APM 08279+5255 field, the slice centered before the redshift of the DLA (z_{DLA}), P_-, is shown in Fig. 5.15a. The slice centered on the redshift of the DLA (z_{DLA}), P_{DLA}, is shown in Fig. 5.15b. The slice centered after the redshift of the DLA (z_{DLA}), P_+, is shown in Fig. 5.15c. The dots represent galaxies detected in the four

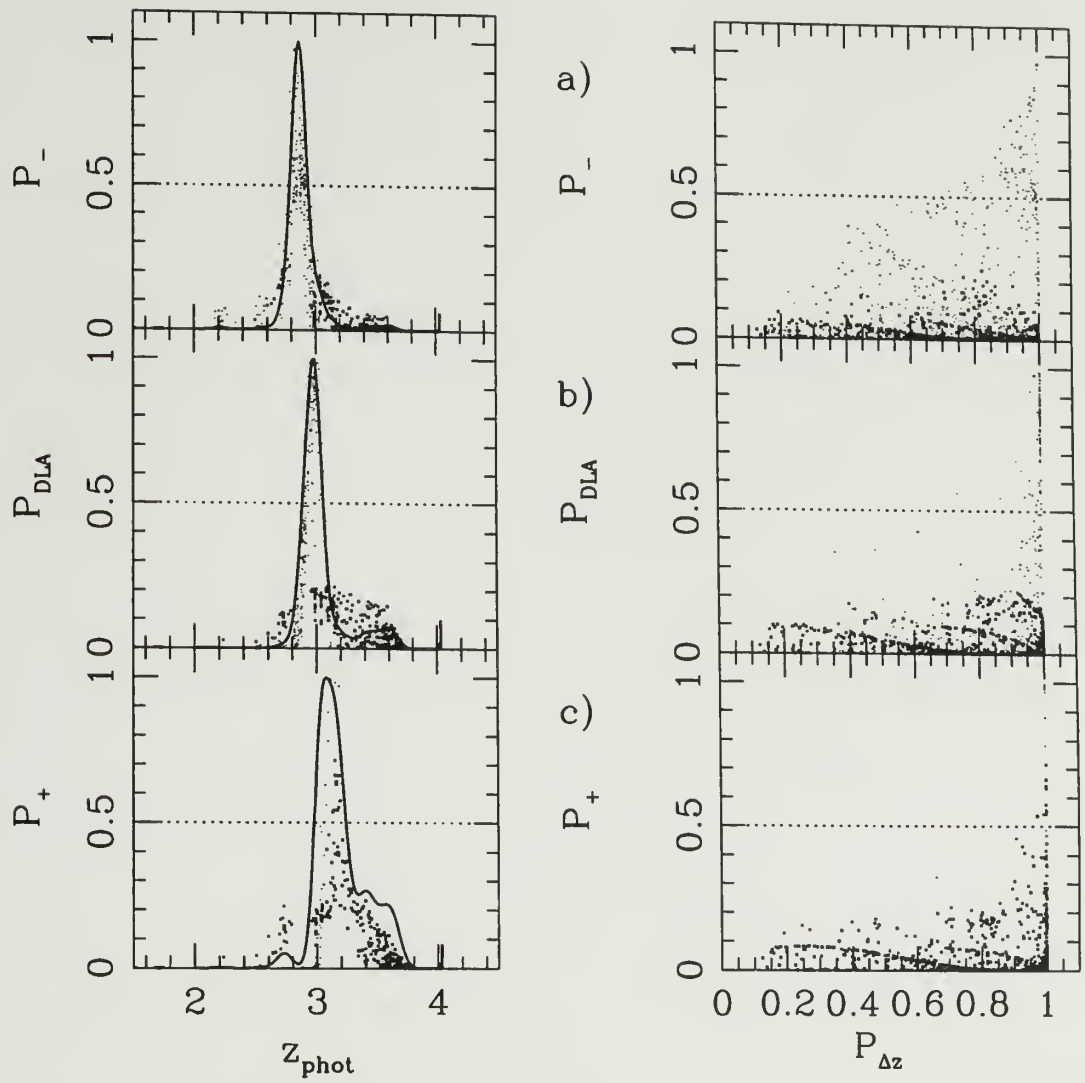

Figure 5.15 For the APM 08279+5255 field, the redshift slices centered before, on, and after the DLA ($z_{\rm DLA}$=2.974) is shown in rows a), b) and c) respectively. Each dot represent a galaxy that was detected in the four UBV& I bands. The filled squares indicate objects that are not detected in the U band. The left column shows the probability distribution as a function of photometric redshift. The continuous line shows the smoothed distribution (arbitrarily scaled to peak). The right column shows the probability to be in that particular slice as a function of the 'goodness' of the photometric redshift $P_{\Delta z}$ defined in Eq. 5.8. The dotted line shows the minimum threshold used in selecting LBG candidates in each of the slices.

Table 5.4. Number of galaxies in the different redshift slice.

Field	P_-	$P_{\rm DLA}$	P_+
APM 08279+5255	89	84	17
PC 1233+4752	20	70	83
J0124+0044	184	65	22
Q2342+3417	65	177	129

UBV& I bands. The filled squares indicate objects that are not detected in the U band. The left column shows the probability distribution. Smoothing the distributions using a Gaussian kernel (e.g. Wand & Jones, 1995) produced the continuous lines in Fig. 5.15. The right column shows the probability distribution as a function of $P_{\Delta z}$, which measures the reliability of the photometric redshift, defined in Eq. 5.8. The dotted line shows the minimum threshold used in selecting LBG candidates in each of the slices. From the right panels, galaxies that have a high probability of being in a redshift slice also have reliable photometric redshifts, indicated by the fact that they have a low probability to be an outlier, i.e. $P_{\Delta z}(= 1 - P(\text{outlier})) \sim 1$.

The number of galaxies, for each redshift slice, that met the threshold of 0.5 is shown in Table 5.4. Fig. 5.16 shows four examples of galaxies at $z_{\rm phot} \simeq 3$. For each row, the four columns correspond to the four imaging bands, U, B, V & I, respectively. Because the Lyman break enters the U band at $z \sim 3$, a galaxy drops out of the U band. This effect is at the root of the photometric redshift technique.

For the remainder of this work, our sample of LBGs that are in the redshift slice centered on the DLA comprises galaxies that meet the probability threshold of 0.5 shown in Fig. 5.15b:

$$P_{\rm DLA} > 0.5,$$

and accordingly for the other slices.

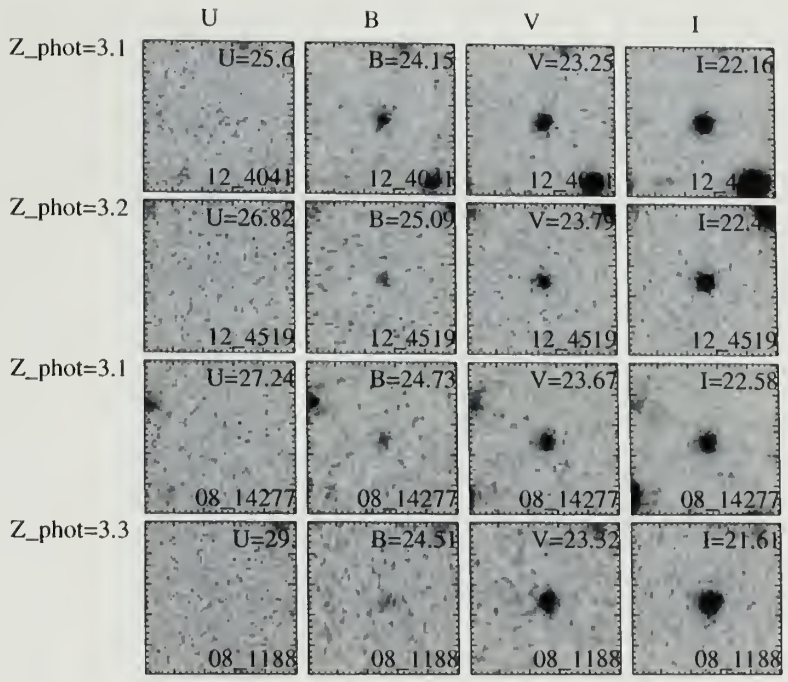

Figure 5.16 Four examples of Lyman break galaxies at $z_{\rm phot} \simeq 3$. Each column corresponds to the four imaging bands, U, B, V & I. At that redshift, the Lyman break is redshifted in the U filter and beyond and the galaxy drops out of the U band.

5.7 On the robustness of photometric redshifts

We close this chapter with a note on the robustness of the photometric redshift technique. In section 4.6, we discussed a possible systematic offset in the zeropoint in our photometric calibration for run I. In order to test the effect of this systematic offset, we ran our photometric redshift code on the APM 08279+5255 field with the magnitudes U and B offset by $U' = U + 0.074$ and $B' = B - 0.044$. The offsets 0.074 and -0.044 represents the possible systematic. This worst-case scenario gave photometric redshifts that, at redshift $z \leq 3$, were slightly offset on average by 0.03, the minimum redshift step. At higher redshifts $z \geq 3$, it did not change the photometric redshifts.

Finally, we note that we have very recently obtained spectra of 80 LBGs in the field towards Q2342+3417 using the multi-object spectrograph GMOS on the Gemini telescope. This will provide us with spectroscopic redshifts of a small number of galaxies. The comparison between spectroscopic and photometric redshift will in the future allow us to refine the Bayesian approach further.

CHAPTER 6

THE LBG/DLA CONNECTION III. CLUSTERING ANALYSIS

This chapter is the last chapter on the LBG/DLA connection. Chapter 4 presented the data. Chapter 5 showed our selection of galaxies based on photometric redshifts. Finally, this chapter presents our clustering analysis since, as motivated in section 1.2.1, our scientific goal is to answer "are Lyman break galaxies correlated with the presence of the DLA?".

Section 6.1 presents preliminary considerations. Section 6.2 lays the foundations of correlation functions and their properties. Section 6.3 shows the results of the DLA-LBG cross-correlation in numerical simulations. Finally, section 6.4 shows the results in our data using the same methods.

6.1 Clustering analysis: motivations

Since massive galaxies are likely to attract others and affect their environment, they are more likely to be found in regions of high density, whereas low-mass galaxies are more uniformly distributed throughout the universe. This produces an enhancement of high-mass galaxies since they tend to form near each other in rare over-dense regions. Therefore, such clustering (predicted by hierarchical models of galaxy formation, see section 6.2.8) is a probe of the mass distribution of galaxies. For instance, if Lyman break galaxies cluster around DLAs, it would indicate that DLAs reside in massive dark halos.

Saslaw (2000) has written one of the most comprehensive discussions on correlation functions. Here, we summarize the key ideas that are relevant to the clustering analysis of this work.

An intuitive question to describe the clustering of a sample of galaxies is to ask: 'if one sits in a galaxy, are the neighbor galaxies closer than average?' The average number density $\overline{n}$ is the basic quantity to address this, so it is the first quantity to find. However, one must specify whether $\overline{n}$ is the unconditional density $\overline{n}_u$ averaged over randomly placed volumes, or whether it is the conditional density $\overline{n}_c$ averaged over volumes known a priori to contain a galaxy at their center. Naturally, if the galaxies are randomly distributed, then they have a Poisson distribution, and $\overline{n}_c = \overline{n}_u$. This means that for Poisson distribution, it does not matter whether the volume is centered on a galaxy or not. In this case, the expected number of galaxies in a small volume dV centered on any galaxy is $N = \overline{n}_u \mathrm{dV}$. On the other hand, if galaxies are clustered, the expected number of galaxies on a small volume dV centered on a galaxy will not be $\overline{n}_u \mathrm{dV}$, but $\overline{n}_c \mathrm{dV}$. The galaxies are now correlated because either, the central galaxy has perturbed its environment, or the environment has made the galaxy more likely to exist.

A statistic widely used to quantify this perturbation of $\overline{n}$ is the correlation function ξ, which we used here. (Other statistics include 'the nearest neighbor distance', e.g. Bogart et al., 1973).

6.2 Correlations: definitions and properties

This section contains useful results on correlation functions. The first four sections (§ 6.2.1-6.2.4) are a general presentation of correlation functions. Sections 6.2.1 and 6.2.2 lay down the basic definitions of the two-point and three-point correlation functions. Section 6.2.3 presents the expectation values of counts in cells and the variance

of those counts. Section 6.2.4 generalizes correlation functions to cross-correlations functions.

Section 6.2.5 presents the estimator of the cross-correlation function that we will use in section 6.3 and 6.4, along with the expected variance associated with them. Another estimator of the cross-correlation function is presented in section 6.2.6. Section 6.2.7 shows the relationship between angular correlation functions and spatial correlation function. Finally, section 6.2.8 shows the intimate relationship between correlation functions and the theory of galaxy formation.

6.2.1 The two-point correlation function

The two-point correlation function, $\xi(r)$, can be defined in several ways. One definition states that if there is a galaxy at the position r_1 centered in the volume dV_1, then the conditional probability of finding another galaxy in a volume dV_2 (centered at r_2) at a distance $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ is

$$P(2|1) = \overline{n}(1 + \xi_{gg}(\mathbf{r}))\mathrm{dV}_2. \tag{6.1}$$

where $\overline{n}$ is the *unconditional* galaxy density. The two-point correlation function $\xi(r)$ describes the influence of the galaxy in dV_1 on the probable presence of a galaxy in dV_2. ξ_{gg} quantifies the excess or deficit correlation compared to a Poisson distribution. If $\xi > 0$, there will be an excess of neighbor galaxies (correlation); if $\xi < 0$ there is a deficit of neighbors (anti-correlation) from a given galaxy.

Another definition involves the joint probability of finding a galaxy in dV_1 and also one in dV_2. This joint probability is the product of $P(1) = \overline{n}_g \mathrm{dV}_1$ and $P(2|1)$:

$$P_{1,2} = \overline{n}^2(1 + \xi_{gg}(\mathbf{r}))\mathrm{dV}_1\mathrm{dV}_2. \tag{6.2}$$

If the distribution of galaxies can be approximated by a continuous density distribution $n(\mathbf{r})$, then each random volume element dV has a probability $\mathrm{d}P = n(\mathbf{r})\mathrm{dV}$ of

containing a galaxy. Thus, the joint probability is $P_{1,2} = n(\mathbf{r}_1)\mathrm{d}V_1 n(\mathbf{r}_2)\mathrm{d}V_2$ and can be written as an average over all the galaxies:

$$P_{1,2} =< n(\mathbf{r}_1)n(\mathbf{r}_1+\mathbf{r}) > \mathrm{d}V_1\mathrm{d}V_2, \tag{6.3}$$

where $\mathbf{r}_1 + \mathbf{r} = \mathbf{r}_2$.

Comparing Eq. 6.2 and 6.3 gives a third definition:

$$\xi(\mathbf{r}) = \frac{< n(\mathbf{r}_1)n(\mathbf{r}_1+\mathbf{r}) >}{\overline{n}^2} - 1 \tag{6.4}$$

which is the auto-correlation of $n(\mathbf{r})$.

A fourth way to define ξ is in terms of the overdensity Δn at each point where $n(\mathbf{r}_1) = \overline{n} + \Delta n(\mathbf{r}_1)$. Since $< \Delta n >= 0$, Eq. 6.4 becomes

$$\xi(\mathbf{r}) \;=\; \frac{< \Delta n(\mathbf{r}_1)\Delta n(\mathbf{r}_1+\mathbf{r}) >}{\overline{n}^2}. \tag{6.5}$$

If the distribution of galaxies is isotropic ξ is just a function of $r = |\mathbf{r}_2 - \mathbf{r}_1|$, $\xi(r)$. The galaxy two-point auto-correlation function is usually expressed as $\xi_{gg}(r) = (r/r_o)^{-\gamma}$, where r_o is the scale length at which the probability of finding a galaxy is doubled compared to a random distribution, and $\xi(r_o) = 1$.

6.2.2 The three-point correlation function

Similarly to Eq. 6.5, one can define the three-point correlation function as

$$\zeta(\mathbf{r}_1, \mathbf{r}_2) = \frac{< \Delta n(\mathbf{r}_o)\Delta n(\mathbf{r}_o+\mathbf{r}_1)\Delta n(\mathbf{r}_o+\mathbf{r}_2) >}{\overline{n}^3}. \tag{6.6}$$

This expression can be expanded using $n(\mathbf{r}_o) = \overline{n} + \Delta n(\mathbf{r}_o)$ and Eq. 6.4 to give the joint probability of finding three galaxies, one at $\mathbf{r}_o$ and two neighbors at r_1 and r_2:

$$P_{0,1,2} \;=\; < n(\mathbf{r}_o)n(\mathbf{r}_o+\mathbf{r}_1)n(\mathbf{r}_o+\mathbf{r}_2) > \mathrm{d}V_o\mathrm{d}V_1\mathrm{d}V_2$$

$$= \overline{n}^3[1 + \xi(r_1) + \xi(r_2) + \xi(|\mathbf{r}_1 - \mathbf{r}_2|) + \zeta]\mathrm{d}V_o\mathrm{d}V_1\mathrm{d}V_2. \quad (6.7)$$

This three-dimensional relation is usually difficult to measure directly as it requires precise distance measurements of a large sample of galaxies. However, one can rewrite ζ as shown by Peebles (1980) as a product of two-point correlations:

$$\begin{aligned} \zeta(\mathbf{r}_1, \mathbf{r}_2) &= Q\,[\xi(r_1)\xi(r_2) + \xi(r_1)\xi(|\mathbf{r}_1 - \mathbf{r}_2|) + \xi(r_2)\xi(|\mathbf{r}_1 - \mathbf{r}_2|)], \\ &= Q\,[\xi_1\xi_2 + \xi_1\xi_{12} + \xi_2\xi_{12}] \end{aligned} \quad (6.8)$$

where $r_1 = |\mathbf{r}_o + \mathbf{r}|$, $r_2 = |\mathbf{r}_o + \mathbf{r}_2|$ and $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$. Q is observed to be $Q \simeq 1.3$ (Groth, 1977).

6.2.3 Properties: Counts in cell & variance

For the interpretation of our results, it is important to understand the sources of the variance inherent in correlation functions. In other words, given the correlation function, what is the average expected number of counts and their variance? In this section, we show the expected number counts and their variance (from Peebles, 1980) in (A) random cells and in (B) cells centered on a galaxy.

A. Randomly placed cells

For randomly placed cells, the expected mean number of objects within a random volume $\mathrm{d}V$ is simply:

$$< N_r >= \overline{n}_g \Delta V = \overline{N}. \quad (6.9)$$

The variance, however, is larger if the galaxies are clustered. From Peebles (1980), section 36, the variance $V(N) =< (\Delta N)^2 >=< N^2 > - < N >^2$ is

$$V(N_r) = \overline{N} + \overline{N}^2 \frac{1}{(\Delta V)^2} \int_{\Delta V} \int_{\Delta V} \xi(|r_2 - r_1|)\mathrm{d}V_1\mathrm{d}V_2 \quad (6.10)$$

which is the sum of the Poisson shot noise, V_{sn}, and the clustering variance, V_{cl} :

$$V(N_r)_{sn} = N_r = \overline{N}, \tag{6.11}$$

$$V(N_r)_{cl} = \overline{N}^2 \frac{1}{(\Delta V)^2} \int_{\Delta V} \int_{\Delta V} \xi_{12}\, dV_1 dV_2. \tag{6.12}$$

In the case of no clustering, $\xi = 0$. Eq. 6.10 reduces to the Poisson variance $V(N) = \overline{N}$.

B. Counts of neighbors

When sitting in a galaxy at $\mathbf{r}_o = 0$, the expected mean number of neighbors at a distance $r \pm dr$ is given by Eq. 6.1 over the shell ΔV:

$$\begin{aligned} < N_p > &= \overline{n}_g \Delta V \frac{1}{\Delta V} \int_{\Delta V} (1 + \xi(r)) dV \\ &= \overline{N} + \overline{N}\, \overline{\xi}(r). \end{aligned} \tag{6.13}$$

where n_g is the unconditional galaxy density. As expected, the mean expected number of neighbors N_p is increased by the clustering ξ. N_p is $N_r + N_c$, where N_c gives the contribution of the correlation and N_r is what would be expected for a Poisson distribution.

From the second moment $< N_p^2 >$, given in (Peebles, 1980), section 36, the variance $V(N_p) =< (\Delta N)^2 >=< N_p^2 > - < N_p >^2$ is

$$V(N_p) = \overline{N}_p + \overline{n}_g^2 \int \int [\xi_{12} - \xi_1 \xi_2 + \zeta] dV_1 dV_2, \tag{6.14}$$

which, as before, can be separated into the Poisson shot noise V_{sn}, and the clustering variance, V_{cl} :

$$V(N_p)_{sn} = \overline{N}_p = \overline{N} + \overline{N}\, \overline{\xi}(r), \tag{6.15}$$

$$V(N_p)_{\rm cl} = \overline{N}^2 \frac{1}{(\Delta V)^2} \int_{\Delta V} \int_{\Delta V} [\xi_{12} - \xi_1 \xi_2 + \zeta] {\rm d}V_1 {\rm d}V_2. \quad (6.16)$$

The clustering variance $V_{\rm cl}$ is also a sum of two terms, $V_{\rm cl} = V_{\rm 2pt} + V_{\rm 3pt}$, due to the two-point correlation and the three-point correlation, respectively:

$$V_{\rm 2pt} = \overline{N}^2 \frac{1}{(\Delta V)^2} \int_{\Delta V} \int_{\Delta V} \xi_{gg}(|\mathbf{r}_2 - \mathbf{r}_1|) {\rm d}V_1 {\rm d}V_2 \quad (6.17)$$

$$V_{\rm 3pt} = \overline{N}^2 \frac{1}{(\Delta V)^2} \int_{\Delta V} \int_{\Delta V} [\zeta_{ggg}(\mathbf{r}_1, \mathbf{r}_2) - \xi_{gg}(r_1)\xi_{gg}(r_2)] {\rm d}V_1 {\rm d}V_2. \quad (6.18)$$

In general, these equations require numerical integration. Peebles (1980) shows analytical solutions in the case of a spherical cell. Eq. 6.18 requires the use of Eq. 6.8 for the three-point correlation function ζ.

6.2.4 A simple estimator of the cross-correlation ξ_{dg}

In the case of two populations (in this case DLAs and Lyman break galaxies), one can define the cross-correlation similarly to the galaxy-galaxy auto-correlation (Eq. 6.1). One can define the cross-correlation ξ_{dg} between DLAs and LBGs from the conditional probability of finding a galaxy in a volume ${\rm d}V_2$ at a distance $r = |\mathbf{r}_2 - \mathbf{r}_1|$, given that there is a DLA centered on the volume ${\rm d}V_1$:

$$P(LBG|DLA) = \overline{n}_g (1 + \xi_{dg}(r)) {\rm d}V_2. \quad (6.19)$$

Therefore, if LBGs and DLA are correlated ($\xi_{dg} \neq 0$), the mean number of galaxies in a cell of volume ΔV is given by the counts of neighbors found in Eq. 6.13:

$$N_p = \overline{N}(1 + \overline{\xi}_{dg}(r)). \quad (6.20)$$

The variance in the counts of neighbors (Eq. 6.14) can be extended to cross-correlations. The variance of the number of neighbor galaxies near a DLA is the sum

of the shot noise V_{sn} and the clustering variance $V(N_p)_{cl} = V_{2pt} + V_{3pt}$:

$$V(N_p)_{sn} = N_p \tag{6.21}$$

$$V(N_p)_{2pt} = \overline{N}^2 \frac{1}{(\Delta V)^2} \int_{\Delta V} \int_{\Delta V} \xi_{gg}(|r_2 - r_1|) \tag{6.22}$$

$$V(N_p)_{3pt} = \overline{N}^2 \frac{1}{(\Delta V)^2} \int_{\Delta V} \int_{\Delta V} [\zeta_{dgg}(r_1, r_2) - \xi_{dg}(r_1)\xi_{dg}(r_2)] dV_1 dV_2. \tag{6.23}$$

These equations are at the center of the following discussions in assessing the variance in galaxy counts in cells around a DLA. Note that the variance is augmented not only by the cross-correlation between the galaxies and the DLA, but also by the auto-correlation of the galaxies ξ_{gg}, such that in the case when there is no cross-correlation $\xi_{dg} = 0$ the galaxy-galaxy auto-correlation ξ_{gg} can increase the variance significantly. In that case, the two-point variance V_{2pt} dominates the clustering variance $\overline{\Delta N^2}_{cl}$ since the three-point variance $\overline{\Delta N^2}_{3pt}$ is zero because $\xi_{dg} = \zeta_{dgg} = 0$.

6.2.5 A robust estimator of the cross-correlation ξ_{dg}

An estimator of the cross-correlation ξ_{dg} that does not depend on $\overline{N} = n_g dV$ can be constructed in the following way. It is given by the ratio of the number of neighbor galaxies observed in a cell of volume $dV(r)$ at distance r from a DLA, which is given by Eq. 6.13:

$$N_{\rm obs} = \overline{N} + \overline{N}\,\overline{\xi}(r),$$

and of the number of neighbor galaxies in the same cell $dV(r)$ expected for a random distribution of galaxies, which is given by Eq. 6.9

$$N_{\rm rand} = \overline{N}.$$

The ratio of these equations gives :

$$1 + \xi_{dg}(r) = \frac{N_{\rm obs}(r)}{N_{\rm rand}(r)}.$$

This is for one DLA, and thus the average over the number of DLAs ($N_{\rm DLA}$) gives the estimator of $\xi_{dg}(r)$:

$$1+\bar{\xi}_{dg}(r) \;=\; \left\langle \frac{N_{\rm obs}(r)}{N_{\rm rand}(r)} \right\rangle, \tag{6.24}$$

where the brackets <> indicate the average over $N_{\rm DLA}$. In the case of the auto-correlation ξ_{gg}, the average is performed over the galaxies. In practice, we prefer to average over the ratio, i.e. $\left\langle \frac{N_{\rm obs}(r)}{N_{\rm rand}(r)} \right\rangle$ rather than $\frac{<N_{\rm obs}(r)>}{<N_{\rm rand}(r)>}$ because the latter method increases the variance of ξ. Indeed, in our galaxy sample, certain regions are masked and not all of the volume elements $dV(r)$ are of equal size. As a consequence, the variance of $< N_{\rm obs}(r) >$, and therefore the variance of ξ will be much higher than it actually is. Naturally, when there are no masks, i.e. all the volumes $dV(r)$ have the same sizes, the two methods are equivalent.

This definition (Eq. 6.24) reflects very well the fact that ξ is a measure of the excess of neighbors compared to the number of neighbors expected in a random distribution of galaxies. $N_{\rm rand}$ can be estimated from generating random catalogs of galaxies. In practice, the random catalog is much larger than the galaxy catalog and is scaled to the galaxy sample, such that the variance of $N_{\rm rand}$ is not dominated by shot noise. Effectively, this method measures the volume fraction of ΔV to the total volume and is equivalent to using Eq. 6.20.

The variance $V(\xi)$ of the estimator of ξ can be computed analytically. From Landy & Szalay (1993), it is

$$\begin{aligned} V(\xi) \;&=\; V\left(\left\langle \frac{N_{\rm obs}}{N_{\rm rand}} \right\rangle\right) \\ &\simeq\; \frac{V(<N_{\rm obs}>)}{<N_{\rm rand}>^2} + \frac{V(<N_{\rm rand}>)<N_{\rm obs}>^2}{<N_{\rm rand}>^4} \\ &\simeq\; \left[\frac{V(<N_{\rm obs}>)}{<N_{\rm obs}>^2} + \frac{V(<N_{\rm rand}>)}{<N_{\rm rand}>^2}\right](1+\bar{\xi}_{dg})^2. \end{aligned} \tag{6.25}$$

The shot noise of the random sample in Eq. 6.25, $V(< N_{\rm rand} >)/ < N_{\rm rand} >^2$, can be neglected because the random sample of galaxies is intentionally much larger than the sample of observed galaxies. Thus, the rms (1σ) of ξ is

$$\sigma_\xi \simeq \frac{\sigma_{<N_{\rm obs}>}}{< N_{\rm obs} >}[1 + \overline{\xi}_{dg}] \tag{6.26}$$

where the variance to the mean of $N_{\rm obs}$ is given by Eqs. 6.21-6.23. From Eq. 6.26, one sees that the correlation itself increases the variance by a factor $(1 + \xi)$.

Several approximations to Eq. 6.26 can be made to see the various contributions at play. If we neglect the clustering variance of $N_{\rm obs}$ and approximate $\sigma_{<N_{\rm obs}>}$ by the Poisson variance, that is, $\sigma_{<N_{\rm obs}>} = \sqrt{\frac{1}{N_{\rm DLA}} < N_{\rm obs} >}$ since the average $<>$ is performed over the number of DLAs ($N_{\rm DLA}$), then Eq. 6.26 becomes:

$$\sigma_\xi \simeq \frac{1}{\sqrt{N_{\rm DLA}}} \frac{1}{\sqrt{< N_{\rm obs} >}}[1 + \overline{\xi}_{dg}]. \tag{6.27}$$

which gives the Poisson errors to the cross-correlation function.

If, however, we approximate the clustering variance of $N_{\rm obs}$ (Eqs. 6.22–6.23) by $V_{cl} = A\overline{\xi}_{gg} + B\overline{\xi}^2_{dg}$, where A and B are constants (see Peebles, 1980, section 59), then $\sigma_{<N_{\rm obs}>}$ becomes

$$\begin{aligned} \sigma_{<N_{\rm obs}>} &= \frac{1}{\sqrt{N_{\rm DLA}}} \sqrt{V(N_p)_{sn} + V(N_p)_{cl}} \\ &= \frac{1}{\sqrt{N_{\rm DLA}}} \sqrt{< N_{\rm obs} > + A\overline{\xi}_{gg} + B\overline{\xi}^2_{dg}}, \end{aligned} \tag{6.28}$$

and therefore, the expected rms of the cross-correlation function σ_ξ, Eq. 6.26, becomes:

$$\sigma_\xi = \frac{1}{\sqrt{N_{\rm DLA}}} \frac{1}{\sqrt{< N_{\rm obs} >}} \cdot \sqrt{1 + \frac{A\overline{\xi}_{gg} + B\overline{\xi}^2_{dg}}{< N_{\rm obs} >}} \cdot (1 + \overline{\xi}_{dg}). \tag{6.29}$$

This expression is proportional to

$$\sigma_\xi \propto \frac{1}{\sqrt{N_{\rm DLA}}} \frac{1}{\sqrt{N_g}} \tag{6.30}$$

as one might have expected, since $< N_{\rm obs} >$ is proportional to the number of galaxies, N_g. Thus, the noise in $< \xi >$ goes as the inverse of the square root of the number of DLAs, $N_{\rm DLA}$, and as the inverse of the square root of the number of galaxies N_g. Landy & Szalay (1993) and Mo, Jing, & Boerner (1992) discussed extensively the error of auto-correlation functions. In the case of auto-correlation functions, ω_{gg}, Landy & Szalay (1993) proposed an estimator that has purely Poissonian variance, that is, it reduces to Eq. 6.27. As noted in section 6.2.4, when there is no cross-correlation between DLAs and LBGs ($\xi_{dg} = 0$), the galaxy auto-correlation ξ_{gg} increases the variance of the observed number of neighbors $V(< N_{\rm obs} >)$ and, hence, of σ_ξ.

6.2.6 An estimator of the volume average of the cross-correlation

Eisenstein (2002) presented another useful estimator of the spatial cross-correlation ξ_{dg} between a galaxy sample and a spectroscopic object using weights from a window function. In our case, we want to measure the cross-correlation of a DLA with known redshift (spectroscopic object) and our galaxy sample. The estimator is:

$$\overline{\xi}_W = \frac{1}{V} \int_0^\infty 4\pi r^2 {\rm d}r \xi_{dg}(r) W(r), \tag{6.31}$$

where $V = \int 4\pi r^2 {\rm d}r W(r)$ and $W(r)$ is the window function.

The result is a simple sum over the galaxy sample

$$\overline{\xi}_W = \frac{1}{\overline{n} V} \sum_{k \in LBGs} G(R_k), \tag{6.32}$$

where R_k is the angular distance of the kth galaxy in the LBG sample from the DLA. In the case of a Gaussian window $W(r) = \exp(-r^2/2a^2)$, with a volume V is $(2\pi)^{3/2}a^3$, $G(R)$ is (Eisenstein, 2002):

$$G(R) = e^{-s}[I_o(s) - 2sI_o(s) + 2sI_1(s)], \tag{6.33}$$

where $s = R^2/2a^2$ and I_n are the modified Bessel functions of the first kind.

The variance to this estimator ($\overline{\xi}_W$) has been worked out by Eisenstein (2002), and includes the usual shot noise V_{sn} and clustering variance V_{cl}.

6.2.7 Angular correlation functions

The cross-correlation $\xi_{dg}(r)$ is a 3-dimensional quantity. It requires the knowledge of the galaxies in the three dimensions x, y, z. In chapter 5, we selected galaxies in a slice of depth $W_z = 0.15$ centered on the redshift of the DLA. Unfortunately, our photometric redshift technique is not accurate enough to pinpoint the exact distance to each galaxy. Indeed, the typical uncertainty of our photometric redshifts is $\sigma_z = 0.15$, which at redshift $z = 3$ corresponds to $100h^{-1}$ Mpc. This is much larger than the width of our images, which are $40h^{-1}$ Mpc wide. This means that we are sensitive to small angular scales $0.05 - 40h^{-1}$ Mpc, but not sensitive to scales less than $100h^{-1}$ Mpc along the line of sight. Therefore, we must use angular correlation functions, i.e. projected on the sky.

All the derivations of the expected number counts and their variance for the spatial correlation and cross-correlation functions in sections 6.2.3–6.2.5 can be extended to angular correlations $\omega(R_\theta)$ using the surface density Σ_g instead of n_g, and defining the probability of finding a galaxy in an annulus of radius $r_\theta, r_\theta + \mathrm{d}r_\theta$ as

$$P(LBG|DLA) = \overline{\Sigma}(1 + \omega_{dg}(r_\theta))\mathrm{d}\Omega, \tag{6.34}$$

where $d\Omega$ is the solid angle of the annulus.

All estimators are biased

As noted in section 6.2.1, the two-point correlation function measures an excess number of neighbors compared to the 'unconditional' galaxy density n_u. In practice, this is not a measurable quantity since the entire sample is generally correlated, and therefore, the observed galaxy density is used to estimate n_u and the correlation function. Because the unconditional galaxy density n_u is always lower than the observed galaxy density, all estimates of ω are biased low. Indeed, if n_u is approximated by $\overline{n} = \frac{N_g}{V}$, then the sum over all the cross-pairs, which is

$$\int_V \overline{n}(1 + \xi_{dg}) \mathrm{d}V,$$

must be equal to the total number of galaxies N_g. This implies that

$$\int_V \overline{n}\, \xi_{dg} \mathrm{d}V = 0 \tag{6.35}$$

since $\int_V \overline{n} \mathrm{d}V = N_g$. This will bias ξ or ω, and on large scales, the cross-correlation will be negative to satisfy Eq. 6.35. The bias is often referred to as the 'integral constraint'.

If one assumes a model of the angular correlation function $\hat{\omega}$, the bias $\Delta\omega$ can be computed analytically (Groth, 1977; Landy & Szalay, 1993; Saslaw, 2000). For the auto-correlation, the bias is $\Delta\omega = \frac{1}{\Omega^2} \int \mathrm{d}\Omega_1 \int \mathrm{d}\Omega_2\, \hat{\omega}_{gg}(|r_2 - r_1|)$ where Ω is the survey area. On scales of 0.5 deg, for instance, ω_{gg} is still appreciable, which leads to a 5% overestimate of the unconditional galaxy density and a 10% equivalent underestimate of the clustering. For the cross-correlation function, the bias $\Delta\omega$ follows from the normalization Eq. 6.35, and is (Smith, Boyle, & Maddox , 1995):

$$\Delta\omega = \frac{1}{\Omega} \int \mathrm{d}\Omega\, \hat{\omega}_{dg}(r_\theta). \tag{6.36}$$

Relation to spatial correlation functions

Angular correlation functions ω and spatial correlation functions ξ are related to each other, as one might have expected. If we express $\omega(r_\theta)$ as a function of physical scale r_θ and define our coordinates $\mathbf{r}$ as $(\mathbf{R}_\theta, Z)$, where $\mathbf{r}_\theta$ is the 2-D position on the sky and Z is the coordinate (in units of Mpc) along the line of sight centered on some redshift[1], then the angular correlation function $\omega(R_\theta)$ is given by:

$$\omega(r_\theta) = \int_{-\infty}^{\infty} \mathrm{d}Z\ \phi(Z)\xi\left(\sqrt{(Z^2 + \mathbf{r}_\theta^2)}\right), \tag{6.37}$$

where $\phi(Z)$ is the redshift selection function per unit Z, normalized to one.

If the selection function $\phi(Z)$ is constant over the scale on which ξ is not negligible (as in our case), i.e. $\phi(Z) \simeq \phi(0) \equiv 1/W_z$ within $[-W_z/2, W_z/2]$ and zero otherwise, then Eq. 6.37 can be approximated as

$$\begin{aligned} \omega(r_\theta) &\simeq \phi(0) \int_{-W_z/2}^{W_z/2} \mathrm{d}Z\ \xi\left(\sqrt{(Z^2 + \mathbf{r}_\theta^2)}\right) && (6.38) \\ &\simeq \frac{2}{W_z} \int_{0}^{W_z/2} \mathrm{d}Z\ \xi\left(\sqrt{(Z^2 + \mathbf{r}_\theta^2)}\right). && (6.39) \end{aligned}$$

This integral can easily be performed numerically. If $\xi = (r/r_o)^{-\gamma}$, Eq. 6.39 can be computed analytically. The result is (Adelberger et al., 2003):

$$\begin{aligned} \omega(r_\theta) &\simeq \frac{1}{W_z} r_o^\gamma r_\theta^{1-\gamma}\ H_\gamma\ I_x(1/2, (\gamma-1)/2) \\ &\simeq \frac{1}{W_z} r_o \left(\frac{r_\theta}{r_o}\right)^{1-\gamma}\ H_\gamma\ I_x(1/2, (\gamma-1)/2) && (6.40) \\ &= A\left(\frac{r_\theta}{r_o}\right)^{\beta} \end{aligned}$$

[1] In our case, it will be the redshift of the DLA z_{abs}.

where $\beta = 1 - \gamma$ and $H_\gamma = \frac{\Gamma(1/2)\Gamma((\gamma-1)/2)}{\Gamma(\gamma/2)} \simeq 4.8$ for $\gamma = 1.55$. $I_x(1/2, (\gamma-1)/2)$ is the incomplete beta function following the notation of Press et al. (1992), where $x = \frac{W_z^2}{W_z^2 + 4 \cdot r_\theta^2}$.

Going from ω to ξ is in general a more complicated matter. It involves the inversion of Eq. 6.39 and is the subject of an extensive literature (e.g. Groth, 1977). For instance, using an Abel integral (e.g. von Zeipel, 1908), the inversion is

$$\xi(r) = -\frac{1}{\phi(0)\pi} \int_r^\infty \frac{\mathrm{d}R}{\sqrt{R^2 - r^2}} \frac{\mathrm{d}\omega(R)}{\mathrm{d}R} \tag{6.41}$$

which is noisy since it involves a derivative of $\omega(r)$.

6.2.8 Clustering in hierarchical models

In the final subsection of 6.2, we present how correlation functions relate to galaxy formation models and structure formation.

In the very early universe, small gaussian quantum perturbations on very small scales were stretched out to cosmological scales by inflation. If we describe the universe as an expanding fluid, the amplitude of these small fluctuations $\frac{\delta\rho}{\rho}(r)$ can be worked out (e.g. Longair, 1998) in the various epochs between the Big Bang and today. It turns out that, in a matter dominated universe, the fluctuations grow linearly with the expansion

$$\delta\rho/\rho \propto R^\alpha, \tag{6.42}$$

with $\alpha = 1, 2$, and R is the scaling factor of universe. As a side note, if the universe were not expanding, the fluctuations would grow exponentially with time. This rate would be too rapid to form the observed structure in the universe (stars, galaxies, and galaxy clusters).

Following many textbooks, such as Longair (1998), the relation to correlation functions is as follows. Recall that one definition of the auto-correlation function is given by Eq. 6.5,

$$\xi(r) =< \frac{\delta\rho}{\rho}(\mathbf{x})\frac{\delta\rho}{\rho}(\mathbf{x}+\mathbf{r}) >\equiv \left\langle \left(\frac{\delta\rho}{\rho}\right)^2 \right\rangle,$$

where the brackets indicate the average over a volume V. Then, using the Fourier transform of $\delta\rho/\rho$:

$$\frac{\delta\rho}{\rho}(\mathbf{r}) = \frac{V}{(2\pi)^3}\int \frac{\delta\rho}{\rho}(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{r}}\mathrm{d}^3x, \tag{6.43}$$

the auto-correlation function becomes

$$\xi(r) = \frac{V}{(2\pi)^3}\int |\frac{\delta\rho}{\rho}(k)|^2 e^{-i\mathbf{k}\cdot\mathbf{r}}\mathrm{d}^3k, \tag{6.44}$$

$$= \frac{V}{(2\pi)^3}\int P_E(k)e^{-i\mathbf{k}\cdot\mathbf{r}}\mathrm{d}^3k \tag{6.45}$$

where $P_E(k)$ is the power spectrum of the fluctuations. In other words, the auto-correlation is the fourier transform of the power spectrum P_E.

Indeed, large galaxy surveys, such as the Two Micron All Sky Survey (2MASS, Jarrett et al., 2000), are very useful because they provide us with a good measurement of the angular correlation function ω_{gg}. Eq. 6.45 can then be inverted to give the power spectrum $P_E(k)$ accurately (e.g. Maller et al., 2003, and reference therein).

It is important to realize that the power spectrum $P_E(k)$ in Eq. 6.45 is the power spectrum evolved from the initial power spectrum of perturbations. Indeed, inflation predicts that the power spectrum of initial perturbations be nearly scale invariant, i.e. $P(k) \propto k^n$ with $n \simeq 1$. This initial power spectrum gives the amplitude of the initial fluctuations on a given scale $\lambda \propto 1/k$. As the universe expands, it goes from being radiation dominated to matter dominated. The shape of the power spectrum evolves, and the fluctuations continue to grow as in Eq. 6.42. These fluctuations eventually form galaxies, and galaxy clusters until the overdensity reaches $\delta\rho/\rho \sim 1$. At this point, the fluctuations become non-linear, i.e. they do not follow Eq. 6.42

anymore. After they become non-linear, they continue to grow until the structure becomes stable and virializes which occurs at around $\delta\rho/\rho \sim 200$.

The only piece of information that needs to be specified in order to predict the correlation function ξ from Eq. 6.45 is the evolved power spectrum $P_E(k)$ at a given epoch. As mentioned, the initial power spectrum $P(k)$ evolves first linearly, and then non-linear effects alter its shape.

The linear growth rate α in Eq. 6.42 changes with time depending on whether the universe is matter- or radiation-dominated and on whether the fluctuations are larger or smaller than the size of the horizon of the universe. Shortly after the Big Bang, the fluctuations are larger than the horizon. But, as the universe expands, the size of the horizon catches up with the size of the fluctuations. After that, the fluctuations do not grow much; they grow only logarithmically until the universe is matter dominated. All these effects are combined into what is often called the 'transfer function' $T(k)$. The initial power spectrum $P(k)$ can then be evolved to give the linear power spectrum $P_l(k)$:

$$P_l(k) \propto kT^2(k). \tag{6.46}$$

The transfer function $T(k)$ is usually parameterized with one parameter Γ, (e.g. Mo, Jing, & Boerner, 1997, and references therein). The effect of T is mainly to suppress the amplitude of fluctuations on small scales (large k) by four powers of k. $P_l(k)$ goes as $\sim k^{-3}$.

When the perturbation becomes non-linear, i.e. $\delta\rho/\rho \sim 1$, the fluctuations do not grow as in Eq. 6.42 anymore. Several authors have worked out a fitting formula that relates the evolved spectrum $P_E(k_E)$ to the initial power spectrum (e.g. Hamilton, 1991; Padmanabhan, 1996; Mo, Jing, & Boerner, 1997) where $P_E(k_E) = f[P_l(k)]$.

Fig. 6.1 shows the auto-correlation function and the effect of the non-linear regime on small scales (from Mo, Jing, & Boerner, 1997).

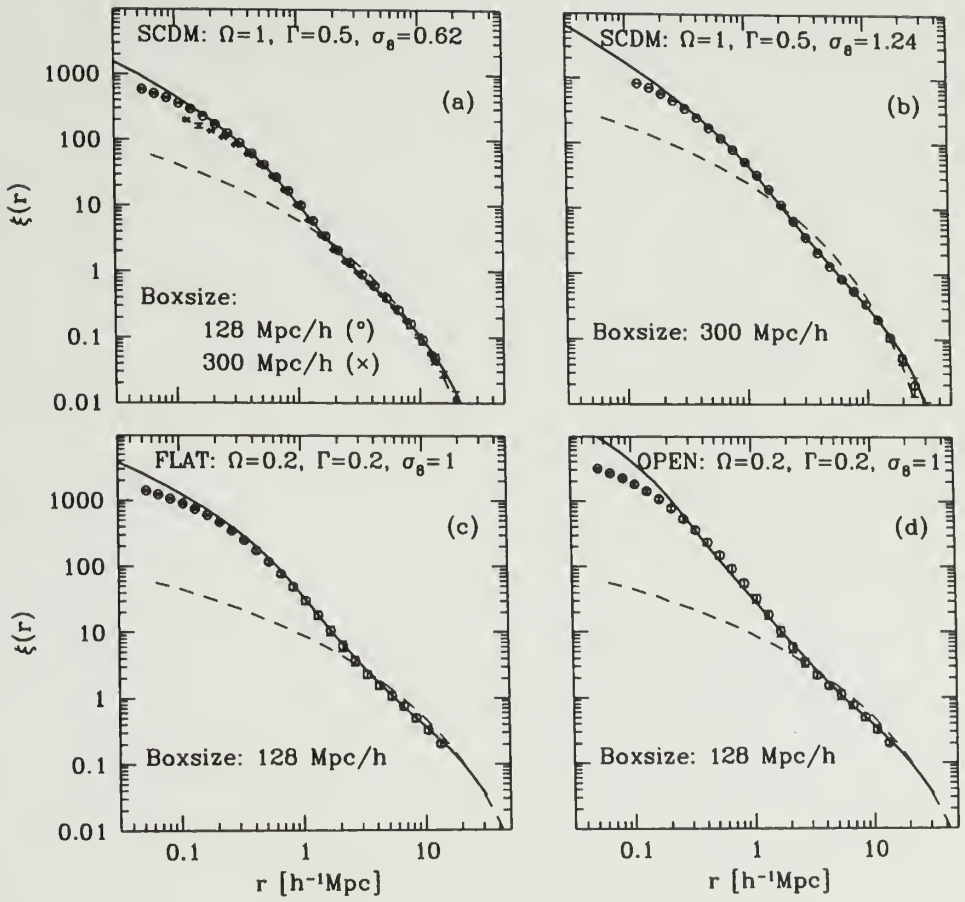

Figure 6.1 The two-point correlation function predicted by analytical models (solid curve) compared to the results derived from N-body simulations (symbols). The dashed curve shows the correlation function given by the linear power spectra P_l in different cosmologies. The model parameters and the simulation box sizes are indicated in each panel (reproduced from Mo, Jing, & Boerner, 1997).

So far, the correlation function ξ refers to the auto-correlation of the smoothed matter density $\rho(r)$. In the universe, however, galaxies are discrete entities. As a consequence, Kaiser (1984) argued that galaxies are likely to form in the highest density peaks of the density distribution. Thus, if we require the density perturbation to exceed some value $\rho_{\rm crit}$ in order to form a galaxy, galaxy formation will be biased towards the highest density perturbations over the mean background density (Longair, 1998; Peacock, 1999). This has been worked out by Peacock & Heavens (1985). The result, in a cold dark matter dominated universe, is:

$$\xi_{\rm gg}(r) = b^2 \xi_{\rm DM}(r) \tag{6.47}$$

where b is the bias factor and $\xi_{\rm DM}$ is the auto-correlation of the underlying dark matter distribution. As pointed out by Peacock (1999), it is important to realize that this bias occurs even if galaxy light traces mass exactly. The bias is due to the fact that the perturbations have a Gaussian distribution of amplitudes with variance $\overline{\delta}^2 \equiv (\delta\rho/\rho)^2$. Thus, the probability of finding an overdensity $\delta\rho/\rho \equiv \delta$ at some point in space is proportional to $\exp(-\delta^2/2\overline{\delta}^2)$ and the clustering is biased.

The exact behavior of this bias depends also on the detailed physics of galaxy formation, and is difficult to model, but is within the reach of current numerical simulations (e.g. Katz, Hernquist, & Weinberg, 1999; Weinberg et al., 2003). Qualitatively, the bias b is a linear function of the mass enclosed in the fluctuation on a particular scale (Efstathiou et al., 1988). Using a different approach, Mo & White (1996) give an analytical expression for the bias a function of mass, $b(M)$. Thus, the amplitude r_o of the correlation function $\xi = \left(\frac{r}{r_o}\right)^{-\gamma}$ is related to the mass of the population. To illustrate this point, clusters have r_o on the order of $\sim 20h^{-1}$ Mpc, local galaxies have $r_o \sim 4 - 6h^{-1}$ Mpc.

The bottom line of this section is that correlation functions do not give a straightforward answer to the mass of the structure involved. It is highly model dependent,

and thus requires a comparison to models of galaxy formation. This is the purpose of the following section.

6.3 The LBG/DLA connection. Results from the simulated universe

Here, we present the results of our analysis of the cross-correlation between DLAs and LBGs in cosmological simulations. § 6.3.1 presents the simulations and also describes the selection criteria adopted to reproduce the observations. § 6.3.2 presents the LBG-LBG auto-correlation. § 6.3.3 presents the results of the cross-correlation between DLAs and LBGs, and in § 6.3.4 we summarize the results from the simulated universe.

6.3.1 The simulations and the selection of DLAs at $z = 3$

Thanks to the advances in computational power in the past decade or so, numerical simulations of the universe have become more and more common. Typically, these numerical codes simulate a small volume of 10, 20 or 50 Mpc on a side, depending on the resolution. In all cases, they include dark matter, which dominates the universe. More recently, the baryons, which initially behave as a gas, have been added to numerical codes.

We used the TreeSPH simulations of Katz, Weinberg, & Hernquist (1996b), which have three kinds of particles: dark matter, gas, and stars. Dark matter particles are collisionless and only influenced by gravity, while gas particles are influenced by pressure gradients and shocks, in addition to gravity. The code combines smoothed particle hydrodynamics (SPH) (Lucy, 1977) with the tree algorithm for computation of the gravitational force (Hernquist, 1987). This formulation is completely Lagrangian, i.e. it follows each particle in space and time. A consequence of the SPH code is that, at each point in space, gas properties are known by smoothing over nearby particles,

typically 32. In this implementation of the SPH, the gas can cool radiatively. Star formation is included in the following way: gas particles are transformed into stars when a region's density reaches a certain threshold and is part of a convergent flow that is Jeans unstable. Also, photoionization by a spatially uniform UV background (Haardt & Madau, 1996) was included in the simulation.

The simulations that we used were run from redshift $z = 49$ to redshift $z = 0$. We used only the simulations at $z = 3$. The simulation were run with the following cosmological parameters: $\Omega_M = 0.4, \Omega_\Lambda = 0.6, h = 0.65, n = 0.93, \sigma_8 = 0.8$, which are close to the recent results from the Wilkinson microwave anisotropy probe (WMAP) (Bennett et al., 2003). It had 128^3 dark matter particles and the same number of gas particles in a periodic box of $22.222h^{-1}$ Mpc (comoving) with a gravitational softening length of $3.5h^{-1}$ kpc.

In summary, the standard picture for the formation of galaxies, and of large-scale structure, occurs as a result of the action of gravity on the dark matter since the universe is dominated by it. Gravity amplifies the small density fluctuations given by the linear power spectrum $P_l(k)$ (Eq. 6.46). Thus, bound structures of dark matter—called dark matter halos—that will form clusters, groups, and galaxies are formed by the gravitational collapse of the density perturbations (e.g. White & Rees, 1978). The baryons condense in the dark matter halos, cool, and then form stars wherever the local physical conditions permit.

Galaxy identification

To find the galaxies in the simulated universe, we needed an algorithm that can detect dense clumps of stars and cold gas. The algorithm that was used in this simulation is the group finding algorithm 'spline kernel interpolative denmax' (SKID) (Katz, Weinberg, & Hernquist, 1996b). To ensure the reliability of this method, a mass resolution of 6.8×10^9 $M_\odot$ was adopted. This corresponds to 64 SPH particles. 651 galaxies are resolved and detected by the algorithm at redshift $z = 3$.

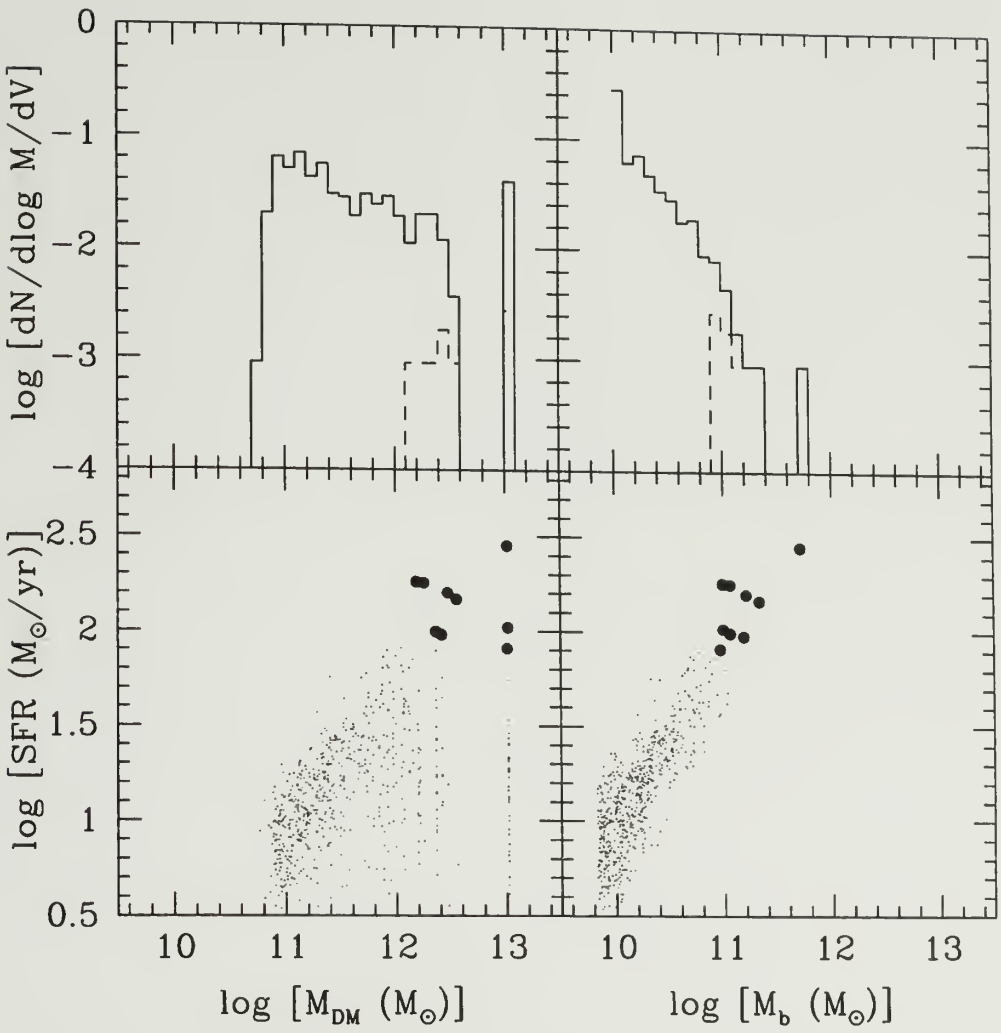

Figure 6.2 Top left: the dark matter mass function of the SKID-identified galaxies with baryonic masses M_b larger than the resolution 6.8×10^9 $M_\odot$, corresponding to 64 SPH particles. Top right: the baryonic mass function for the same galaxies. The dashed histogram shows the mass functions for the simulated LBGs. Bottom: the star formation rate (SFR) as a function of dark matter (left), baryonic mass (right). Because of the prescription for the SFR, galaxies with a large amount of baryons have large SFRs. The filled circles are our 'LBG' sample. They are selected according to their high SFR, and we matched the volume density to the observed volume density of LBGs (see text).

The top panels of Fig. 6.2 show the dark matter mass function (left) and baryonic mass function (right) of all the 651 SKID-identified galaxies that have a baryonic mass M_b larger than the resolution 6.8×10^9 M$_\odot$. The bottom panels of Fig. 6.2 show the star formation rate for the 651 galaxies as a function of their dark matter mass (left), and their baryonic mass (right). As expected from the prescription that turns gas into stars, galaxies with a large amount of baryons have large star formation rates.

Selection of LBGs in the simulations

Lyman break galaxies are forming stars actively, that is the main reason they are so bright after all. We selected galaxies in the simulations in a way to reproduce the observations: from the 651 galaxies, the simulated LBG sample is made of galaxies that had the highest star formation rates (SFRs). To simulate the effect of a flux-limited survey (as in our observations), we matched the number of galaxies per unit volume to the observed volume density. In our observations, there are approximately 80 to 120 LBGs in a volume of $40 \times 40 \times 100h^{-3}$ Mpc^3. Since, the simulations are $22.222h^{-1}$ Mpc on a side, only ~ 9 galaxies in the simulated universe will reproduce the observed volume density of our fields. These nine galaxies constitute our sample of simulated LBGs and is represented by the filled circles in Fig. 6.2. Since, this sample was made of the tip of the luminosity distribution, we selected another sample of the 100 most star forming galaxies. This constitutes the 'large' galaxy sample.

Gas column density maps

Since our interest was to find damped absorbers in the simulations, the H I density projected along one dimension onto a uniform grid is the closest analogy to real damped absorbers. In this case, each pixel represents a potential line of sight of a quasar. The prescription used to create such column density distributions was the same as in Gardner et al. (2001). From the intrinsic aspect of the SPH code, the gas properties are known at each point in space. Thus, the column density can be averaged (i.e. smoothed) around each galaxy before being projected on a uniform grid

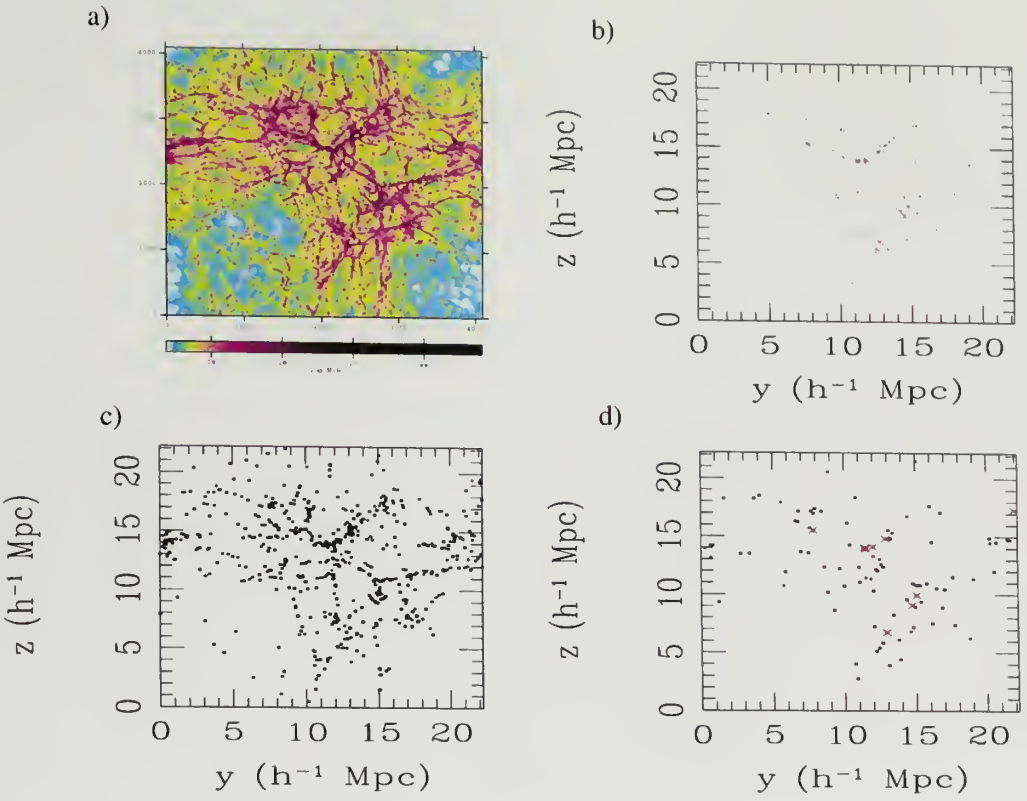

Figure 6.3 (a): Column density map of HI in the $22.222h^{-3}$ Mpc^3 volume projected along the x axis on a 4096^2 pixel grid. Potential DLAs with $N_{\rm HI}$ larger than $> 10^{20.3}$cm^{-2} appear black. (b): Position of potential DLAs projected along the x axis. The red crosses show the positions of the nine LBGs. (c): Position of the 651 galaxies that have a baryonic mass M_b larger than the resolution 6.8×10^9 M$_\odot$. (d): Position of the 100 most star forming galaxies. The red crosses show the positions of the nine LBGs.

as follows: first, around each SKID-identified galaxy, the column density of collapsed gas was computed by enclosing each halo with a sphere centered on the most bound particle. This sphere was of sufficient size to include all the gas particles that might contribute to high column density absorption within the halo. Then, the column density within this sphere was projected onto a uniform grid with 4096^2 pixels, each 5.43 kpc in size.

The HI column density was first calculated assuming the gas is optically thin and then corrected using a self-shielding correction as in Katz et al. (1996a). The HI column density projected along the x axis is shown in Fig. 6.3(a). A DLA was

selected from the column density map if its H I column density ($N_{\rm H\,I}$) was higher than $> 10^{20.3}{\rm cm}^{-2}$. In this simulation, there were approximately 115,000 pixels that met this criteria, and are shown in Fig. 6.3(b). Each pixel is a potential DLA, where one would need a QSO shining behind. This is in sharp contrast to our observations, which contain only four DLAs. Fig. 6.3(b) shows also the distribution of the simulated LBGs as red crosses. Fig. 6.3(c) shows the position of the 651 galaxies that have a baryonic mass M_b larger than the resolution 6.8×10^9 $M_\odot$. Fig. 6.3(d) shows the position of the 100 galaxies with the most star formation rate, and the position of the simulated LBGs as red crosses (see § 6.3.2). From Fig. 6.3, one sees that the galaxies and the DLAs are correlated.

If the sample of simulated DLAs was large enough to perform statistical analyses, our sample of simulated galaxies was too small: it contained only nine galaxies. To simulate the observations, where we have a volume $40 \times 40 \times 100 h^{-3}$ Mpc^3 (comoving), we used the same simulation several times, each time using a projection along a different axis, and randomizing the center of the additional simulations. We used four additional volumes, this brought the survey volume of the simulations to $22 \times 22 \times 111 h^{-3}$ Mpc^3 (comoving) and the simulate LBG sample to 45 galaxies.

6.3.2 The LBG auto-correlation

Katz, Hernquist, & Weinberg (1999) and more recently Weinberg et al. (2003) computed the 3-D galaxy-galaxy auto-correlation and the evolution of the bias b (see Eq. 6.47) in simulations similar to the one used here, but with different volumes. Katz, Hernqnist, & Weinberg (1999) used a $11.111 h^{-1}$ Mpc box, and Weinberg et al. (2003) used a $50 h^{-1}$ Mpc box. Fig. 6.4 shows the galaxy auto-correlation $\xi_{\rm gg}$ for the complete sample of resolved galaxies (top panel), for the sample of 500 most massive galaxies (to simulate an L^* sample) (middle panel), and for the dark matter halos. It shows that the galaxy auto-correlation $\xi_{\rm gg}$ does not evolve with redshift. The dark matter

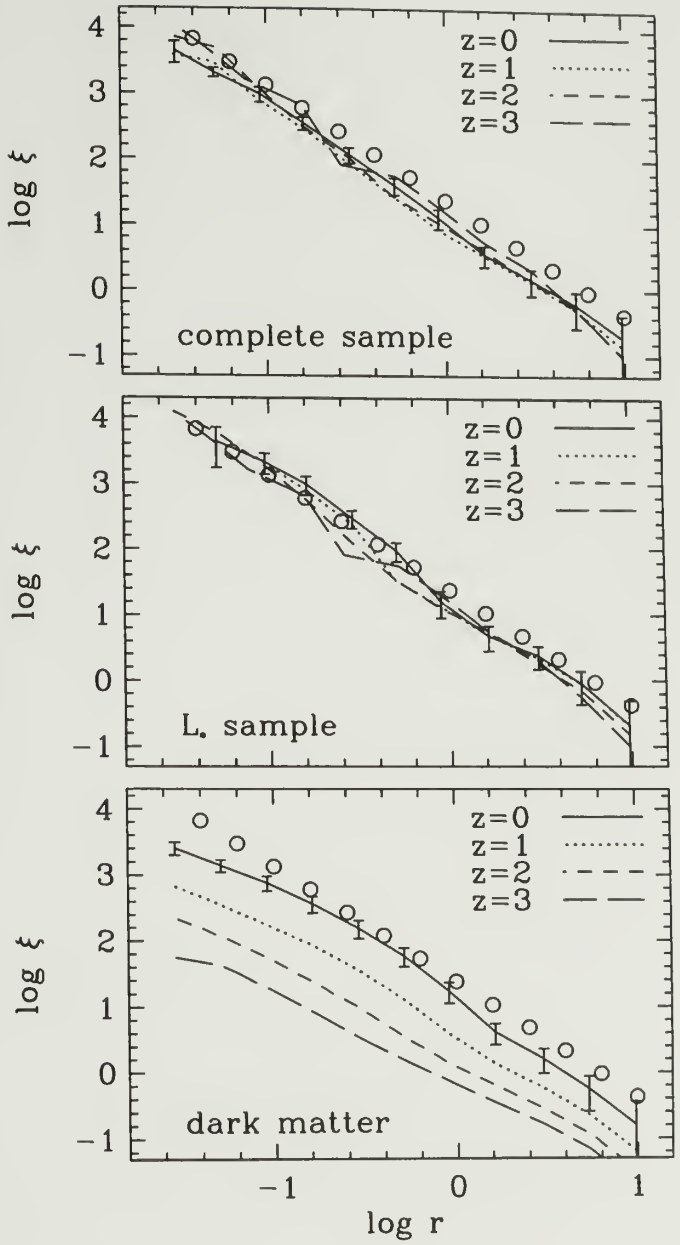

Figure 6.4 In the larger simulation of Weinberg et al. (2003), the galaxy-galaxy autocorrelation $\xi(r)$ is shown for the complete galaxy sample (top panel), the L^* sample (middle), and the dark matter halos (bottom), and redshifts $z = 0, 1, 2, 3$. Open circles show a power law with the parameters measured from the SDSS by Zehavi et al. (2002), $r_o = 6.1h^{-1}$ Mpc and $\gamma = 1.75$. (reproduced from Weinberg et al., 2003)

auto-correlation $\xi_{\rm DM}$, however, does evolve strongly with redshift as structures grow. The reason for this is that the bias b evolves with time. It is stronger at high redshifts, and decreases to ~ 1 at $z = 0$. Weinberg et al. (2003) fitted the auto-correlation with $\xi_{gg} = \left(\frac{r}{r_o}\right)^{-\gamma}$ and, at $z = 3$, found $r_o \simeq 4 \pm 0.3 h^{-1}$ Mpc, $\gamma = 1.90 \pm 0.07$.

We computed the angular galaxy auto-correlation in the simulations used here with the estimator $1 + \omega_{gg} = \langle DD/RR \rangle$. This estimator was similar to Eq. 6.24, which is the one used for the cross-correlation in § 6.3.3 and 6.4.1.

Fig. 6.5 shows our results of the auto-correlation. The filled triangles show the auto-correlation for our simulated LBG sample of 5×9 galaxies, which corresponds to the most massive galaxies. The filled circles show the auto-correlation for the larger sample made of 5×100 galaxies that have the most star formation rate. The error bars are computed by resampling the data, a method known as 'bootstrap resampling' (e.g. Mo, Jing, & Boerner, 1992). Fig. 6.5 shows that the simulated LBGs (filled triangles) are more clustered than the larger sample (filled circles). This reflects the fact that the nine simulated LBGs are among the most massive dark matter halos in the simulations.

How does this compare with the real world?

Adelberger et al. (2003) measured the auto-correlation $\xi_{gg} = \left(\frac{r}{r_o}\right)^{-\gamma}$ on a sample of ~ 500 spectroscopically confirmed LBGs, and found $r_o = 3.96 \pm 0.29$ Mpc and $\gamma = 1.55 \pm 0.15$. We used Eq. 6.37 (equivalent to Eq. 6.40) to convert the 3-D auto-correlation ξ_{gg} to the projected auto-correlation ω_{gg}, taking into account the volume of the simulations ($W_z = 111.11 h^{-1}$ Mpc). Fig. 6.5 compares the observed angular auto-correlation ω_{gg} of Adelberger et al. (2003) (dotted line) with the angular auto-correlation of the simulated LBGs (filled triangles). The auto-correlation of the simulated LBG sample is much stronger than the observed angular auto-correlation of Adelberger et al. (2003). This strong auto-correlation of the simulated LBGs may indicate that they are not representative of the real LBGs.

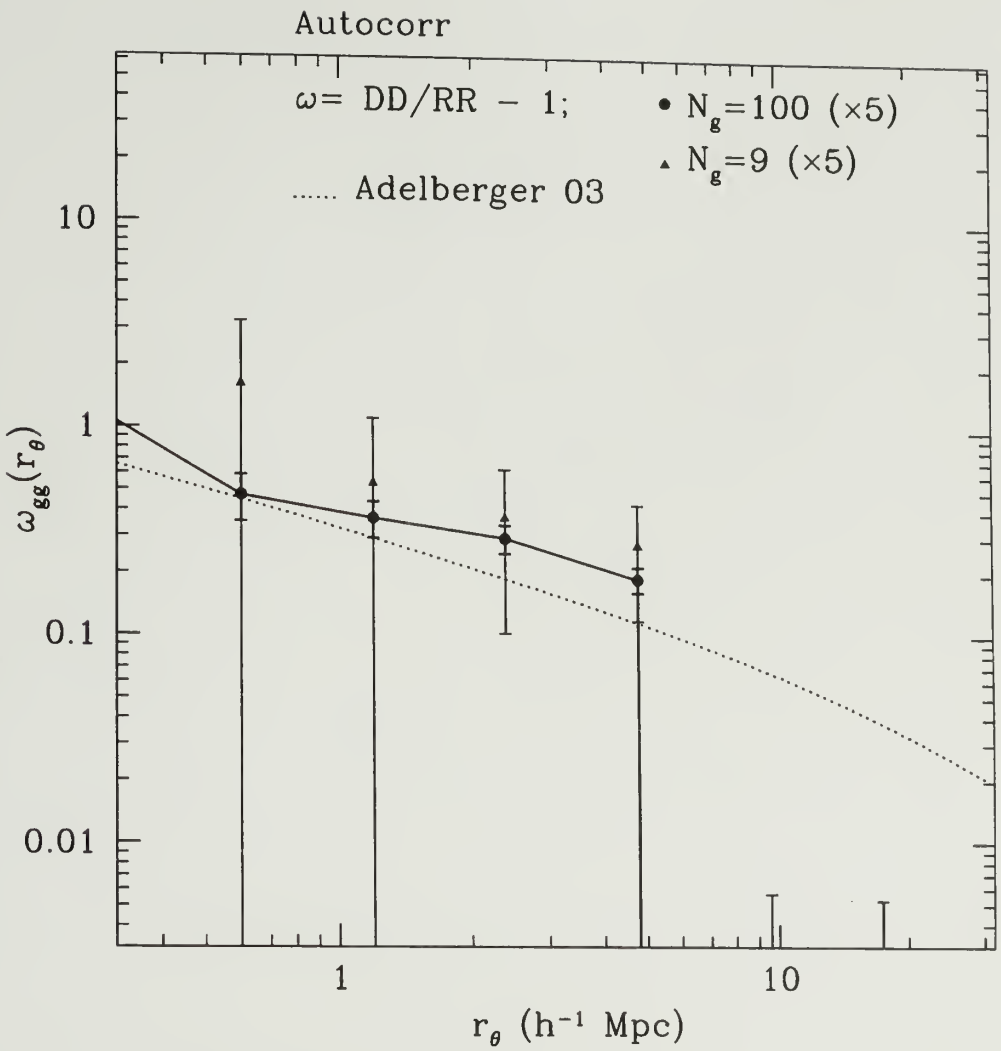

Figure 6.5 The angular LBG-LBG auto-correlation ω_{gg} for the sample of 9 simulated LBGs (filled triangles) that have the most star formation rate (or the most baryonic mass). The filled circles show the auto-correlation for the larger sample of 100 galaxies. The auto-correlation was computed using the estimator $< DD/RR > -1$ on five simulations. The center of the additional boxes was randomized. The error bars are computed using the bootstrap resampling method. The clustering is stronger for the more massive galaxies. We used $C = 0.04$ for the integral constraint. The dotted line show the observed LBG auto-correlation (from Adelberger et al., 2003, using Eq. 6.37).

As a side note, the exact value of the observed correlation length r_o quoted above is still being debated. For instance, Giavalisco & Dickinson (2001) reported $r_o = 5 \pm 0.7h^{-1}$ Mpc in a spectroscopic sample, and $r_o = 3.2 \pm 0.7h^{-1}$ Mpc in the HDF. And Porciani & Giavalisco (2002) reported $r_o = 4 \pm 1.2h^{-1}$ Mpc. Recently, the Canada-France deep field survey published their results of a large sample of LBGs (~ 1300), and found $r_o = 4.9 \pm 0.5h^{-1}$ Mpc and $\gamma = 1.8 \pm 0.2$ (Foucaud et al., 2003). Regardless of the exact value of r_o, the observed clustering of LBG at redshift ~ 3 is similar to the clustering of L^* galaxies today. Indeed, two recent surveys, the two degree field galaxy redshift survey (2dFGRS) and the Sloan digital sky survey (SDSS) have shown that local galaxies have $r_o = 4.9 \pm 0.3$ and $r_o = 6.1 \pm 0.2$ respectively (Norberg et al., 2001; Zehavi et al., 2002).[2]

That the clustering of LBGs at redshift ~ 3 is similar to the clustering of L^* galaxies today has an important consequence: the clustering bias b has to go from $\sim 3-4$ at $z = 3$ to ~ 1 at $z = 0$ because the clustering of dark matter halos (predicted in theoretical models) is much less at $z = 3$—as shown in Fig. 6.4. i.e. if the dark matter picture presented in § 6.2.8 and in § 6.3.1 has anything to do with the real universe. A stronger bias at redshift $z = 3$ means that LBGs reside in massive dark matter halos. This does not directly imply, however, that LBGs are massive galaxies. For instance, Somerville et al. (2001) argued that the strong star formation rate observed in LBGs is triggered by galaxy interactions (e.g. mergers) of small galaxies falling into a deep dark matter potential.

---

[2]The difference values reflects the difference between the bluer and redder selection of the two surveys since red galaxies are known to be more clustered.

6.3.3 The DLA-LBG cross-correlation with hundreds of DLAs

We are now ready to discuss the cross-correlation between galaxies and DLAs in the simulated universe. § 6.3.3.1 presents the main results. We discuss the errors of the cross-correlation functions in § 6.3.3.2 and summarize our results in § 6.3.4

6.3.3.1 The DLA-LBG cross-correlation

We used the estimator presented in section 6.2.5 (Eq. 6.24) to compute the angular cross-correlation between the 115.000 potential DLAs and the nine simulated LBGs in the $22.222^3\ h^{-3}$ Mpc^3. As mentioned in section 6.3, we used four additional simulated boxes to reproduce the observed volume. We selected random lines of sight from the 115.000 DLAs. For each line of sight, the galaxies in the additional volumes were projected along a random direction (x, y, or z) and the centers of the additional volumes were randomized.

Naturally nine galaxies is not a fair sample of the galaxy luminosity function. [3] This is an inherent problem due to the size of the simulation, rendering the comparison between the observed and the simulated cross-correlation difficult. A larger simulation is needed to produce a fairer sample of galaxies. Not to mention that the simulated LBGs are the brightest, while the observed LBGs are color selected.

In an attempt to qualitatively predict the observations where we have only a handful of DLAs, we cross-correlated the nine simulated LBGs (plus four additional volumes) with ten DLAs selected at random. That we have more DLAs than in the observations (four) should compensate the sparse sample of simulated LBGs. Fig. 6.6 (top panel) shows the DLA-LBG cross-correlation for ten DLAs. The error bars show the variance for such a sample and were computed by repeating the calculation 100

[3] Nine simulated galaxies correspond to a sample of 120 galaxies in the observations, which is a maximum size. For instance, a sample of 80 corresponds to five galaxies or so. The cross-correlation with such a small sample can be done, and the results presented in the chapter are not affected. The errors will be larger by a factor $\sim \sqrt{2}$.

Table 6.1. The number of DLA-galaxy pairs $N_{\rm obs}$ and the number of random-galaxy pairs $N_{\rm exp}$ per DLA, for the cross-correlation shown in Fig. 6.6, computed in the case of $N_{\rm DLA} = 10$.

r_θ	$N_{\rm obs}$ $N_{\rm gal} = 5 \times 9$	$N_{\rm exp}$ $N_{\rm gal}({\rm rand}) = 7000$	$\omega_{\rm dg}$	$\sigma(\omega)$
0.15	0.023	0.00843	1.66788	5.5500
0.3	0.071	0.03484	1.13934	2.6471
0.6	0.227	0.13675	0.68678	1.3157
1.2	0.754	0.55506	0.35753	0.5695
2.4	2.761	2.22817	0.23918	0.3025
4.8	9.638	8.88839	0.08437	0.1391
9.6	33.442	35.0389	-0.04561	0.0662
17.4	90.412	91.6030	-0.01301	0.0345

times, each time selecting ten DLAs at random. The dotted line is the LBG angular auto-correlation from Adelberger et al. (2003) calculated using Eq. 6.40. The solid line is a simple fit to the slope of the cross-correlation. The bottom panel shows the cross-correlation of ten random positions. The clustering signal vanished since the positions are now not correlated with the galaxies.

Several important points that will apply to our observational results are worth noting from Fig. 6.6.

- First, because of the size of the simulations, we were not sensitive to clustering beyond $10h^{-1}$ Mpc or so.

- Second, we were sensitive to clustering only on scales of $r_\theta \sim 3 - 7h^{-1}$ Mpc. Indeed, only on those scales is the error bar not consistent with zero (top panel). Equivalently, only on those scales is the error bar smaller than the auto-correlation (bottom panel).

- Third, the bias of the cross-correlation due to the integral constraint, which was pointed out in section 6.2.7, is nicely illustrated in the top panel of Fig. 6.6.

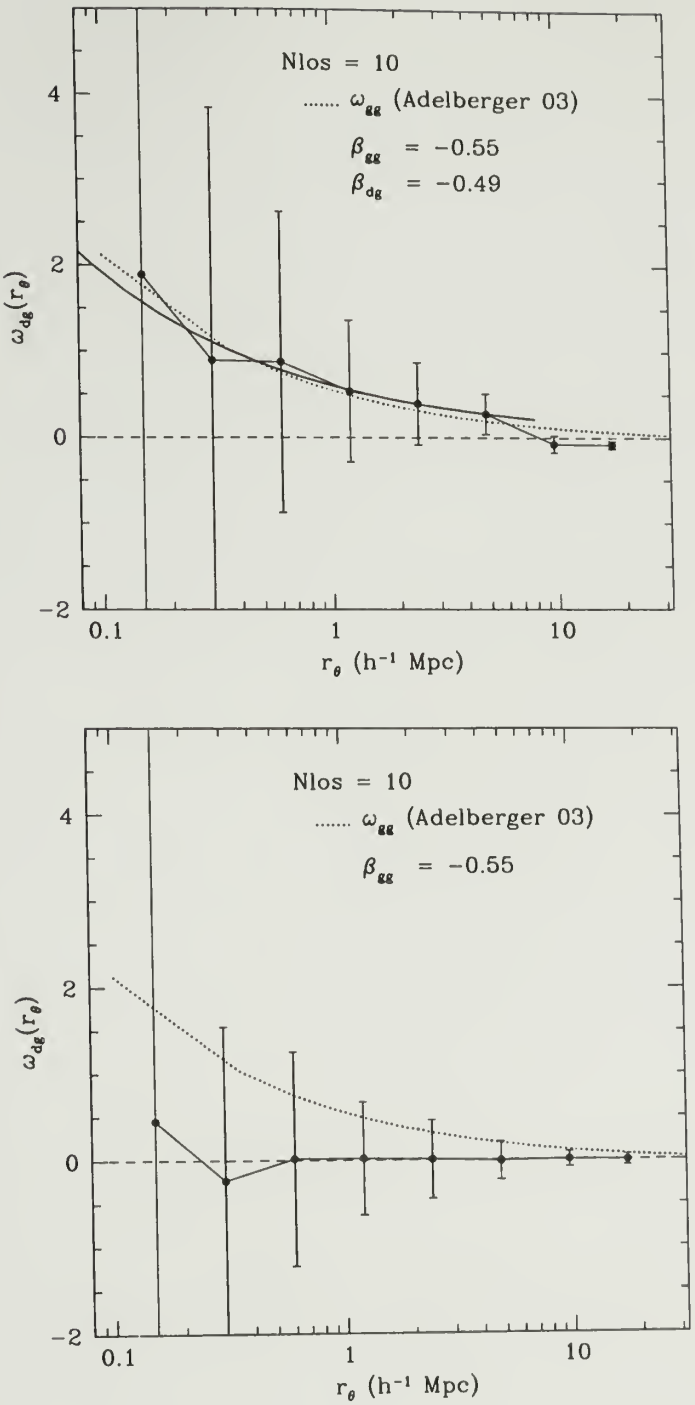

Figure 6.6 Top: The angular cross-correlation ω_{dg} between DLAs and LBGs in the simulated universe with only 10 lines of sight. A simple fit to the slope of ω_{dg} is shown by the continuous line. Bottom: The angular cross-correlation between 10 random positions and the simulated LBGs. The error bars represent the rms of 100 repetitions. The observed LBG auto-correlation has a slope $\beta_{gg} = -0.55$ and is shown as the dotted line. The dashed line represents the zero level.

Indeed, on the largest scales, $r_\theta > 8h^{-1}$, the cross-correlation tends to a negative value ~ -0.05, similar to the ~ -0.08 expected value from Eq. 6.36 using the auto-correlation of Adelberger et al. (2003).

- Fourth, given the rms of the cross-correlation (shown by the error bars), it will be difficult to differentiate between clustering and no clustering, i.e. between the top and the bottom panels, with just one realization of 10 lines of sight.

Table 6.1 shows the number of DLA-galaxy pairs $N_{\rm obs}$ and the number of random-galaxy pairs $N_{\rm exp}$ per DLA, for the cross-correlation shown in Fig. 6.6.

To reproduce our analysis in the observations, we performed the following modifications to Fig. 6.6: we used the same radial bins as in the observations, we corrected the data for the integral constraint C, and we used log scales. Fig. 6.7 corresponds to the modified top panel of Fig. 6.6. We also fitted the amplitude of the cross-correlation a with respect to the observed LBG auto-correlation ω_{gg}, using the method explained below. The inset shows the χ^2 as a function of a. Fig. 6.8 shows the same as Fig. 6.7, but with 200 DLAs instead of ten. As expected, the errors become significantly smaller.

In other words, we modeled the cross-correlation as a factor a times the auto-correlation ω_{gg}:

$$\hat{\omega}_{dg} = a \cdot \omega_{gg},$$

where a was the amplitude of the cross-correlation function with respect to the galaxy-galaxy auto-correlation ω_{gg}. We, then, corrected the estimated correlation function ω_{dg} for the integral constraint $C \equiv \Delta\Omega$, i.e. we use $\omega + C$ in Eq. 6.49.

To perform the fit shown in Fig. 6.7 and in Fig. 6.8, we needed to use the covariance matrix since each radial bin is correlated with the others. We computed the covariance matrix using the $n_r = 100$ calculations as follows:

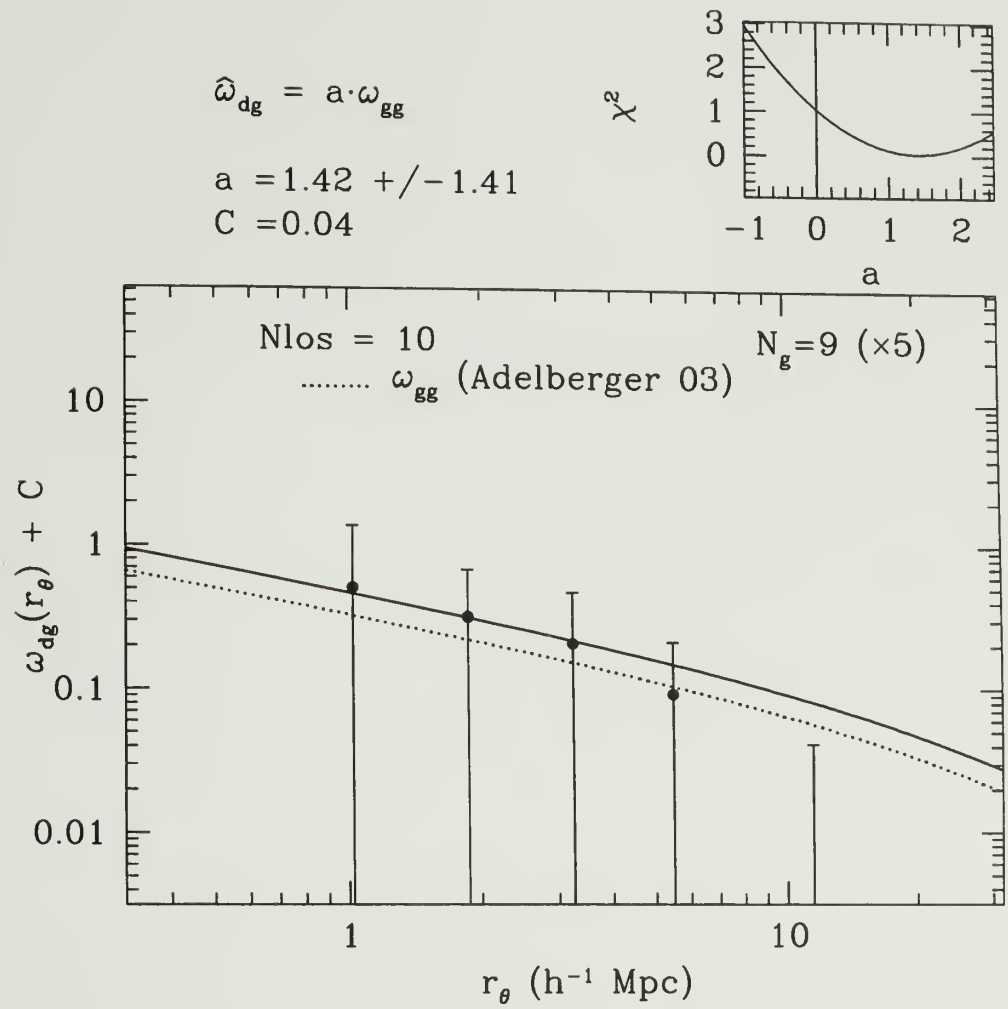

Figure 6.7 The angular cross-correlation between DLAs and LBGs in the simulated universe ten lines of sight. The error bars represent the rms of 100 repetitions. A weighted fit to the cross-correlation, corrected for the integral constraint C, is shown by the continuous line (see text). The fitted parameter a is shown. The inset shows the χ^2 distribution with the 1σ interval. The observed LBG auto-correlation has a slope $\beta_{gg} = -0.55$ and is shown as the dotted line (from Adelberger et al., 2003).

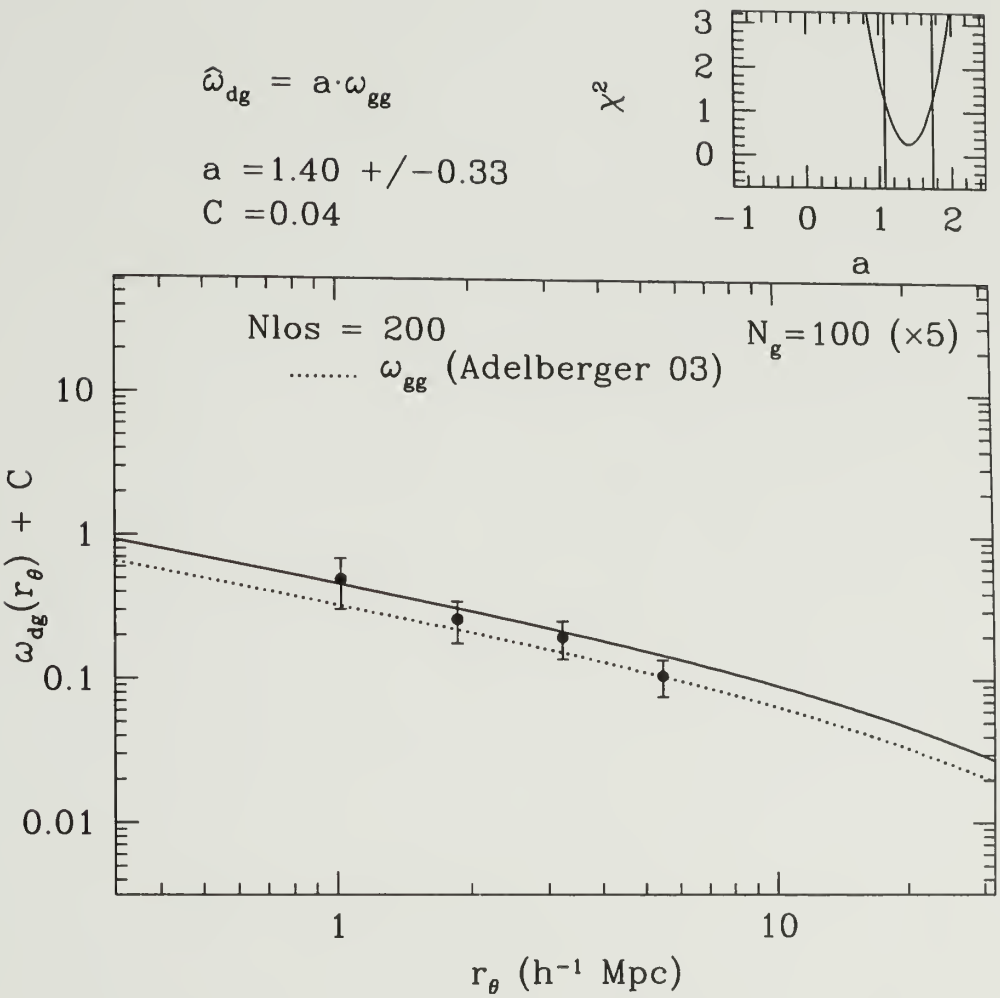

Figure 6.8 Same as Fig. 6.7 for 200 simulated DLAs. A sample larger than anyone could ever dream to achieve since there are only 300 DLAs known at all redshifts. A weighted fit to the cross-correlation is shown by the continuous line (see text). The χ^2 distribution with the 1σ lower limit is shown in the inset.

$$\mathrm{COV}(\omega(r_{\theta_i}),\omega(r_{\theta_j})) \equiv \mathrm{COV}_{ij} = \frac{1}{n_r - 1}\sum_{k=1}^{n_r}[\omega_k(r_{\theta_i}) - \overline{\omega}(r_{\theta_i})]\cdot[\omega_k(r_{\theta_j}) - \overline{\omega}(r_{\theta_j})] \quad (6.48)$$

where ω_k is the kth measurement of the angular cross-correlation and $\overline{\omega}$ is the average of the n_r measurements of the cross-correlation.

This allowed us to fit the cross-correlation using the covariance between the bins by minimizing χ^2 defined as:

$$\chi^2 = \sum_{i,j}[\omega_{\mathrm{dg}}(r_{\theta_i}) - \hat{\omega}_{\mathrm{dg}}(r_{\theta_i})]\mathrm{COV}^{-1}_{i,j}[\omega_{\mathrm{dg}}(r_{\theta_j}) - \hat{\omega}_{\mathrm{dg}}(r_{\theta_j})] \quad (6.49)$$

where $\hat{\omega}$ is a model of the cross-correlation, and COV^{-1} is the inverse of the covariance matrix (Eq. 6.48). In general this inversion is unstable, because $\mathrm{COV}_{i,j}$ is singular (e.g. Berstein, 1994). To avoid large instabilities in the inversion process, we used the single value decomposition (SVD) technique to invert the matrix COV as recommended by Berstein (1994). This involved diagonalizing the matrix and keeping the most stable eigenvalues. In practice, the last two radial bins are not included in the fit.

Since the model, $\hat{\omega}_{dg} = a \cdot \omega_{gg}$, was linear in its parameter a, the solution to Eq. 6.49 was analytic: $a = (\omega^T\mathrm{COV}^{-1}\hat{\omega})/(\hat{\omega}^T\mathrm{COV}^{-1}\hat{\omega})$ and the variance in the fitted parameter was $V(a) = 1/(\hat{\omega}^T\mathrm{COV}^{-1}\hat{\omega})$ (Berstein, 1994), where ω and $\hat{\omega}$ are the vector data and the vector model respectively.

The insets in Fig. 6.7 and in Fig. 6.8 show the χ^2 distribution and the best fitted amplitude a. The continuous line shows the fitted model $\hat{\omega}$. The small number of DLAs used in Fig. 6.7 is reflected by the uncertainty in the parameter a: the best amplitude is $a = 1.42 \pm 1.41$.

Does Fig. 6.8 mean that the DLA-LBG cross-correlation is stronger than the LBG-LBG auto-correlation? No; in the next paragraph, we show that in the simulation, the opposite is true. This comparison to the auto-correlation of Adelberger et al. (2003) was an attempt to reproduce the analysis of our observations, which is presented in section 6.4.

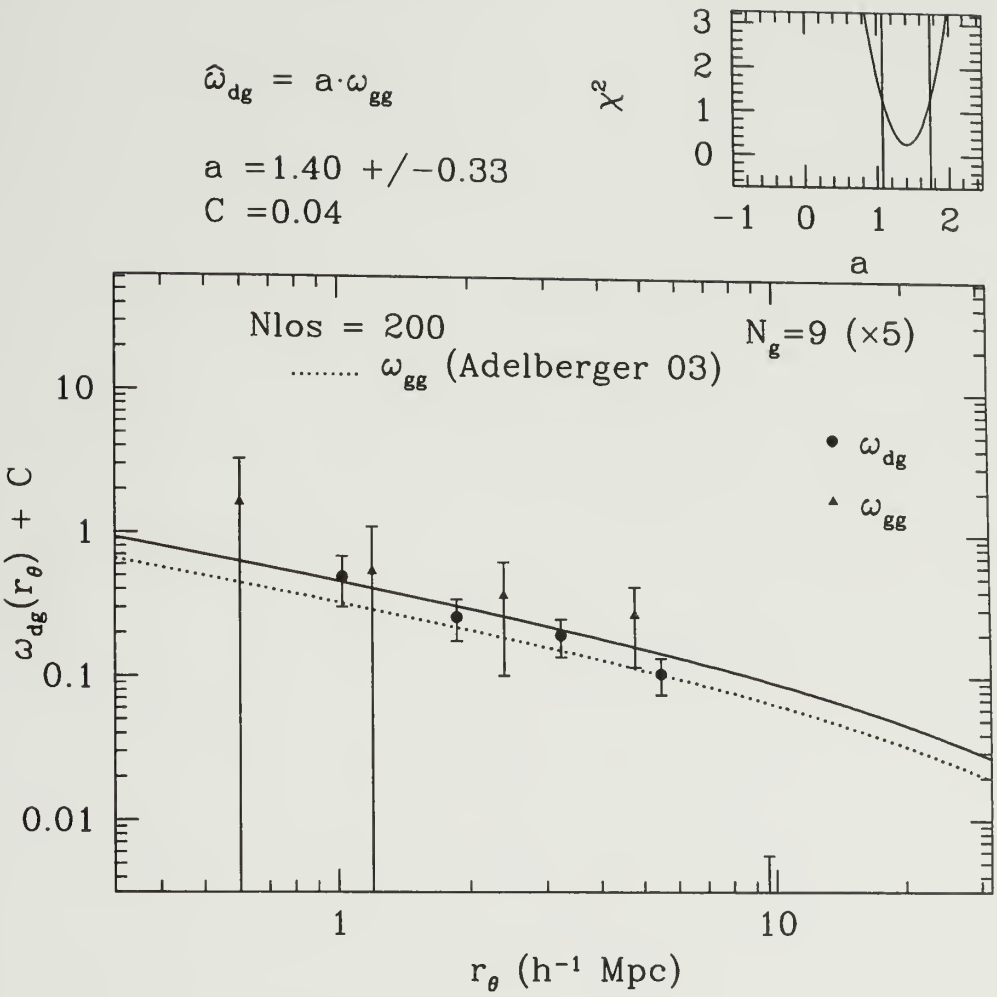

Figure 6.9 Comparison between the DLA-LBG cross-correlation (filled circles) shown in Fig. 6.8 and the LBG-LBG auto-correlation (filled triangles) shown in Fig. 6.5, both computed on the same sample of 45 simulated LBGs. A fit to the amplitude of the cross-correlation is shown by the continuous line. The fitted amplitude is shown a. The auto-correlation is stronger than the cross-correlation. This indicate that in these simulations, DLAs do not reside in dark matter halos as massive as the ones associated with LBGs. The observed auto-correlation of Adelberger et al. (2003) is shown by the dotted line.

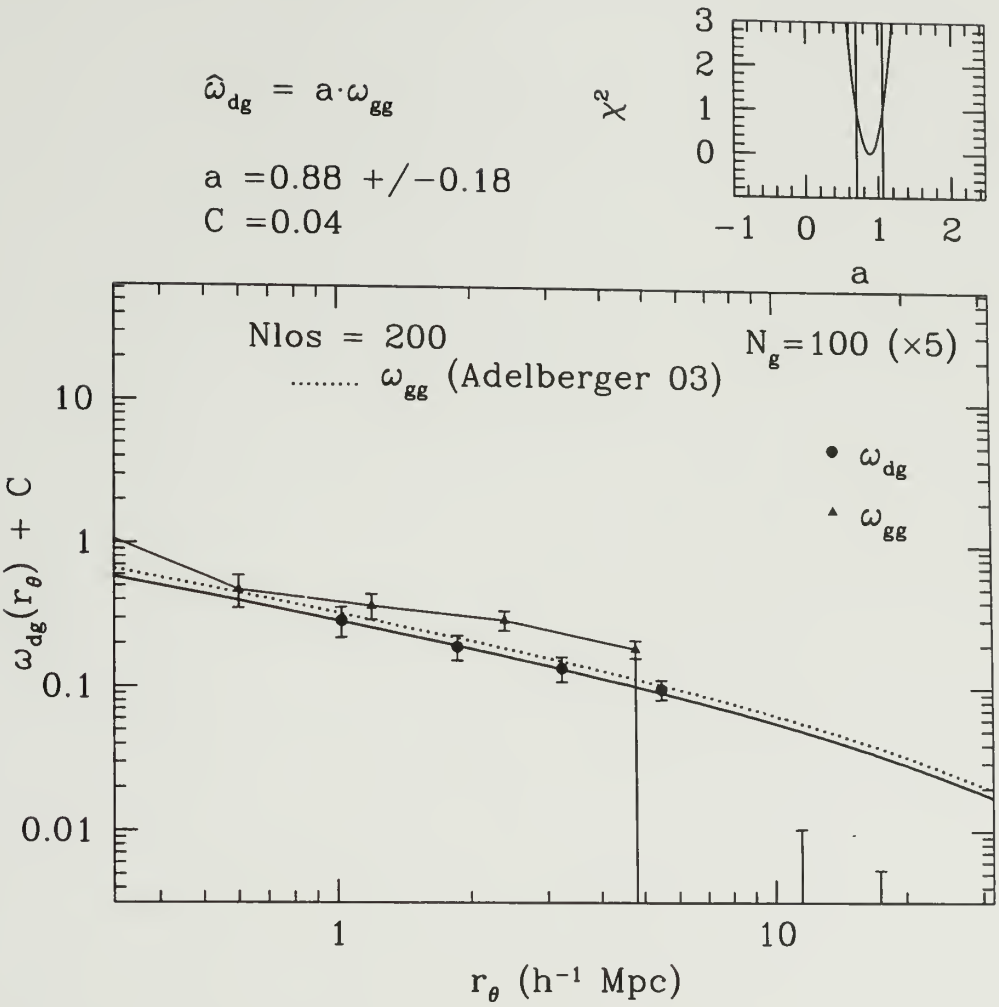

Figure 6.10 Comparison between the DLA-LBG cross-correlation (filled circles) and the LBG-LBG auto-correlation (filled triangles), both computed on the 'larger' sample of simulated LBGs made of the 100 most star forming galaxies. A fit to the cross-correlation is shown by the continuous line as in Fig. 6.8. The auto-correlation is stronger than the cross-correlation by a factor of $\sim 2 - -3$. This shows that in these simulations, DLAs do not reside in dark matter halos as massive as the ones associated with LBGs. The observed auto-correlation of Adelberger et al. (2003) is shown by the dotted line.

We close this section by noting that we are comparing the cross-correlation in the simulations to the auto-correlation in the real world measured by Adelberger et al. (2003). A better way to compare the cross-correlation in the simulation is against the auto-correlation of simulated galaxies (Fig. 6.5). Fig. 6.9 compares the LBG-LBG auto-correlation (filled triangles in Fig. 6.5) to the DLA-LBG cross-correlation (Fig. 6.8). The fitted amplitude a and the χ^2 distribution are shown as in Fig. 6.8. This comparison showed that the cross-correlation was weaker that the LBG-LBG auto-correlation. This meant that in these simulations, DLAs did not reside in dark matter halos as massive as the ones associated with LBGs.

However, as discussed towards the end of last § 6.3.2, the LBG-LBG auto-correlation of the simulated galaxies (filled triangles in Fig. 6.5) is much stronger than the observed auto-correlation from Adelberger et al. (2003). This could imply that the simulated LBGs might not represent the observed LBG population. Thus, we compared the autocorrelation of the larger galaxy sample made of the 100 most star forming galaxies with their cross-correlation with the potential DLAs, shown in Fig. 6.10. The auto-correlation of the larger sample is represented by the filled triangles and the DLA-LBG cross-correlation by the filled circles. This comparison confirms that the cross-correlation is weaker than the LBG-LBG auto-correlation. The fitted amplitude a and the χ^2 distribution are shown as in Fig. 6.8. This comparison, again, meant that DLAs did not reside in dark matter halos as massive as the ones associated with LBGs.

6.3.3.2 Error analysis of the cross-correlation

Fig. 6.11 shows the errors of the cross-correlation presented in Fig. 6.6 (top panel). The Poisson errors for a random DLA sample, i.e. with no cross-correlation $\omega_{dg} = 0$ (see Fig. 6.6, bottom panel), are shown by the dotted line. The Poisson errors of the cross-correlation, represented by the continuous line, are computed using Eq. 6.27.

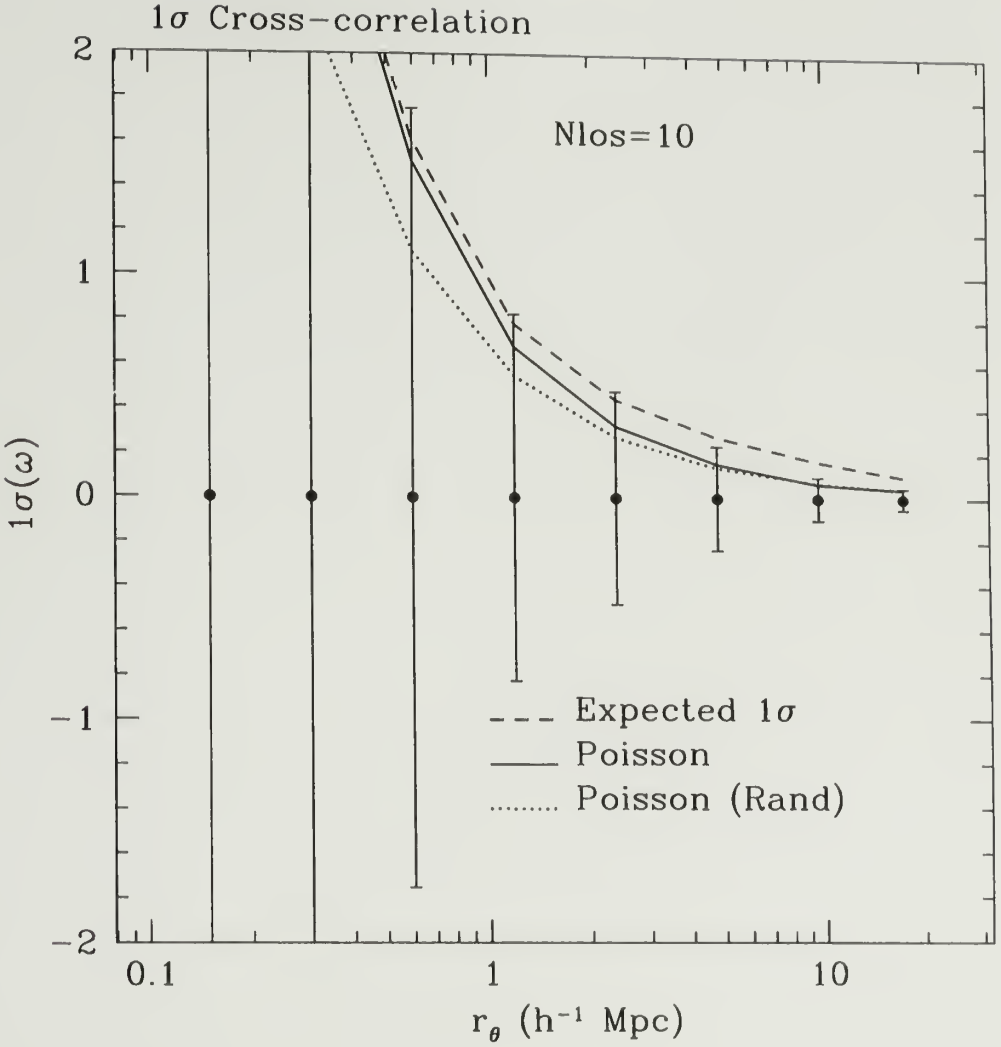

Figure 6.11 The errors to the DLA-LBG cross-correlation are shown for 10 lines of sight. The dotted line shows the rms expected in the case when there is no cross-correlation. The continuous line shows the rms expected from Poisson noise (Eq. 6.27). The difference between the dotted line and the continuous line shows the effect of the cross-correlation on its variance. The expected rms to the cross-correlation (Eq. 6.29) is shown by the dashed line. The difference between the dashed line and the continuous line shows the effect of the clustering variance V_{cl}.

This shows that the cross-correlation increases the variance to ω_* as expected from Eq. 6.27. The dashed line shows the expected errors of the cross-correlation function ω_{dg} computed from Eq. 6.29 using the observed galaxy auto-correlation ω_{gg} from Adelberger et al. (2003).

Because the error bars are larger than the expected Poisson errors (continuous line), Fig. 6.11 shows that the Poisson errors underestimate the variance of the cross-correlation especially on large scales. Indeed, on large scales, the clustering variance affects the variance of ξ_{dg} at the 10-20% level. The good agreement between the analytic prediction (dashed line) and the errors computed from the Monte Carlo realizations indicates that the error budget is understood. Minor differences between the dashed line and the computed error bars are due to approximations in Eq. 6.29.

6.3.4 Summary of the results from the simulations

- The error budget is well accounted for (Fig. 6.11).
- The auto-correlation ω_{gg} of the nine most star forming galaxies is very strong (Fig. 6.5). This indicates that the simulated LBGs reside in the most massive dark matter halos.
- The cross-correlation is weaker than the auto-correlation of the same galaxies by a factor of approximately two (Fig. 6.9). This indicates that DLAs are not among the most massive dark matter halos. Indeed, if they were, the cross-correlation would be the same as the auto-correlation.
- For the larger sample of simulated galaxies, the fitted amplitude of the cross-correlation is $1.40 \pm 0.33\times$ the observed auto-correlation

6.4 The LBG/DLA connection. Results from the observations

We used three estimators of the LBG/DLA cross-correlation ω_{dg}:

1. the estimator presented in § 6.2.4—Eqs. 6.20 or 6.34,

2. the volume averaged estimator of section § 6.2.6, and

3. the more robust estimator of section § 6.2.5

Results for the first two estimators were applied to one field, the APM 08279+5255 field, and resulted in the publication of Bouché & Lowenthal (2003). At the time of the writing of Bouché & Lowenthal (2003), the four fields were not all available. The results with the combined fields using the estimator presented in § 6.2.5, superseded the results of Bouché & Lowenthal (2003). For that reason, we only present our final results of the cross-correlation in § 6.4.1.

6.4.1 The LBG-DLA cross-correlation

Here, we present our results on the observed DLA-LBG cross-correlation using the same method applied to the simulations presented in section 6.3.3.

There are two major differences between the observed and simulated data set. First, in the simulations, the same galaxies were used with all the DLAs since we had the luxury of picking any of the 115,000 available DLAs. This made the calculation of the sum of the DLA-galaxy pairs in Eq. 6.24 straightforward. In our data, we had only a handful of DLAs and the galaxy sample associated with each differed. This was due to the fact that each DLA was at a slightly different redshift and thus our selection was not the same in the four fields. The different selections were reflected by the different number of galaxies in each field (see section 5.6). Thus, in performing the sum in Eq. 6.24, we must weight each field accordingly.

Second, because we have only four DLAs at our disposal, the error bars on the cross-correlation for each individual field must be computed carefully. One way to compute the error bars is to randomly resample the galaxies (bootstrap technique); another way is to randomize the position of the DLA throughout the field. The proper way to compute the errors would be to resample the DLAs. This, however, is impractical with only four DLAs. Each method has its pros and cons. The first method reproduces only the Poissonian errors, but not the clustering variance. The second method takes into account the clustering variance due to the galaxy auto-correlation. The third method takes into account the clustering variance due to the galaxy auto-correlation and due to the cross-correlation. However, on large scales $r_\theta > 5h^{-1}$ Mpc, the clustering variance due to the cross-correlation ($1+\xi_{dg}$ in Eq. 6.29) will be small since $\xi_{dg} << 1$.

Thus, we used the second option, that is, $n_r = 200$ random positions of the DLA. By construction, this Monte Carlo process takes into account the Poisson variance and the clustering variance due to the galaxy auto-correlation. The covariance matrix COV_{ij} was obtained, as in section 6.3.3, from the n_r randomizations. The rms $\sigma_\omega(r_i)$ at the scale r_{θ_i} was naturally given by $\sqrt{\mathrm{COV}_{ii}}$. As mentioned, this is only an approximation to the covariance of the DLA-LBG cross-correlation function. Strictly speaking, it is the covariance for random lines of sight. However, it includes the variance and the co-variance due to the LBG-LBG auto-correlation.

The DLA-LBG cross-correlations for the four fields are shown in Fig. 6.12. In each panel, the dotted line is the LBG-LBG autocorrelation measured by Adelberger et al. (2003) using Eq. 6.37 to account for the volume of our redshift slice. Because in each case there is one and only one DLA, the individual measurements of the cross-correlation shown in Fig. 6.12 are plagued with large errors because the Poisson noise dominates. Thus, one should not pay attention to individual measurements. The cross-correlation is a measurement of an averaged quantity (see Eq. 6.24).

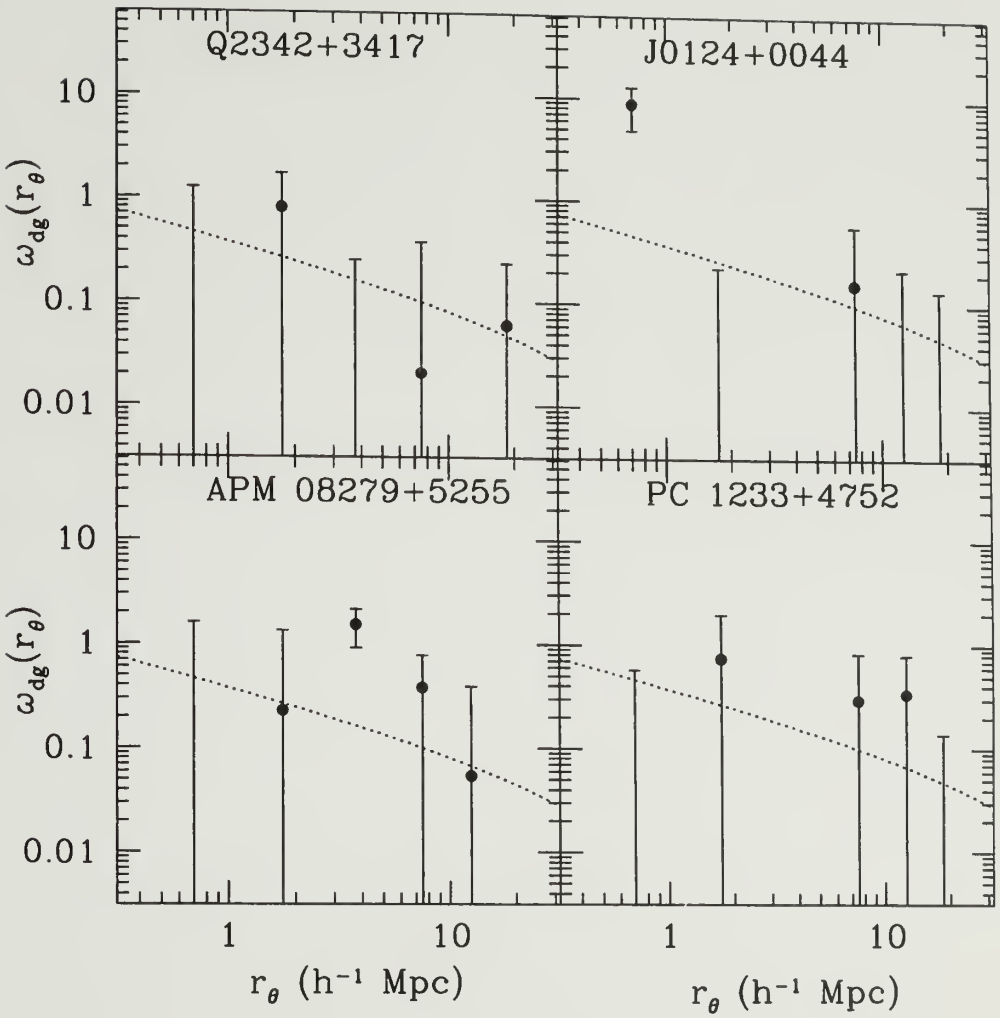

Figure 6.12 The observed cross-correlation ω_{dg} in the four fields. The error bars were computed by randomizing the line of sight 200 times. The dotted line is the galaxy-galaxy auto-correlation measured by Adelberger et al. (2003).

Combining the four fields

Thus far, we have measured $\frac{N_{\rm obs}}{N_{\rm rand}}$ four times. The last step in order to estimate the DLA-LBG cross-correlation is to perform the average indicated in Eq. 6.24 by the brackets. As mentioned, it is necessary to take into account the different selections (such as redshift, and magnitude limits) at play in the different fields. Thus, we weighted each field with its errors using the standard error propagation formulae. Namely, for the mean cross-correlation $< \omega(r_i) >$, the variance to the mean $\sigma_{<\omega>}(r_i)$, and the covariance to the mean $\mathrm{COV}_{<\omega>}(r_i, r_j)$:

$$< \omega(r_i) > \equiv f(r_i) = \frac{\sum_{l=1}^{N_{\rm DLA}} \omega_l(r_i)/\sigma_l(r_i)^2}{\sum_{l=1}^{N_{\rm DLA}} 1/\sigma_l(r_i)^2} \tag{6.50}$$

$$\begin{aligned} \sigma^2_{<\omega>}(r_i) &= \sum_{l=1}^{N_{\rm DLA}} \left(\frac{\partial f(r_i)}{\partial \omega_l(r_i)} \right)^2 \sigma_l(r_i)^2 \\ &= \frac{1}{\sum_{l=1}^{N_{\rm DLA}} 1/\sigma_l(r_i)^2} \end{aligned} \tag{6.51}$$

$$\begin{aligned} \mathrm{COV}_{<\omega>}(r_i, r_j) &= \sum_{l=1}^{N_{\rm DLA}} \frac{\partial f(r_i)}{\partial \omega_l(r_i)} \frac{\partial f(r_j)}{\partial \omega_l(r_j)} \mathrm{COV}_l(r_i, r_j) \\ &= \sum_{l=1}^{N_{\rm DLA}} \frac{1/\sigma_l(r_i)^2}{\sum_l 1/\sigma_l(r_i)^2} \frac{1/\sigma_l(r_j)^2}{\sum_l 1/\sigma_l(r_j)^2} \mathrm{COV}_l(r_i, r_j), \end{aligned} \tag{6.52}$$

where $N_{\rm DLA}$ is the number of DLAs, and we used the following notation for each measurement l of the DLA-LBG cross-correlation ω_l, its error σ_l and its covariance COV_l:

$$\left.\begin{aligned} \omega_l(r_i) &= \frac{N_{\rm obs}(r_{\theta_i})}{N_{\rm rand}(r_{\theta_i})} \\ \sigma_l(r_i) &= \sigma\left(\frac{N_{\rm obs}(r_{\theta_i})}{N_{\rm rand}(r_{\theta_i})} \right) \\ \mathrm{COV}_l(r_i, r_j) &= \mathrm{COV}(\omega_l(r_{\theta_i}), \omega_l(r_{\theta_j})) \end{aligned}\right\} \text{for } l = 1, 2, 3, 4. \tag{6.53}$$

The result of Eq. 6.50 is our best measurement of the DLA-LBG cross-correlation ω_{dg} and is shown in Fig. 6.13. We corrected ω_{dg} for its bias C due to the integral constraint assuming that ω_{dg} is ω_{gg} in Eq. 6.36. The error bars $\sigma_l(r_i)$ were computed

using Eq. 6.51. In applying Eq. 6.50, we have rejected one field, the field Q2342+3417 with the lowest galactic latitude $b = -25\,\mathrm{deg}$, because this field suffers from a large stellar contamination (see discussion in § 4.8). The paucity of our data due to the fact that we have only three DLAs is represented by the large error bars in Fig. 6.13.

Is this result consistent with no clustering, i.e. with $\omega_{\mathrm{dg}} = 0$?

It is clear from Fig. 6.13 that we will not be able to constrain the slope of the cross-correlation, so we assumed that the cross-correlation ω_{dg} has the same shape as the auto-correlation ω_{gg}. That is, we fitted the cross-correlation to the model

$$\hat{\omega}_{\mathrm{dg}} = a \cdot \omega_{\mathrm{gg}},$$

where a is the unknown amplitude.

We performed the same analysis presented earlier in section 6.3.3.1, i.e. we fitted the amplitude of the cross-correlation a using the covariance $\mathrm{COV}_{<\omega>}(r_i, r_j)$ computed in Eq. 6.52, and the χ^2 statistic given by:

$$\chi^2 = \sum_{i,j} [\omega_{\mathrm{dg}}(r_{\theta_i}) - \hat{\omega}_{\mathrm{dg}}(r_{\theta_i})]\mathrm{COV}^{-1}_{i,j}[\omega_{\mathrm{dg}}(r_{\theta_j}) - \hat{\omega}_{\mathrm{dg}}(r_{\theta_j})]. \tag{6.54}$$

where $\hat{\omega}_{\mathrm{dg}}$ is the model of the angular cross-correlation and COV^{-1} is the inverse of the covariance matrix. In performing the inversion of the covariance matrix, we did reject the eigenvalue corresponding to the last bin. The inset in Fig. 6.13 shows the χ^2 distribution as a function of the amplitude a.

We found that an amplitude of $a > 0$ was favored:

$$a = 1.53 \pm 1.19. \tag{6.55}$$

The significance of this result is given by testing the null hypothesis $H_o : a = 0$ against $H_1 : a > 0$. For four degrees of freedom ($N - 1 = 5 - 1$), the null hypothesis

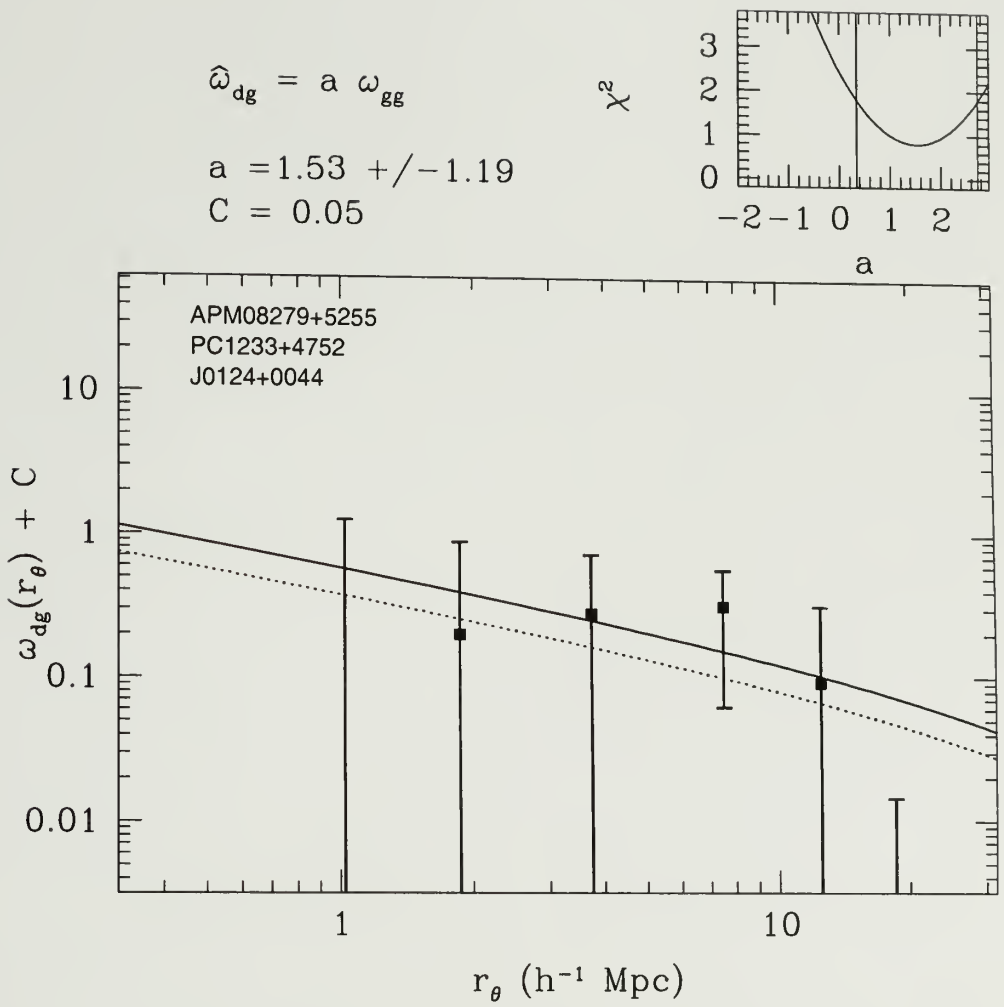

Figure 6.13 The cross-correlation between DLAs and Lyman break galaxies in a redshift slice of width ($W_z = 0.15$) that contain the DLAs. The filled squares show the averaged cross-correlation for the three fields shown (see text for discussion). The dotted line is the LBG auto-correlation (from Adelberger et al., 2003, using Eq. 6.37 to account for the volume of our redshift slice.). The continuous line is a fit to the amplitude of the cross-correlation normalized to the auto-correlation function. The inset shows the χ^2 distribution and the 1σ range.

is rejected with a maximum probability of 87%. Indeed, the student t-distribution ($t_4(\alpha)$) equals 1.285(= 1.53/1.19) for $\alpha = 0.13$. In other words, this gives a 87% confidence level to $a > 0$.

This first measurement of the DLA-LBG cross-correlation is most significant on scales greater than 5-10h^{-1} Mpc. We cannot, however, say much about the amplitude itself, i.e. whether or not the DLA-LBG cross-correlation is stronger or weaker than the LBG-LBG auto-correlation.

Is this result drawn from random lines of sight?

Given the large rms to the fitted amplitude a, could our result simply be a large fluctuation of the set of possible values for random lines of sight? To test this, we computed the cross-correlation for 20 lines of sight selected at random in the redshift slice centered on the DLAs. As before, we used $\left\langle \frac{N_{\rm obs}}{N_{\rm rand}} \right\rangle$ to compute the cross-correlation. We excluded the central $5h^{-1}$ Mpc to ensure that the new lines of sight are not correlated with the one centered on the DLAs.

Fig. 6.14 shows the logarithm of the $\chi^2_{\rm min}$ as a function of the fitted amplitude a for the 20 random lines of sight (filled circles). The open square show the location of the result of Fig. 6.13. The open square do not belong to the distribution defined by the random lines of sight. This shows that the signal measured in Fig. 6.13 is not drawn from a random distribution of lines of sight.

How about other redshift slices?

The result of Fig. 6.13 should be compared with the cross-correlation when there is no DLA in the redshift bin. From our photometric redshift analysis, we selected galaxies in two other redshift slice that did NOT contain the DLA. We chose the slices that were in the foreground and in the background from the DLA, and offset by -0.15 in redshift (see Fig. 5.15). Fig. 6.15 shows the result in the foreground slice. Fig. 6.16 shows the result in the background slice. In each case, the χ^2 fit, shown in the inset, does not favor any clustering. The best amplitude is $a = -0.54 \pm 1.20$

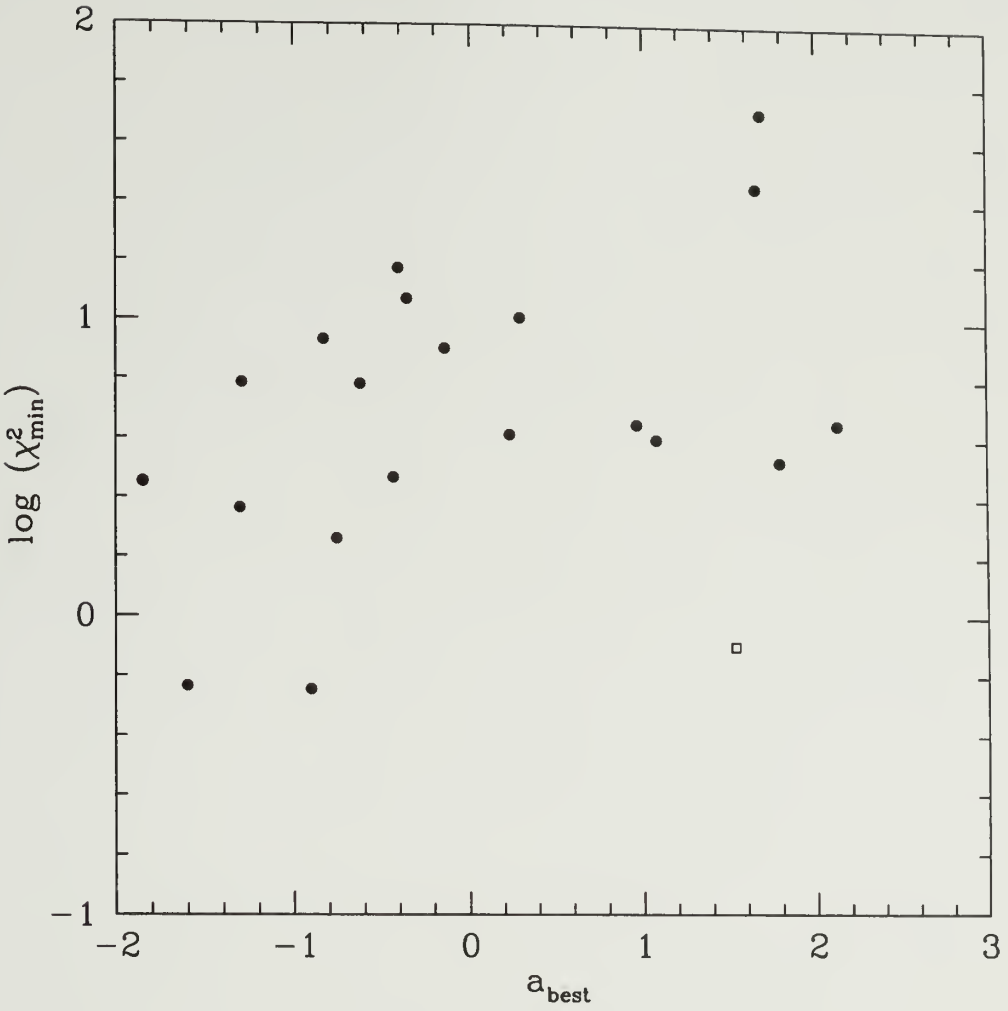

Figure 6.14 The filled points show the logarithm of the $\chi^2_{\rm min}$ as a function of the fitted amplitude a for 20 random lines of sight (excluding the central $5h^{-1}$ Mpc) in the redshift slice centered on the DLAs. The open square shows the situation of Fig. 6.13. Because the open square do not belong to the distribution defined by the random lines of sight, this shows that the signal measured in Fig. 6.13 is not drawn from a random distribution of lines of sight.

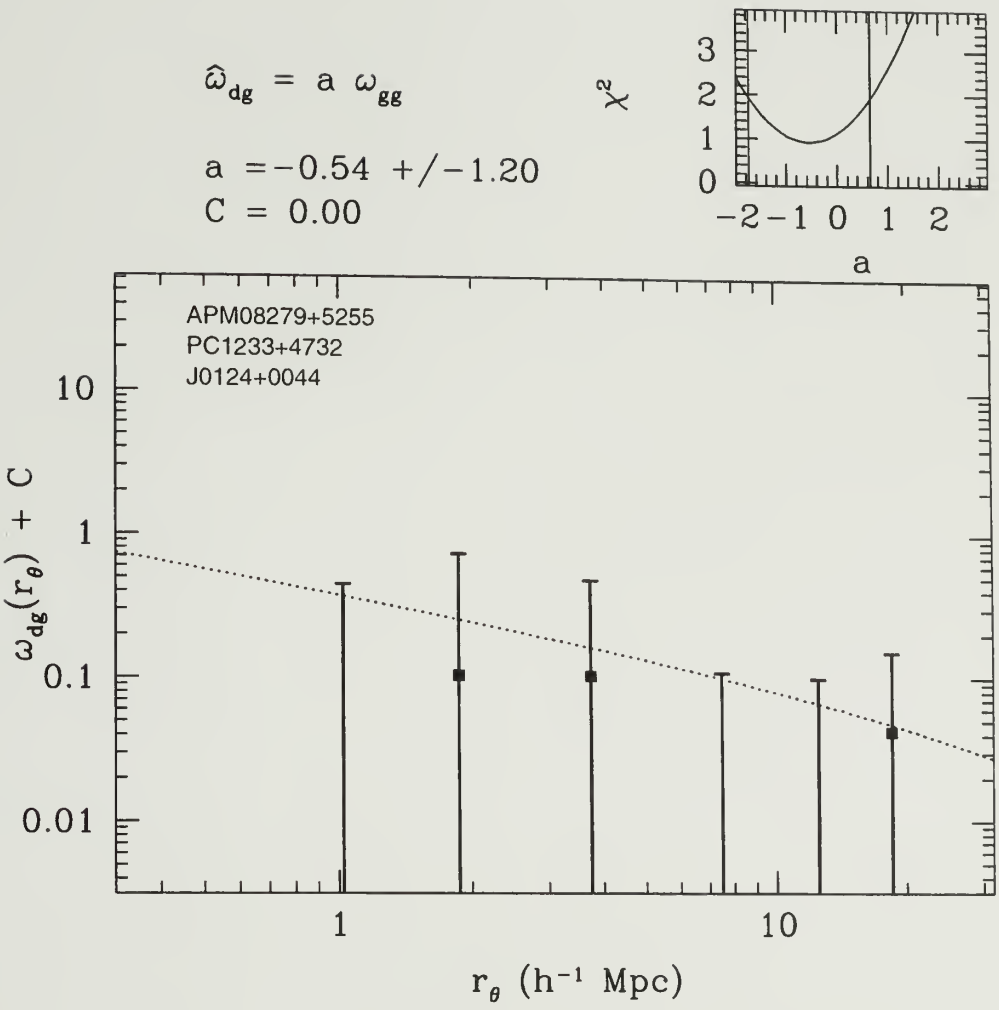

Figure 6.15 Same as Fig. 6.13 for galaxies in a foreground redshift slice of the same width ($W_z = 0.15$). This slice does not contain the DLA, and thus there should be no cross-correlation between the position of the DLA and the galaxies. The inset shows the χ^2 fit to the amplitude of the cross-correlation and the fit is completely consistent with $\omega_{\rm dg} = 0$, $a = -0.54 \pm 1.20$.

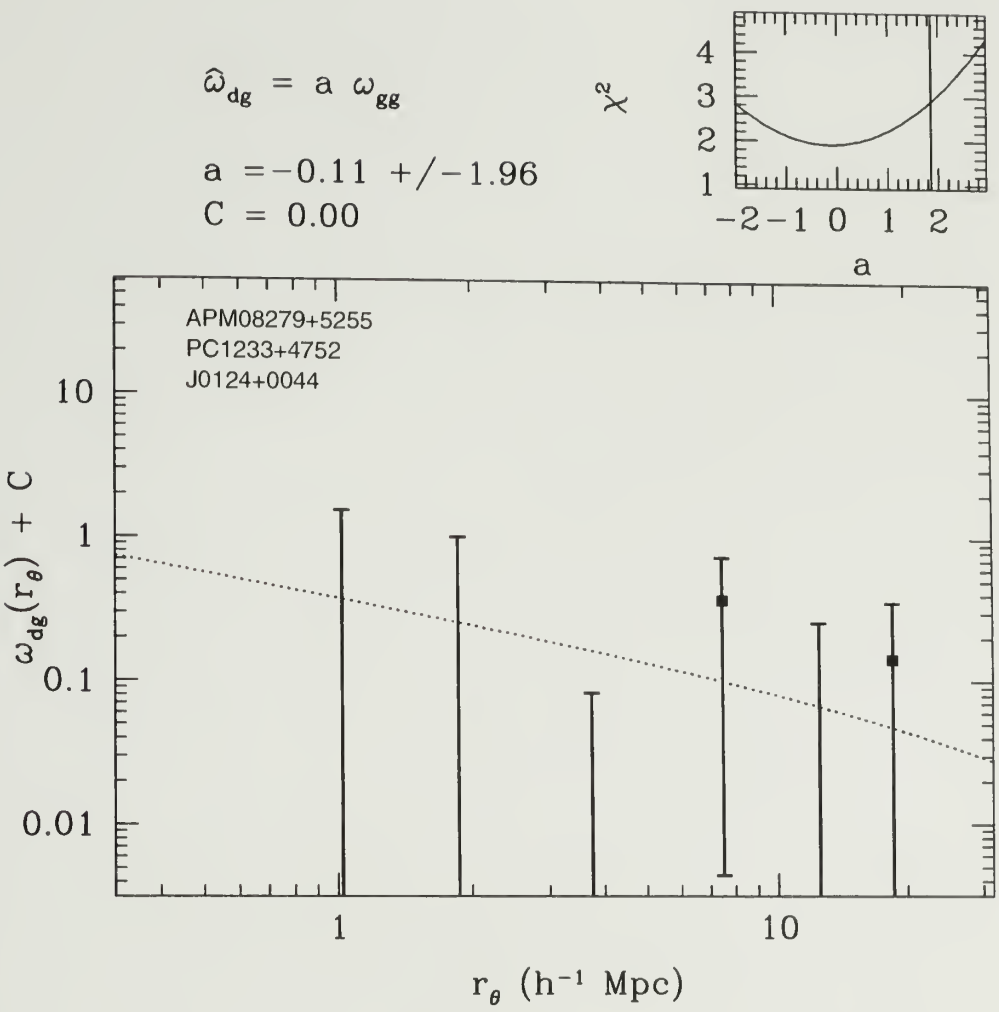

Figure 6.16 Same as Fig. 6.13 for galaxies in a background redshift slice of the same width ($W_z = 0.15$). This slice does not contain the DLA, and thus there should be no cross-correlation between the position of the DLA and the galaxies. The inset shows the χ^2 fit to the amplitude of the cross-correlation and the fit is completely consistent with $\omega_{\rm dg} = 0$, $a = -0.11 \pm 1.96$.

and $a = -0.11 \pm 1.96$, respectively. A clustering signal in this slice would have cast a strong doubt on our results that do contain the DLA in Fig. 6.13. In addition, we performed the same check on a slice at redshift 3.6. The best amplitude for this slice $a = -0.18 \pm 1.55$.

6.4.2 Summary of the observational results.

At redshift $z \sim 3$, we measured the cross-correlation between DLA and LBGs on scales of $\sim 7h^{-1}$ Mpc for the first time. Indeed, the results of Fig. 6.13 indicate that

- we detected the cross-correlation between DLAs and LBGs on large scales $r_\theta > 4h^{-1}$ Mpc at the 87% confidence level. Given the small number of DLA fields, at present, we could not conclude whether or not the DLA-LBG cross-correlation was stronger or weaker than the LBG auto-correlation.

- the χ^2 value to the fit is not drawn from a random distribution of lines of sight,

- in a total of three slices that did not contain the DLAs, there was no evidence for clustering.

It is worth mentioning that our observational results are in qualitative agreement with the results from the simulations, presented in § 6.3.3, given the differences to the galaxy selection and sample sizes.

Our results are not inconsistent with other authors (see § 2.10). For instance, Gawiser et al. (2001) studied the clustering environment towards one DLA at $z = 4$ and did not find any evidence for cross-correlation between LBGs and the DLA. This result is in agreement with ours because there is little hope to detect any clustering signal with one DLA as shown in section 6.3.3. The study of Adelberger et al. (2003), which is very similar to ours, found a lack of galaxies near their four DLAs. As mentioned in § 2.10, these two surveys were not sensitive to clustering on scales larger than $> 5h^{-1}$ Mpc. This lack of galaxies near the DLAs on small scales is also

consistent with our results since we cannot be conclusive on such small scales and they cannot be conclusive on large scales. The lack of galaxies on small scales could, however, be due to more local environmental effects, such as strong galactic winds from star forming galaxies.

CHAPTER 7

CONCLUSIONS

Based on wide-field images, we cross-correlated LBGs with a sample of three damped absorbers. The large area covered by our images, 0.31 deg^2 or $\sim 40 \times 40$ Mpc comoving at redshift $z = 3$, and their depth , $\mu_{I,\mathrm{AB}}(sky) \simeq 27.6$ mag arcsec^{-2}, allowed us to identify ~ 500 LBG candidates (per field) brighter than $I_{AB} = 24.80$ at $2.75 < z < 3.25$ both close (50kpc) to the DLA line-of-sight and up to 20 Mpc (comoving) from the DLA, i.e. physically unrelated.

LBG candidates were selected using photometric redshift techniques that included the I magnitude as a prior estimate in addition to the colors. The two were combined using Bayes' theorem following Benítez (2000). This helped to break the degeneracies that occur in a pure spectral template fitting scheme. The overall rms is $\sigma_z \simeq 0.15$ at $z \sim 3$ based on our analysis of photometric redshifts in the HDF-N. From the redshift likelihood distributions, we selected LBG galaxies within a redshift slice of width $W_z = 0.15 (\simeq \sigma_z)$ centered on the redshift of the DLAs z_{abs}.

Within that slice, we cross-correlated the LBGs with the position of the DLAs. For the first time, we detected the DLA-LBG cross-correlation on scales larger than $4h^{-1}$ Mpc (comoving). We found that this positive detection $a > 0$ was significant at the 87% confidence level. In addition, the clustering signal was not present in three redshift slices that did not contain the DLAs and was not consistent from being a random fluctuation.

The amplitude of the cross-correlation ω_{dg} relative to the auto-correlation ω_{gg} was $\omega_{dg}/\omega_{gg} \equiv a = 1.53 \pm 1.19$. This uncertain is dominated by the small number of DLA

fields in hand; a larger sample should allow us to test whether the cross-correlation is stronger or weaker than the auto-correlation. In the former case, this would favor the large disk hypothesis of Wolfe et al. (1986). In the latter case, it would exclude the large disk hypothesis. Our results cannot put constraints on the model of Maller et al. (2000), which predicts a large number of smaller satellite galaxies within $100h^{-1}$ kpc of the DLA.

In the simulated universe, we found that, at redshift $z \sim 3$, the cross-correlation is weaker than the auto-correlation of the same galaxies by a factor of approximately two. This indicates that DLAs do not reside in the most massive dark matter halos. Indeed, if they did, the DLA-LBG cross-correlation would be the same as the LBG-LBG auto-correlation. This difference could be connected to the difference between the luminosity of DLAs and of the LBGs. In the local universe, luminous galaxies are more clustered than other galaxies (e.g. Brown et al., 2003). This is consistent with the picture presented in § 2.8 where DLAs are made of the galaxies populating the faint end of the luminosity function. Furthermore, on small scales $< 1h^{-1}$ Mpc, the DLA-LBG cross-correlation increases at the high end of the column density distribution indicating that, in the simulations, damped absorbers are related to the LBG galaxy population.

Our observation of clustering on large scales ($> 4h^{-1}$ Mpc) is not inconsistent with previous clustering studies (Gawiser et al., 2001; Adelberger et al., 2003) since these were limited to small scales. A larger sample of DLAs, combined with spectroscopic redshift, should allow us to probe the small scales.

Future work includes the analysis of 80 LBG spectra in one of our fields from our recent multi-object spectroscopy program on the Gemini telescope. This will enable us to refine the parameterization of the Bayesian photometric redshifts and to decrease our errors in the cross-correlation.

APPENDIX A

THE UNIVERSE AS WE UNDERSTAND IT IN 2003

Since this study is focused on a particular type of object (namely hydrogen gas clouds) in the early universe, when it was about 20% of its age, it is important to understand not only the broad context in which this project lies, but also how it fits in the overall evolutionary picture of the universe. This section is an attempt to put this research in the general context of our understanding of the universe in 2003. This section is slightly more detailed than the previous one and is aimed at the more avid reader.

The universe is made of light, matter, dark-matter and an unknown 'dark energy'. In the following section, we present the various contributions of these constituents. Section A.2 describes the foundations of the Big Bang theory. Section A.3 summarizes the first three minutes of the universe, section A.4 summarizes the next 400,000 years, and section A.5 shows how stars and galaxies came about. Finally, section A.6 focuses on the universe's content at two different epochs, 13.7 billion years and 3 billions years after the Big Bang.

A.1 What is the universe made of?

To begin, it is important to present what astronomers think the universe is made of. According to recent measurements of relic radiation from the Big Bang by the Wilkinson microwave anisotropy probe (WMAP), the total content the universe is split into matter (26.8%) and 'dark energy' (73.2%). The two are usually noted as Ω_M and Ω_Λ.

Only 4.4% of the universe is, however, in normal matter (i.e. protons, neutrons and electrons), and the remaining 95.6% is unknown. The normal matter is also called baryonic matter. Some (one fourth) that we see: $0.3 - 1\%$ is in luminous matter (stars and galaxies that emits light), and some (three fourths) that we don't see, called baryonic dark-matter.

Since 4.4% of the universe is in normal matter, the remaining of the 26.8% of matter, or 22.4% is made of a new and unknown type of matter that interacts mainly with gravity, which is called non-baryonic dark-matter. In other words, astronomers don't have a clue of what makes 95.6% of the universe, and should be ready for more surprises! Finally, today, approximately 0.0045% of the total content of the universe is in the form of photons (radiation).

The total (100%) average density of the universe corresponds to about 6 atoms per cubic meter, which, clearly, is not very dense at all. To further put this in perspective, this means that the universe would have to contract by a factor of a billion (10^9) to reach the density of water (which is about 3×10^{28} atoms per cubic meter). Since 4.4% of the total universe's content is in the normal matter form, there are in fact only about 0.25 atom of normal matter per cubic meter on average (or 10^{-30} grams per cubic cm).

On the other hand, all the light (photons) make only $\sim 0.0045\%$ of the universe's content. This may seem insignificant, but since most of them are in radio-waves (cm-waves to be precise), i.e. low energy, it turns out that they have to be very numerous: the universe contains 400 million photons per cubic meter. Thus, there are about 2 billion (2×10^9) more photons than atoms/nuclei in the universe.

A.2 The evolving universe.

From several observations, it is clear the universe has to have had a beginning, and will have an end. Stars don't live forever: They 'burn' hydrogen into helium (up

to Iron in their last phase). Even recycling stars cannot last forever for they produce more and more 'heavy elements'. Eventually, they will use up all the available fuel (elements lighter than Iron).

Today's leading theory of the evolution of the universe is the "Big Bang" theory. This theory has been confirmed by three direct observations.

The first observation is that galaxies are moving away from each other. The famous astronomer Edwin Hubble measured the velocity of nearby galaxies through the shift of spectral lines in their spectrum and discovered that distant galaxies are moving away from us faster the further away they are (e.g. Hubble, 1925). Distant galaxies are moving faster and appear redder because they have their spectrum shifted towards the red. This 'redshift' is then a measure of distance (defined in section 1.3). A redshift (z) of 3 means that the light has been shifted towards the red by a factor $1 + z = 4$. and that galaxies were 4 times closer together.

This outward motion of galaxies discovered by Hubble immediately implies that the universe's density is decreasing with time, and that the universe had to be denser (and therefore hotter) in the past. We can track this contraction back in time using Einstein's theory of General Relativity almost up to the time $t = 0$ when the density becomes infinite and the theory breaks down. The 'Big Bang' refers then to the epoch of the very early universe when we don't understand anything anymore, and let us set the cosmic clock $t = 0$.

The second observation is the discovery of the cosmic microwave background (CMB) by Penzias & Wilson (1965). This relic radiation (now at a temperature $T = 2.73$K) actually dominates the sky at all wavelengths and dates from about 400,000 years after the Big Bang, when the universe was about 3000 K and was sufficiently transparent to allow all the photons to travel freely to us. The existence of this radiation is a direct consequence of the Big Bang and the expanding universe. In addition, this radiation contains a few ripples at the 10^{-5} level. The size and the

amplitude of these ripples contain a lot of information, including the total content of the universe, and the amount of normal baryonic matter. Quite impressively, this measurement of the amount of normal matter is in very good agreement with results from the third observation.

The third observation is the abundances of light elements such as helium, lithium and deuterium. These elements were the only ones produced by fusion in the first three minutes, when the physical conditions in the universe were just right to favor fusion. It turns out that the output of these nuclear reactions, as measured with the abundances of these elements with respect to hydrogen, is a function of of the amount of of normal matter (e.g. Peebles, 1966; Burles & Tytler, 1998). It turns out that the amount of normal matter measured from the primordial elements is in complete agreement with the results from the measurement of the ripples of the cosmic microwave radiation.

A.3 The first three minutes: Big Bang nucleosynthesis.

Astronomers can figure out in what physical condition the universe was in the very distant past and what happened because we know both the matter density and the radiation density today, and how these scale with the expansion of the universe.

Up until 0.1 milliseconds after the Big Bang, there were no stable atomic nuclei. The temperature was greater than 10^{13} K which corresponds to the energy equivalent of the mass of a proton or a neutron. When the temperature dropped below 10^{13} K, almost all the protons/neutrons combined with anti-protons/anti-neutrons and converted their mass into radiation. Luckily for us, there was a small excess of matter over antimatter (one out of every billion protons survived).

At around one second ($T = 3 \times 10^{10}$ K), the same process happens to the electrons. This occurs later because electrons are much less massive. The universe is now made of electrons, protons, but still dominated by photons.

From about ten seconds ($T = 10^9$ K,$z \sim 10^{10}$), and up to three minutes or so, when the universe was slightly denser than water, but very hot, nuclear reactions started. Protons and neutrons combined to form deuterium, helium and other light elements. About three minutes later when the temperature reached $T = 10^8$ K, these nuclear reactions stopped because the universe's temperature and density decreased sufficiently rapidly and no other elements were produced. In other words, the universe simply ran out of time. Only helium, and traces of deuterium, and lithium were produced. As mentioned, a measure of the abundance of these light elements constrains the density of normal matter (baryons).

From that point on, the 4.4% of the universe is made of hydrogen (three fourths), helium (one fourth), traces of deuterium, and lithium nuclei and continues to cool down.

A.4 The universe becomes transparent.

Not much happens after nucleosynthesis. The universe cools down. However, radiation cools faster[1] than the matter. At around 200,000 years ($z = 3000$; $T = 9,000$ K) after the Big Bang, that the matter density started to dominate over the radiation density. However, the universe was still not transparent: photons were absorbed or scattered many times over short distances along their path of travel, and so the universe was totally opaque[2]. Later on, nuclei and electrons combined to form atoms and the universe became neutral.

---

[1]If the matter density evolves as expected as the cube of the expansion, the energy density of radiation, on the other hand, scales as the fourth power, i.e. when the universe was 4 times smaller, the radiation energy density was $256 = 4^4$ larger. The extra factor of 4 comes from the fact that the energy of a photon scales with the expansion of the universe.

[2]matter and radiation were still strongly coupled through interactions between electrons and photons called 'Thompson scattering'.

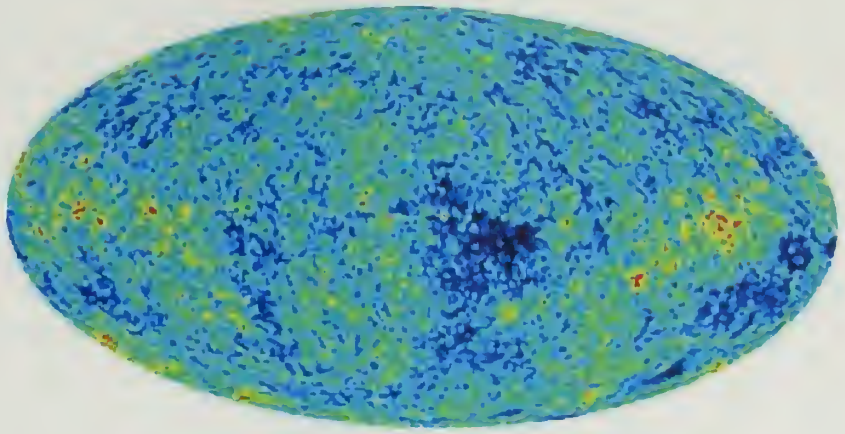

Figure A.1 The ripples in the Cosmic Microwave Background from the recent results of Wilkinson Microwave Anisotropy Probe. The largest fluctuations on this picture are 10^{-5} smaller than the uniform radiation. The shape and amplitude of these fluctuations are a function of various parameters of the Big Bang theory, such as the baryonic matter content, the dark matter content, and the total energy content of the universe From http://map.gsfc.nasa.gov/m_mm.html.

Shortly after this, around 400,000 years after the Big Bang ($z = 1000$; $T = 3000$ K), when the density of the universe was already very very small—comparable to the mean density of matter inside a galaxy *today* or about 10^{-23} grams per cubic cm—, the universe became transparent to light and photons can travel in a straight path without being absorbed or scattered. This happens because the radiation cooled down faster than the matter. From that point on, photons traveled in straight lines until some reach our telescopes. It is those photons—but much cooler, now at 2.73 K—that we see now in centimeter-waves coming from all directions discovered by Penzias & Wilson (1965) and measured by the cosmic background explorer (COBE) satellite in 1992. This confirmed the prediction that the universe was denser and hotter in the past.

A.5 From galaxy seeds to the first stars and galaxies.

A theory yet to be discovered, quantum-gravity, is thought to have produced very small fluctuations in the very early universe ($t < 10^{-33}$ sec) on atomic scales.

Then, inflation occurred, a time when the universe underwent a extremely rapid 'inflationary' expansion. Inflation is thought to (1) explain the uniformity of the microwave background, and (2) expand these fluctuations to very large cosmological scales. Since matter and radiation were strongly coupled up to $t = 400,000$ yr, these small fluctuations in matter density were predicted to be visible in the background radiation and, indeed, were measured by COBE for the first time, and more recently by the Wilkinson Microwave Anisotropy Probe (WMAP, Bennett et al., 2003) (see Fig A.1). These fluctuations grew, under the influence of gravity, and eventually created galaxy clusters and galaxies.

For instance, when the universe just became transparent ($z = 1000$), these small fluctuations imply that, in rough terms, the density of (mainly non-baryonic) matter in a volume of space that now contains the Milky Way, was 0.5% higher than in adjacent regions. Because the density was higher there, gravity held that region together against the expansion of the universe, and this volume of matter expanded more slowly than surrounding regions. As a result, its relative over-density grew. When the universe was about 18 million years old ($z = 100$), that same region of space was 5% denser than the surrounding regions. This gradual rise in local density continued as the universe expanded. When the universe was about 1.2 billion years old ($z = 5$), this region of space was probably twice as dense as neighboring regions.

This might seem like a very slow process, but fluctuations on smaller scales (that formed the first stars) started with a larger amplitude and therefore reached a large over-density at much earlier times. As a consequence, stars were among the first object formed at around 200 million years ($z = 20$) after the Big Bang.

After 500 million years (or $z \sim 10$), the first galaxies were assembled. After three billions years ($z \sim 3$), the epoch of interest for this work, young galaxies dominated the sky. The strong ultraviolet radiation coming from the young stars in these galaxies re-ionized the universe. If small pockets of neutral hydrogen remained in galaxies,

most of the universe is ionized. Finally, at eight billion years ($z \sim 0.5$), our solar system was formed.

A.6 The content of the universe

Now, that the global evolution of the universe was presented, I would like to focus on one of its component, the normal baryonic matter. Section A.6.1 presents *where* most of the baryons are today, 13.7 billion years after the Big Bang, while section A.6.2 contrast the current situation with the one three billions years after the Big Bang.

A.6.1 Today, 13.7 billion years after the Big Bang.

As we saw earlier, only 4.4% of the universe is in the form of normal matter, also called baryonic matter. Today, the nature of a humbling 95.6% remains unknown. From direct observations, the 4.4% can be broken into the following components (still expressed in terms of the total content) (Fukugita et al., 1998): 1.4% is in hot ($T = 10^6$ K) ionized gas in small galaxy clusters or groups[3]; 0.35% is made of stars in galaxies[4]; 0.26% is in ionized gas in galaxy clusters; 0.11% is in massive compact halo objects (MACHOs), such as brown dwarfs; 0.04% is in neutral hydrogen gas (in disks) Rosenberg & Schneider (2002); 0.03% is in molecular hydrogen.

From these numbers, we can see that, interestingly, the bulk of the baryons are in the ionized gas in clusters and groups of galaxies and never made it into galaxies! All these components do not add up to 4.4%. The sum is 2.2%, although some of these numbers are uncertain and are in fact marginally consistent with the 4.4% expected as discussed by Fukugita et al. (1998) and Hogan (1999). However, note that the sum

[3]This is an extrapolation from clusters of galaxies (Fukugita et al., 1998).

[4]Stars in elliptical galaxies and in the bulges of spiral galaxies account for 0.26%, while stars that are in disks make 0.09%.

is not *more* than the 4.4% expected both from the WMAP results. This is another consistency check of the Big Bang theory.

Thus, today, most of the baryons are the hot ionized gas in clusters. It is important for the context of this work to note that the neutral gas amounts to only 1% of the baryons, that the neutral gas and the stellar components differ by a factor of 7 and that most of the neutral hydrogen is in galaxies of some sort.

A.6.2 At $z \simeq 3$, 3 billion years after the Big Bang

Ten billion years ago at redshift $z \simeq 3$, the situation is drastically different. First, most (50%) of the baryons are not in the hot ($T > 10^6$ K) ionized intergalactic gas, but in small colder ($T < 10^4$ K), but still ionized, gas clumps distributed in filaments. Second, the neutral gas at $z \sim 3$ is predominantly in the so-called 'damped Lyα absorbers' which are the focus of this thesis. Indeed, the amount of neutral hydrogen at high redshifts corresponds to today's stars (see Eq. 2.1 in § 2.3). Third, the amount of neutral hydrogen gas at $z \sim 3$ is comparable to today's stellar content . Finally, the amount of neutral hydrogen at $z \sim 3$ is larger (by a factor of two or three) than best estimate of the total hydrogen content seen locally and does not always appear to be associated with galaxies.

This striking contrast between state of the gas at $z > 1$ and today's situation, clearly shows that the universe has strongly evolved during that time. For instance, these differences have led some (e.g. Wolfe et al., 1986) to argue that, between then and now, neutral hydrogen has been transformed. One natural way to use hydrogen gas, is the convert it into stars.

APPENDIX B

DLAS: BASIC RELATIONSHIPS

B.1 The number density per unit redshift $\frac{dn}{dz}$

The number of absorbers per unit redshift $l(z) = \frac{dn}{dz}$ is the product of the cross-section $\sigma = \pi r^2$ and the comoving volume density n_o:

$$dn = l_{\rm DLA}(z)dz = n_o\, \pi r^2\, dS(z), \tag{B.1}$$

where $dS(z)$ is the comoving path element. Noting that $dS(z) = c\frac{dz}{H(z)}$ where $H(z) = H_o\, E(z)$ and that the cross-section is $\pi r^2 = \pi r_o^2(1+z)^2$ in comoving coordinates, we can rewrite $l(z)$ as

$$\begin{aligned} dn &= l_{\rm DLA}(z)dz = n_o \pi r_o^2 \frac{c}{H_o}\,(1+z)^2 \frac{dz}{E(z)} \\ &= n_o \pi r_o^2 \frac{c}{H_o} dX = l_o dX, \end{aligned} \tag{B.2}$$

where $l_o = n_o \pi r_o^2 \frac{c}{H_o}$ and $dX = (1+z)^2 \frac{dz}{E(z)} = (1+z)^2 \frac{H_o}{c} dS$ is referred to as the element of the 'absorption distance' ($X(z)$) (Bahcall & Peebles, 1969; Tytler, 1987). This 'absorption distance' is constructed such that absorption line systems with constant comoving density and constant physical cross-section will have a constant density per unit absorption distance $X(z)$.

From Eq. B.2, any intrinsic evolution of $l_{\rm DLA}(z)$ will come either from the evolution in the comoving volume density n_o or from the evolution of the cross-section $\sigma_o = \pi r_o$, i.e. from the factor $l_o(z)$ (Sargent et al., 1980). For instance, a matter dominated

Universe will have $E(z) \propto (1+z)^{\frac{3}{2}}$, and the number of absorbers per unit redshift will scale as $l_{\rm DLA}(z) \propto (1+z)^{0.5}$ if there is no intrinsic evolution.

Thus, $l_{\rm DLA}(z)$ is usually parameterized as $\propto (1+z)^{\gamma}$. However, as pointed out by Storrie-Lombardi et al. (2000), γ is poorly constrained due to differential evolution in the number density of DLAs with different column densities (already shown by Lanzetta, Wolfe, & Turnshek, 1995). Indeed, $l_o = \frac{{\rm d}n}{{\rm d}X} = n_o \pi r_o \frac{c}{H_o}$ can depend on redshift z (or X), and also on the column density $N_{\rm H\,I}$ via the number density $n_o(N_{\rm H\,I}, z)$ and the absorption cross-section $\sigma_o(N_{\rm H\,I}, z)$.

B.2 The column density distribution $f(N)$

From the previous section, it is natural to define the column density distribution $f(N, X)$ as the number of absorbers per unit column density (N), per unit absorption distance (X):

$$ {\rm d}n \;=\; f(N,X){\rm d}N{\rm d}X. \qquad \text{(B.3)} $$

This definition implies that the number density per unit X, $l_o(X)$, is $\int f(N,X){\rm d}N$. Thus, the column density distribution, $f(N)$, must be seen as an average over the range X, $X + \Delta X$: $f(N) =< f(N,X) >= \frac{1}{\Delta N}\frac{m}{\sum \Delta X}$, where m is the number of absorbers that have a column density in a bin $N, N+\Delta N$ and $\frac{1}{\Delta N}$ is the normalization factor (Tytler, 1987).

B.3 Total content in DLAs, $\Omega(z)$

From the column density distribution $f(N, X)$, one can infer the mean Hydrogen comoving density $\overline{n}_{\rm H\,I}(X)$ in DLAs via the mean column density $\overline{N}(X) = \int N f(N,X){\rm d}N\Delta X$ (e.g. see Tytler, 1987). Converting this mean number of atoms per unit physical area

to comoving area, $\overline{N}_c(X) = \frac{\overline{N}}{(1+z)^2}$, the mean number of atoms per unit comoving volume is $\overline{n}(X) = \frac{\overline{N}_c(X)}{\Delta S} = \frac{\int Nf(N,X)\mathrm{d}N\Delta S(1+z)^2\frac{H_o}{c}}{(1+z)^2\Delta S} = \frac{H_o}{c}\int Nf(N,X)\mathrm{d}N$.

Therefore, in terms of the critical density, ρ_c, the total neutral gas density contained in absorption systems with column density between $N_{\rm min}$ and $N_{\rm max}$ is:

$$\begin{aligned}\Omega_{\rm H\,I}(X) &= \frac{m_{\rm H\,I}}{\rho_c}\overline{n}(X) \\ &= \frac{H_o}{c}\frac{m_{\rm H\,I}}{\rho_c}\int_{N_{\rm min}}^{N_{\rm max}} Nf(N,X)\mathrm{d}N. \end{aligned} \tag{B.4}$$

and the total gas content in DLAs, $\Omega_{\rm DLA}(z)$, which is $\mu\Omega_{\rm H\,I}$ where μ accounts for the mass fraction of Helium, is given by:

$$\Omega_{\rm DLA}(z) = \frac{H_o}{c}\frac{\mu m_H}{\rho_c}\int_{N_{\rm min}}^{N_{\rm max}} Nf(N,X)\mathrm{d}N. \tag{B.5}$$

Thus, if $f(N,z) \propto N^{-\beta}$, then $\Omega_{\rm DLA}(z)$ is given by

$$\Omega_{\rm DLA}(z) = \frac{H_o}{c}\frac{\mu m_H}{\rho_c}\frac{B}{2-\beta}\left(N_{\rm max}^{2-\beta} - N_{\rm min}^{2-\beta}\right), \tag{B.6}$$

where B takes into account the normalization of $f(N)$. This shows that the integral is dominated by the absorbers with the largest column density, $N_{\rm max}$ if $\beta < 2$.

B.4 Luminosity evolution of l_o

We already mentioned that l_o can be a function of redshift or of column density N. But it can also be a function of luminosity L. Following Wolfe et al. (1986) and Tytler (1987), first let's assume that both the volume number density of absorbers (n_o) and the cross section (σ_o) are a function of the luminosity L, $n_o\sigma_o = \phi(L)\sigma(L)\mathrm{d}L$, and Eq. B.2 becomes:

$$\mathrm{d}n = \phi(L)\sigma(L)\frac{c}{H_o}\mathrm{d}L\mathrm{d}X. \tag{B.7}$$

Then, assuming a direct correspondence between the H I disk size ($R_{\rm H\,I}$), and the galaxy size R ($R_{\rm H\,I} = s \times R$, $s \simeq 1 - 1.2$), and a scaling relation between the galaxy size R and the luminosity L, $R/R^* = (L/L^*)^t$, known as the Holmberg relation for disk galaxies (Holmberg, 1975), one finds that the averaged cross-section is $\sigma(L) = \pi R_{\rm H\,I}^2/2 = \frac{\pi}{2} R^{*2} s^2 (L/L^*)^t$ (e.g. Wolfe et al., 1986), where the factor of 2 accounts for the reduced projected area of randomly inclined disks (Tytler, 1987). Therefore, the number of absorbers per unit redshift is:

$$\begin{aligned} \frac{{\rm d}n}{{\rm d}z} &= \int \phi(L)\sigma(L){\rm d}L \, \frac{c(1+z)^2}{H_o E(z)}, \\ &= \Gamma(1+\alpha+2t)\phi^* \, \frac{\pi}{2} R^{*2} \, s^2 \, \frac{c(1+z)^2}{H_o E(z)}, \\ &= K(t) \times \frac{\pi}{2} R^{*2} s^2, \end{aligned} \tag{B.8}$$

where $K(t) = \Gamma(1+\alpha+2t)\phi^* \frac{c(1+z)^2}{H_o E(z)}$. If one assumes a luminosity function, $\phi(L)$, one can solve this equation for $\frac{\pi}{2} R^{*2} \, s^2$, and use the Holmberg relation once more to find:

$$\begin{aligned} \sigma(L,t) &= \frac{\pi}{2} R^{*2} s^2 (L/L^*)^t \\ &= \frac{{\rm d}n}{{\rm d}z} \frac{1}{K(t)} \left(\frac{L}{L^*}\right)^t = \pi b(L,t)^2 \end{aligned} \tag{B.9}$$

which gives an impact parameter distribution as a function of luminosity and the scaling exponent t. This is the methodology followed by Fynbo, Møller, & Warren (1999) discussed in section 2.8.

APPENDIX C

K-CORRECTIONS

This matter is subject to lots of confusion. It has been described in the literature since Oke & Sandage (1968). In an attempt to clarify some of it, here is our approach to this very common problem.

C.1 Introduction of the problem

When one observes a galaxy at redshift z in a filter (say R), we measure its flux in that filter band m_R^{obs}. Now, one would like to know its *intrinsic* flux in some other filter (called Q here) if the galaxy were at redshift $z = 0$. In other words, given the observed magnitude m_R^{obs}, what is the intrinsic magnitude m_Q^{rest}?

The K-correction is defined as the difference between the two:

$$K_{QR}(z) = m_R^{obs} - m_Q^{rest}. \tag{C.1}$$

Three different effects are at play here:

- First, the two filter bands are sampling two different parts of the spectral energy distribution of the galaxy.
- Second, the 'effective' filter width is being squeezed with redshift.
- Third, the two magnitudes use two different zero-points, or reference flux densities.

C.2 Formalism

As we have defined in section 1.3, the *observed* magnitude is

$$m_R^{obs} = -2.5 \log \left[\frac{\int_0^\infty F_\lambda^{obs}(\lambda_o) R(\lambda_o) \mathrm{d}\lambda_o}{\int_0^\infty g_\lambda^R(\lambda_o) R(\lambda_o) \mathrm{d}\lambda_o} \right] \tag{C.2}$$

where λ_o is the observed wavelength, $F_\lambda^{obs}(\lambda_o)$ is the *observed* flux density, $g_\lambda^R(\lambda_o)$ is the standard-source flux density, and $R(\lambda_o)$ is the filter transmission curve.

Meanwhile, the *intrinsic* magnitude, or the magnitude in the galaxy *rest* frame, is

$$m_Q^{rest} = -2.5 \log \left[\frac{\int_0^\infty f_\lambda^{rest}(\lambda_e) Q(\lambda_e) \mathrm{d}\lambda_e}{\int_0^\infty g_\lambda^Q(\lambda_e) Q(\lambda_e) \mathrm{d}\lambda_e} \right] \tag{C.3}$$

where λ_e is the intrinsic wavelength, $f_\lambda^{rest}(\lambda_e)$ is the *intrinsic* flux density, $g_\lambda^Q(\lambda_e)$ is the standard-source flux density [1], and $Q(\lambda_e)$ is the filter transmission curve of interest.

Of course, the two reference frames are related by $\lambda_o = (1+z)\lambda_e$. Flux conservation (energy if using luminosities L_ν) implies $F_\lambda^{obs}(\lambda_o)\mathrm{d}\lambda_o = f_\lambda^{rest}(\lambda_e)\mathrm{d}\lambda_e$, thus

$$F_\lambda^{obs}(\lambda_o)\mathrm{d}\lambda_o = f_\lambda^{rest}\left(\frac{\lambda_o}{1+z}\right)\frac{\mathrm{d}\lambda_o}{1+z}. \tag{C.4}$$

So, m_R^{obs} (Eq.C.2) can be expressed in terms of the *rest-frame* SED f_λ^{rest}:

$$m_R^{obs} = -2.5 \log \left[\frac{\int_0^\infty f_\lambda^{rest}\left(\frac{\lambda_o}{1+z}\right) R(\lambda_o) \frac{\mathrm{d}\lambda_o}{1+z}}{\int_0^\infty g_\lambda^R(\lambda_o) R(\lambda_o) \mathrm{d}\lambda_o} \right]. \tag{C.5}$$

Using the definition of the K-correction (Eq. C.1), we then have

$$K_{QR}(z) = -2.5 \log \left[\frac{\int_0^\infty f_\lambda^{rest}\left(\frac{\lambda_o}{1+z}\right) R(\lambda_o) \frac{\mathrm{d}\lambda_o}{1+z} \times \int_0^\infty g_\lambda^Q(\lambda_e) Q(\lambda_e) \mathrm{d}\lambda_e}{\int_0^\infty f_\lambda^{rest}(\lambda_e) Q(\lambda_e) \mathrm{d}\lambda_e \times \int_0^\infty g_\lambda^R(\lambda_o) R(\lambda_o) \mathrm{d}\lambda_o} \right]. \tag{C.6}$$

[1] In order to be as general as possible, this need not be the same as g^R in Eq. C.2.

Eq. C.6 is similar to Eq. 12 in Hogg et al. (2002), except they have defined R and Q such that $\lambda R(\lambda)$ is equal to the transmission curve $R(\lambda)$ used here. A major difference, however, is that their f_λ refers to the observed flux density F_λ^{obs} which will depend on the redshift of each galaxy.

If one is interested in expressing Eq. C.6 in terms of f_ν and frequencies, one has to use the following transformations:

$$f_\nu \mathrm{d}\nu = f_\lambda \mathrm{d}\lambda \tag{C.7}$$

$$g_\nu \mathrm{d}\nu = g_\lambda \mathrm{d}\lambda \tag{C.8}$$

$$R(\nu)\mathrm{d}\nu = R(\lambda)\mathrm{d}\lambda \tag{C.9}$$

$$Q(\nu)\mathrm{d}\nu = Q(\lambda)\mathrm{d}\lambda. \tag{C.10}$$

i.e. $f_\nu \propto \lambda^2 f_\lambda$, etc. Assuming $f_\nu \propto \nu^\alpha$, and $f_\lambda \propto \lambda^\beta$ implies $\alpha = -(2+\beta)$.

C.3 Case $R = Q$

In the classical case where the two filters are the same, i.e. $R = Q$, and $g^R = g^Q$, Eq. C.6 is equivalent to the original formulation of Oke & Sandage (1968):

$$K_{RR}(z) = -2.5 \log \left[\frac{\int_0^\infty f_\lambda^{rest}(\frac{\lambda_o}{1+z}) R(\lambda_o) \frac{\mathrm{d}\lambda_o}{1+z}}{\int_0^\infty f_\lambda^{rest}(\lambda_e) R(\lambda_e) \mathrm{d}\lambda_e} \right]. \tag{C.11}$$

We checked Eq. C.11 against the published results of Coleman, Wu, & Weedman (1980) and found good agreement.

When one makes the following approximation (discussed further in section C.7) $\int f_\lambda R(\lambda)\mathrm{d}\lambda \simeq f_\lambda(\lambda_{eff}^R) W^{'R}$, where λ_{eff}^R is the effective wavelength[2] and $W^{'R}$ is the width of the filter R, and if $f_\lambda \propto \lambda^\beta$ then Eq. C.11 becomes

[2] $\lambda_{eff} = \int \lambda R(\lambda)\mathrm{d}\lambda / \int R(\lambda)\mathrm{d}\lambda$.

$$K_{RR}(z) = +(1+\beta)\cdot 2.5\log(1+z). \tag{C.12}$$

In this case, since the two filters are the same, only the first two effects mentioned in section C.1 are causing the K-correction. Indeed, the β term comes from the shape of the distribution while the additional term comes from the fact that the width of the filter in the rest frame is different than in the observed frame by a factor $1+z$.

C.4 Case $R \neq Q$

To get a feeling for Eq. C.6 in the case of two different filters, let's make the same approximation as above: the filter transmission curve is a square function, and $\int f_\lambda R(\lambda)\mathrm{d}\lambda \simeq f_\lambda(\lambda^R_{eff})W^R$. In the AB system, all flux measurements are made with respect to the constant $g_\nu = 3631$Jy where 1Jy=10^{-23} erg s^{-1} cm^{-2} Hz^{-1}, i.e. $\mathrm{AB}_{\nu_{eff}} = m^{obs}_{\mathrm{AB}}(R) \simeq -2.5\log[F^{obs}_\nu(\nu^R_{eff})(Jy)] - 48.60$. However, from the transformation Eq. C.8. g_λ is not constant, $g_\lambda \propto \lambda^{-2} g_\nu$! Therefore, if $f_\lambda \propto \lambda^\beta$, Eq. C.6 becomes:

$$\begin{aligned}
K_{QR}(z) &= -2.5\log\left[\frac{(\lambda^R_{eff}/(1+z))^\beta \cdot W^R/1+z}{(\lambda^Q_{eff})^\beta \cdot W^Q} \times \frac{(\lambda^Q_{eff})^{-2}\cdot W^Q}{(\lambda^R_{eff})^{-2}\cdot W^R}\right] \\
&= -2.5\log\left[\left(\frac{\lambda^R_{eff}}{\lambda^Q_{eff}}\right)^{2+\beta}\right] + (1+\beta)\cdot 2.5\log(1+z) \qquad \text{(C.13)} \\
&= \left[-\beta\cdot 2.5\log[\lambda^R_{eff}/\lambda^Q_{eff}] + (1+\beta)\cdot 2.5\log(1+z)\right] - 5\log[\lambda^R_{eff}/\lambda^Q_{eff}] \\
&= \mathrm{SED_{corr}} + \mathrm{ZP_{corr}}
\end{aligned}$$

which is the sum of two terms. The first term ($\mathrm{SED_{corr}}$) is due to the fact that we are sampling two different regions of the observed galaxy spectrum and takes into account the first two effects mentioned in section C.1. The second term ($\mathrm{ZP_{corr}}$) is due to the fact that the two filters have different zeropoints. For example, the MOSAIC filters I and U have $\lambda^I_{eff}/\lambda^U_{eff} = 8000\text{Å}/3600\text{Å} = 2.22$, so in the AB system, the K-correction $K_{UI}(z=3)$ is:

$$
\begin{aligned}
K_{UI}(3) &= [1.73 - 1.51] - 1.73 = -1.51 && \text{for } \beta = -2 \\
&= [\ 0.87 + 0\ ] - 1.73 = -0.86 && \text{for } \beta = -1 \\
&= [\ 0 + 1.51\ ] - 1.73 = -0.21 && \text{for } \beta = 0 \\
&= [\ -0.87 + 3.01\ ] - 1.73 = 0.41 && \text{for } \beta = 1
\end{aligned}
$$

C.5 Case matching filters

Even if one has a B filter (m^{rest}) that matches another one at a given redshift, say the I filter at $z = 0.8$. i.e. $\lambda^I_{eff}/\lambda^B_{eff} \simeq 1 + 0.8$, the K-correction (still in the AB system) is independent of β, but can be significant:

$$K_{BI}(z) = (1 + \beta - 2 - \beta) \cdot 2.5 \log(1 + z) = -2.5 \log(1 + z) \qquad \text{(C.14)}$$

which at $z = 0.8$ equals to -0.64.

C.6 Remarks

-The 'filter curve' $R(\lambda)$ should really include the filter transmission, the CCD efficiency, the atmosphere transmission and the telescope optics.

-The description presented above on the 'k-correction' of observed fluxes does not include the effect of absorption of the intergalactic medium. This effect can be significant at high redshifts.

C.7 Note on the approximation $\int f_\lambda R(\lambda) \mathrm{d}\lambda \simeq f_\lambda(\lambda^R_{eff}) W^R$

Here, the 'width' of the filter is defined by: $W_R = \int R(\lambda) \mathrm{d}\lambda$, and it should be called 'effective width' as it accounts for the entire shape of the filter. That said, in the case when the width of the filter W_R is much less than λ^R, $(1 + W/\lambda)^\beta \simeq (1 + \beta W/\lambda)$, and the first order approximation $\int f_\lambda R(\lambda) \mathrm{d}\lambda \simeq f_\lambda(\lambda^R_{eff}) W^R$ is valid irrespectively of the shape of the filter and of the slope of the intrinsic SED f^{rest}_λ. When $W_o \ll \lambda_o$ is not verified, then the approximation breaks down.

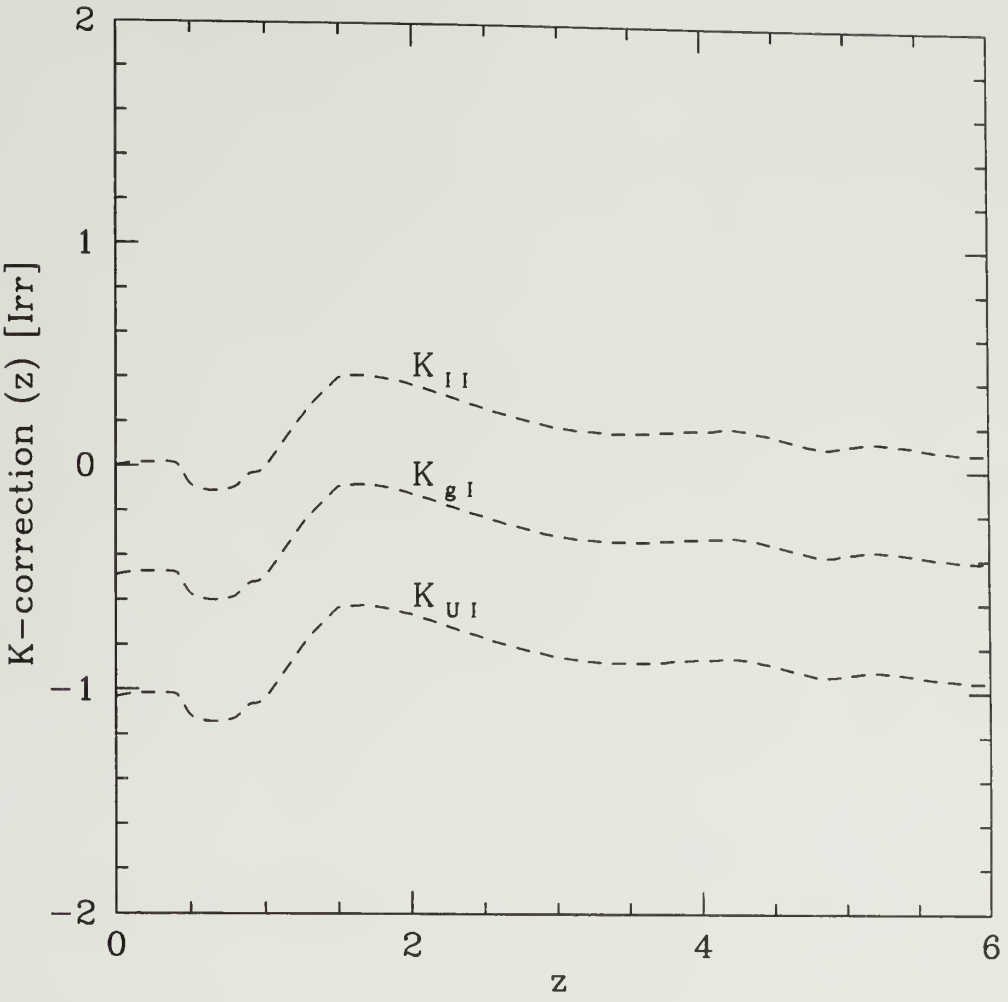

Figure C.1 K-correction as a function of redshift for a galaxy of type Irr, using the Irr SED of Coleman, Wu, & Weedman (1980) in the AB magnitude system. The three curves show K_{II}, K_{gI}, and K_{UI}, the transformation between the observed filter, I, and three other filters, the I, the Sloan g, and U, respectively. The fact that K_{II} is close to zero at all redshifts follows from Eq. C.12 because the Irr SED has a slope of approximately $\beta \simeq -1$. The reason why the K-correction becomes more negative is due to the combination of the two terms in Eq. C.13. $\mathrm{SED}_{\mathrm{corr}} \propto -\beta 2.5 \log[\lambda^{\mathrm{I}}/\lambda]$, which is positive since $\beta \simeq -1$. On the other hand, $\mathrm{ZP}_{\mathrm{corr}} \propto -5 \log[\lambda^{\mathrm{I}}/\lambda]$ since $g_\lambda \propto \lambda^{-2}$ and becomes negative more rapidly.

C.8 Absolute magnitudes

Another important question to compare measurement in a meaningful way: Given m_R, what is the absolute magnitude M_Q?

Given the luminosity distance D_L of a galaxy, its absolute magnitude M is by definition: $M_R = -2.5 \log \left(\frac{L_R^{obs}}{4\pi(10 \text{ pc})^2} \right) = m_R - DM(z)$, where L_R^{obs} is the observed luminosity density and is equal to observed flux density F_R^{obs} times $\times 4\pi D_L^2$. The 'distance modulus' is as usual $DM(z) = 5 \log(D_L) - 5$.

Therefore, using K_{QR} given by Equation C.6, M_Q is

$$M_Q^{rest} = m_R^{obs} - DM(z) - K_{QR}(z) \tag{C.15}$$

since $m_Q = M_Q + DM(z)$.

As a check, let's compute $M_{\text{AB}}^{rest}(\nu_e)$ in the case $\lambda_e = \frac{\lambda_o}{1+z}$, i.e. with matching filters, without referring to the K-correction definition (Eq. C.6). Using the fact that νL_ν =cst (Eq. C.7),

$$\begin{aligned}
M_{\text{AB},\nu}^{rest}\left(\frac{\lambda_o}{1+z}\right) &= M_{\text{AB},\nu}^{rest}(\nu_e) = -2.5 \log \left[\frac{L_\nu^{rest}(\nu_e)}{4\pi(10 \text{ pc})^2} \right] - 48.60 \\
&= -2.5 \log \left[\frac{\nu_o L_{\nu_o}^{obs}}{\nu_e 4\pi(10 \text{ pc})^2} \right] - 48.60 \\
&= -2.5 \log \left[\frac{4\pi D_L^2 \times F_\nu^{obs}(\nu_o)}{4\pi(10 \text{ pc})^2 \times (1+z)} \right] - 48.60 \\
&= 5 - 5 \log(D_L) - 2.5 \log[F_\nu^{obs}(\nu_o)] - 48.60 + 2.5 \log(1+z) \\
&= m_{\text{AB}}^{obs}(\nu_o) - DM(z) + 2.5 \log(1+z). \qquad \text{(C.16)}
\end{aligned}$$

Eq. C.16 is equivalent to Eq. C.15 using the K-correction for matching filters Eq. C.14, which completes the check.

Another example: If one wants to compute the absolute magnitude $M_{\text{AB}}(4400)$ at the same wavelength for a sample of objects at various redshifts z, and all observed

through the same R filter (λ_{eff}^{R}), then one should use Eq. C.15 with $K_{QR}(z)$ from Eq. C.13 (assuming $f_\lambda \propto \lambda^\beta$). The result is:

$$\begin{aligned} M_{\mathrm{AB},\nu}^{rest}(4400\mathrm{A}) &= m_{\mathrm{AB},\nu}^{obs}(\nu^R) - DM(z) + (2+\beta)\cdot 2.5\log\left(\frac{\lambda^R}{4400}\right) - \\ &\quad (2+\beta)\cdot 2.5\log(1+z) + 2.5\log(1+z) \\ &= M_{\mathrm{AB},\nu}^{rest}\left(\frac{\lambda^R}{1+z}\right) + (2+\beta)\cdot 2.5\log\left(\frac{\lambda^R/(1+z)}{4400}\right) \end{aligned} \quad \text{(C.17)}$$

where we used Eq. C.16 for $M_{\mathrm{AB}}^{rest}\left(\frac{\lambda^R}{1+z}\right)$. Eq. C.17 is Eq. 10 in Schmidt, Schneider, & Gunn (1995) using $\alpha = -(2+\beta)$ (Eq. C.7).

APPENDIX D

SOURCE EXTRACTION

Here are listed the parameters used by the detection algorithm SExtractor (see section 4.5).

```
DETECT_MINAREA  5       # minimum number of pixels above threshold
DETECT_THRESH   1.5     # <sigmas> or <threshold>,<ZP> in mag.arcsec-2
ANALYSIS_THRESH 1.5     # <sigmas> or <threshold>,<ZP> in mag.arcsec-2
FILTER          Y       # apply filter for detection ("Y" or "N")?
FILTER_NAME     gauss_2.0_3x3.conv
        # name of the file containing the filter
DEBLEND_NTHRESH 32      # Number of deblending sub-thresholds
DEBLEND_MINCONT 0.0001  # Minimum contrast parameter for deblending
CLEAN           Y       # Clean spurious detections? (Y or N)?
CLEAN_PARAM     1.0     # Cleaning efficiency
#------------------------------ Photometry -------------------------
GAIN            3.0     # detector gain in e-/ADU.
PIXEL_SCALE     .258    # size of pixel in arcsec (0=use FITS WCS info).
#------------------------- Star/Galaxy Separation -------------------
SEEING_FWHM     1.2     # stellar FWHM in arcsec
STARNNW_NAME    default.nnw # Neural-Network_Weight table filename
#------------------------------ Background -------------------------
BACK_SIZE       64      # Background mesh: <size> or <width>,<height>
BACK_FILTERSIZE 3       # Background filter: <size> or <width>,<height>
```

```
BACKPHOTO_TYPE  LOCAL  # can be "GLOBAL" or "LOCAL" (*)
BACKPHOTO_THICK 24     # thickness of the background LOCAL annulus (*)
#------------------------ Check Image ----------------------------
CHECKIMAGE_TYPE BACKGROUND  # can be one of "NONE", "BACKGROUND",
                        # "MINIBACKGROUND", "-BACKGROUND", "OBJECTS",
                        # "-OBJECTS", "SEGMENTATION", "APERTURES",
                        # or "FILTERED" (*)
CHECKIMAGE_NAME object.fits # Filename for the check-image (*)
```

BIBLIOGRAPHY

Adelberger, K. L., Steidel, C. C., Giavalisco, M., Dickinson, M., Pettini, M., & Kellogg, M. 1998, ApJ, 505 18
A Counts-in-Cells Analysis Of Lyman-break Galaxies At Redshift $z \sim 3$

Adelberger, K. L., Steidel, C. C., Shapley A. E., & Pettini M. 2003, ApJ, 584, 45
Galaxies and Intergalactic Matter at Redshift z 3: Overview

Alonso-Herrero, A., Aragón-Salamanca, A., Zamorano, J., & Rego, M. 1996, MNRAS, 278, 417
Star formation history in a sample of starburst galaxies

Anantharamaiah, K. R., Viallefond, F., Mohan, N. R., Goss, W. M., & Zhao, J. H. 2000, ApJ, 537, 613
Starburst in the Ultra-luminous Galaxy Arp 220: Constraints from Observations of Radio Recombination Lines and Continuum

Asimov, I. 1996, in 'The final magic fantasy collection', Harper Collins Publishers, 1996

Bachall, J. N., & Peebles, P. J. E. 1969, 156, 7L
Statistical Tests for the Origin of Absorption Lines Observed in Quasi-Stellar Sources

Bechtold, J., Elston, R., Yee, H. K. C., Ellington, E., & Cutri, R. M. 1998, in 'The Young Universe: Galaxy Formation and Evolution at Intermediate and High Redshift', ed. S. D'Odorico, A. Fontana, and E. Giallongo, ASP Conf. Series, (PASP), 146, 241

Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100,32
Measures of location and scale for velocities in clusters of galaxies - A robust approach

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. 2003, ApJ, submitted (astro-ph/0302543)
The Optical and Near-Infrared Properties of Galaxies: I. Luminosity and Stellar Mass Functions

Bertin, E., & Arnouts, S. 1996, A&A, 117, 39
SExtractor: Software for source extraction

Bernstein, G. M. 1994, ApJ, 424, 577
The variance of correlation function estimates

Benítez, N. 2000, ApJ, 536, 571
Bayesian Photometric Redshift Estimation

Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Spergel, D. N., Tucker, G. S., Wollack, E., Wright, E. L., Barnes, C., Greason, M. R., Hill, R. S., Komatsu, E., Nolta, M. R., Odegard, N., Peiris, H. V., Verde, L., & Weiland, J. L. 2003, AJ, submitted (astro-ph/0302207)
First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results

Blanton, M. R., Dalcanton, J., Eisenstein, D., Loveday, J., Strauss, M. A., SubbaRao, M., Weinberg, D. H., Anderson, J. E., Jr., Annis, J., Bahcall, N. A. and the SDSS collaboration 2001, AJ, 121, 2358
The Luminosity Function of Galaxies in SDSS Commissioning Data

Boesgaard, A. M., & Steigman, G. 1985, ARA&A, 23, 319
Big bang nucleosynthesis - Theories and observations

Boissé, P., Le Brun, V., Bergeron, J., & Deharveng, J-M. 1998, A&A, 333, 841
A HST spectroscopic study of QSOs with intermediate redshift damped Lyalpha systems

Bogart, Richard S., Wagoner, R. V. 1973, ApJ, 181, 609
Clustering Effects among Clusters of Galaxies and Quasi-Stellar Sources

Boissier, S., Péroux, C., & Pettini, M. 2003, MNRAS, 338, 131
Damped Lyman-alpha: systems and disc galaxies: number density, column density distribution and gas density

Bolzonella, M., Miralles, J-M., & Pelló, R. 2000, A&A, 363, 476
Photometric redshifts based on standard SED fitting procedures

Bouché, N., Lowenthal, J. D., Charlton, J. C., Bershady, M. A., Churchill, C. W., & Steidel, C. C. 2001, A&A550, 585
H-alpha; Imaging with Hubble Space Telescope-NICMOS of an Elusive Damped Ly-alpha; Cloud at $z = 0.6$

Bouché, N., & Lowenthal, J. D. 2003, ApJ, accepted (astro-ph/0307033)
Clustering of Galaxies at $z \sim 3$ around the probable Damped Ly-alpha Absorber towards QSO APM 08279+5255

Brown, M. J. I., Dey, A., Jannuzi, B. T., & Lauer, T. R 2003, ApJ, accepted for publication (astro-ph/0306128)
Red galaxies clustering in the NOAO Deep Wide Field Survey

Bruzual, G., & Charlot, S., 1993, ApJ, 405, 538
Spectral evolution of stellar populations using isochrone synthesis

Budavári, T., Connoly, A. J., Sazlay, A. S., et al 2003, ApJ, submitted (astro-ph/0305603)
Angular Clustering with photometric redshifts

Bunker, A. J., Warren, S. J., Clements, D. L., Williger, G. M., & Hewett, P. C. 1999, MNRAS, 309, 875
Limits on the star formation rates of $z > 2$ damped Ly-alpha: systems from H-alpha spectroscopy

Burles, S., & Tytler, D. 1998, ApJ, 507, 732
The Deuterium Abundance toward QSO 1009+2956

Carilli, C. L., Lane, Wendy, de Bruyn, A. G., Braun, R., Miley, G. K.
Redshift H1 *21 cm Line Observations of Damped Ly alpha Absorption Systems*

Carilli, C. L., Bertoldi, F., Rupen, M. P., Fan, X., Strauss, M, A., Menten, K. M., Kreysa, E., Schneider, D, P., Bertarini, A., Yun, M. S., Zylka, R. 2001, ApJ, 555, 625
A 250 GHZ survey of High Redshift Quasars from the Sloan Digital Sky Survey

Churchill, C. W., Mellon, R. R., Charlton, J. C., Jannuzi, B. T., Kirhakos, S., Steidel, C. C., and Schneider, D. P. 1999, ApJ, 519, L43
The C IV Absorption-Mg II Kinematics Connection in ¡Z¿ 0.7 Galaxies

Churchill, C. W., Mellon, R. R., Charlton, J. C., Jannuzi, B. T., Steidel, C. C., & Schneider, D. P. 2000, ApJ, 543, 577
Low- and High-Ionization Absorption Properties of Mg II Absorptionselected Galaxies at Intermediate Redshifts. II. Taxonomy, Kinematics, and Galaxies

Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koorneef, J., & Storchi-Bergmann, T. 2000, ApJ, 533, 682
The Dust Content and Opacity of Actively Star-forming Galaxies

Coleman, G. D., Wu, C-C., & Weedman, D. W. 1980, ApJS, 43, 493
Colors and magnitudes predicted for high redshift galaxies

Cohen, J. G., Hogg, David W., Blandford, R., Cowie, L. L., Hu, E., Songaila, A., Shopbell, P., & Richberg, K. 2000, AJ, 538, 29
Caltech Faint Galaxy Redshift Survey. X. A Redshift Survey in the Region of the Hubble Deep Field North

Connolly, A. J., Csabai, I., Szalay, A. S., Koo, D. C., Kron, R. G., Munn, J. A. 1995, AJ, 110, 2655
Slicing Through Multicolor Space: Galaxy Redshifts from Broadband Photometry

Curran, S. J., Webb, J. K., Murphy, M. T., Bandiera, R., Corbelli, E. , & Flambaum, V.V. 2002, PASP, 19, 455
A catalog of Damped Lyman Alpha absorption systems and radio flux densities of the backgroand quasars

Dav, R., Hernquist, L., Katz, N., & Weinberg, D. H. 1999, 511, 521
The Low-Redshift LYalpha Forest in Cold Dark Matter Cosmologies

d'Odorico, V., Petitjean, P., & Cristiani, S. 2002, A&A, 390, 13
High matter density peaks from UVES observations of QSO pairs: Correlation properties and chemical abundances

Djorgovski, S. G., Pahre, M. A., Bechtold, J., & Elston, R. 1996, Nature, 382, 234
Identification of a galaxy responsible for a high-redshift Lyman-alpha absorption system

Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1988, MNRAS, 235, 175
Gravitational clustering from scale free initial conditions

Efstathiou, G., Schaye, J., & Theuns, T. 2000, RSPTA, 358, 2049
Lyα *absorption systems and the intergalactic medium*

Eisenstein, D. J. 2002, preprint (astro-ph/0212084)
Deprojecting Densities from Angular Cross-Correlations

Ellingson, E., Yee, H. K., & Green, R. F. 1991, ApJ, 371, 49
Quasars and active galactic nuclei in rich environments

Ellison, S. L., Lewis, G. F., Pettini, M., Sargent, W. L. W., Chaffee, F. H., Folts, C. B., Rauch, M., & Erwin, M. J. 1999a, PASP, 111, 946
KECK HIRES Spectroscopy of APM 08279+525

Ellison, S. L., Lewis, G. F., Pettini M., Chaffee, F. H., & Irwin, M. J. 1999, ApJ, 520, 456
HIRES Spectroscopy of APM 08279+5255: Metal Abundances in the Ly-alpha Forest

Ellison, S. L., Yan, L., Hook, I. M., Pettini, M., Wall, J. V., & Shaver, P. 2001, A&A, 379, 393
The CORALS Survey I: New Estimates of the Number Density and Gas Content of Damped Lyman Alpha Systems

Ellison, S. L., Pettini, M., Steidel, C. C., & Shapley, A. E. 2001, ApJ, 549, 770
An Imaging and Spectroscopic Study of the zabs=3.38639 Damped Ly-alpha; System in Q0201+1120: Clues to Star Formation at High Redshift

Fan, X. & SDSS collaboration 1999, AJ, 118,1
High-redshift Quasar Found in Sloan Digital Sky Survey Commissioning Data

Fernández-Soto, A., Lanzetta K.M., Chen H-W., Pascarelle S. M. & Yahata, N. 2001, ApJS, 135, 41
On the Compared Accuracy and Reliability of Spectroscopic and Photometric Redshift Measurements

Foucaud, S., McCracken, H. J., Le Fèvre, O., Arnouts, S., Brodwin, M., Lilly, S. J., Crampton, D., & Mellier, Y. 2003, A&A, (astro-ph/0306585)
The Canada-France deep fields survey-II: Lyman-break galaxies and galaxy clustering at $z \sim 3$.

Francis, P. J., & Hewett, P. C. 1993, AJ, 105, 1633
Superclustering of damped Lyman-alpha absorption systems at redshifts above two

Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, ApJ, 503, 518
The Cosmic Baryon Budget

Fynbo, J. P. U., Møller, P., & Warren, S. J. 1999, MNRAS, 305, 849
Extended Lyα emission from a damped Ly absorber at $z = 1.93$, and the relation between damped Ly absorbers and Lyman-break galaxies

Fynbo, J. P. U., Burud, I., & Møller, P. 2000, A&A, 358, 88
The sources of extended continuum emission towards Q0151+048A: the host galaxy and the damped Lyalpha absorber

Fynbo, J. P. U., Ledoux, C, Mooller, P., Thomsen, B., & Burud, I. 2003, A&A, (submitted, astro-ph/0305282)
The Building the Bridge survey for z=3 Ly-alpha emitting galaxies I: method and first results

Gardner, J. P., Katz, N., Hernquist, L., & Weinberg, D. H. 2001, ApJ, 559, 131
Simulations of Damped Ly-alpha; and Lyman Limit Absorbers in Different Cosmologies: Implications for Structure Formation at High Redshift

Gawiser, E., Wolfe, A. M., Prochaska, J. X., Lanzetta, K. M., Yahata, N., & Quirrenbach, A. 2001, ApJ, 563, 628
First Investigation of the Clustering Environment of Damped Ly-alpha; Absorbers at $z = 4$

Giavalisco, M., & Dickinson, M. 2001, ApJ, 550, 177
Clustering Segregation with Ultraviolet Luminosity in Lyman Break Galaxies at $z \sim 3$ and Its Implications

Giavalisco, M. 2002, ARA&A, 40, 579
Lyman-Break Galaxies

Giovanelli, R., & Haynes, M. P. 1989, ApJ, 346, L5
A protogalaxy in the local supercluster

Groth, E. J., & Peebles, P. J. E. 1977. ApJ, 217, 385
Statistical analysis of catalogs of extragalactic objects. VII - Two- and three-point correlation functions for the high-resolution Shane-Wirtanen catalog of galaxies

Haardt, F., & Madau, P. 1996, ApJ, 461, 20
Radiative Transfer in a Clumpy Universe. II. The Ultraviolet Extragalactic Background

Haehnelt, M. G., Steinmetz, M., & Rauch, M. 1998 ApJ, 495, 647
Damped Ly alpha Absorber at High Redshift: Large Disks or Galactic Building Blocks?

Hamilton, A. J. S., Matthews, A., Kumar, P., Lu, E. 1991, ApJ, 374, 1
Reconstructing the primordial spectrum of fluctuations of the universe from the observed nonlinear clustering of galaxies

Hernquist, L. 1987, ApJS, 64, 715
Performance characteristics of tree codes

Hibbard, J. E., & Yun, M. S. 1999, ApJ, 118, 162
A 180 Kiloparsec Tidal Tail in the Luminous Infrared Merger ARP 299

Hogan, C. J. 1999, astro-ph/9912107
Where the Baryons Are?

Hogg, D. W., Baldry, I. K., Blanton, M. R., & Eisenstein, D. J. 2002, astro-ph/0210394
The K-correction

Holmberg, E. B. 1975, in "Galaxies and the Universe", eds A. Sandage, M. Sandage, & J. Kristian, Chicago University Press, p151

Hubble, E. 1925, ApJ, 62, 409
NGC 6822, a remote stellar system

Hyland, A. R., & Allen, D. A. 1982, MNRAS, 199, 943
An infrared study of quasars

Ibata R. A. Lewis G. F., Irwim M. J., Lehar J., & Tottens, E. 1999, AJ, 118, 1922
NICMOS and VLA Observations of the Gravitationally Lensed Ultraluminous BAL Quasar APM 08279+5255: Detection of a Third Image

Irwin, M. J., Ibata, R. A., Lewis, G. F., & Totten, E. J. 1998, ApJ, 505, 529
APM 08279+5255: an Ultraluminous Broad Absorption Line Quasar at a Redshift Z = 3.87

Jacoby G. H., Liang M., Vaughnn, D., Reed, R., & Armandroff, T. 1998, SPIE 3355, 721
New wide-field corrector for the Kitt Peak Mayall 4-m telescope

Jarrett, T. H., Chester, T., Cutri, R., Schneider, S., Skrutski, M., & Huchra, J. P. 2000, AJ, 119, 2498
2MASS Extended Source Catalog: Overview and Algorithms

Jannuzi B. T., Claver, J., & Valdes, F. 2000, available at http://www.noao.edu/noao/noaodeep/ReductionOpt/frames.html
The NOAO Deep Wide-Field Survey MOSAIC Data Reductions

Jimenez, R., Bowen, D. V., & Matteucci, F. 1999, ApJ, 514, 83L
On the Origin of Damped LYalpha Systems: A Case for Low Surface Brightness Galaxies?

Kaiser, N. 1984, ApJ, 284, 9L
On the spatial correlations of Abell clusters

Katz, N., Weinberg, D. H., Hernquist, L., & Miralda-Escudé, J. 1996a, ApJ, 457, 57L
Damped Lyman-Alpha and Lyman Limit Absorbers in the Cold Dark Matter Model

Katz, N., Weinberg, D. H., Hernquist, L., 1996b, ApJS, 105, 19
Cosmological Simulations with TreeSPH

Katz, N., Hernquist, L., & Weinberg, D. H. 1999 ApJ, 523, 463
The Clustering of High-Redshift Galaxies in the Cold Dark Matter Scenario

Kanekar, N., & Chengalur, J. 2003
A deep search for 21-cm absorption in high redshift damped Lyman-alpha systems

Kennefick, J. D., Djorgovski, S. G., & Meylan, G. 1996, AJ, 111, 1816
A Multicolor CCD Survey for Faint $z > 4$ Quasar

Kennefick, J. D., Djorgovski, S. G., de Carvalho, R. R. 1995, AJ, 110, 2553
The Luminosity Function of $z \gtrsim 4$ Quasars from the Second Palomar Sky Survey

Kennicutt, R. C. Jr 1983, ApJ, 272, 54
The rate of star formation in normal disk galaxies

Kennicutt, R. C. Jr 1998, ApJ, 498, 541
The Global Schmidt Law in Star-forming Galaxies

Khersonsky, V. K., & Turnshek, D. A. 1996, ApJ, 471, 657
Damped LY alpha Absorbers as Tracers of the Evolution of the Mass Spectrum of Interstellar Matter in the Galaxy Population

Kinney, A. L., Bohlin, R. C., Calzetti, D., Panagia, N., & Wyse, R. F. G. 1993, ApJS, 86, 5
An atlas of ultraviolet spectra of star-forming galaxies

Kinney, A. L., Calzetti, D., Bohlin, R. C., McQuade, K., Storchi-Bergmann, T., Schmitt, H. R.1996, ApJ, 467, 48
Template Ultraviolet to Near-Infrared Spectra of Star-forming Galaxies and the Application to K-Corrections

Kobayashi, N., Terada, H., Goto, M., & Tokunaga, A. 2002, ApJ, 569, 676
MgII Absorption Lines in $z = 2.974$ Damped Lyman-alpha System toward Gravitationally Lensed QSO APM 08279+5255: Detection of Small-scale Structure in MgII Absorbing Clouds

Kulkarni, V. P., Hill, J. M., Schneider, G., Weymann, R. J., Storrie-Lombardi, L. J., Rieke, M. J., Thompson, R. I., & Jannuzi, B. T. 2000, ApJ, 536, 36
NICMOS Imaging of the Damped Ly-alpha; Absorber at $z = 1.89$ toward LBQS 1210+1731: Constraints on Size and Star Formation Rate

Kulkarni, V. P., & Fall, S. M. 2002, ApJ, 580, 732
Metallicity Evolution of Damped Ly Galaxies

Landolt, A. 1992, AJ, 104, 340
UBVRI photometric standard stars in the magnitude range 11.5-16.0 around the celestial equator

Landy, S. D., & Szalay, A. 1993, ApJ, 412, 64
Bias and variance of angular correlation functions

Lane, W., Smette, A., Briggs, F., Rao, S., Turnshek, D., & Meylan, G. 1998, AJ, 116, 26
Detection of Warm and Cold Phases of the Neutral ISM in a Damped Ly-alpha Absorber

Lane, W. M., Briggs, F. H., & Smette, A. 2000, ApJ, 532, 146
Detection of Warm and Cold Phases of the Neutral ISM in a Damped Ly-alpha: Absorber

Lanzetta, K. M., McMahon, R. G., Wolfe, A. M., Turnshek, D. A., Hazard, C., & Lu, L. 1991, ApJS, 77, 1
A new spectroscopic survey for damped Ly-alpha absorption lines from high-redshift galaxies

Lanzetta, K. M., Wolfe, A. M., & Turnshek, D. A. 1995, ApJ, 440, 435
The IUE Survey for Damped Lyman- alpha and Lyman-Limit Absorption Systems: Evolution of the Gaseous Content of the Universe

Lanzetta K. M., Yahil A., & Fernández-Soto A. 1996, Nature, 381, 759
Star-forming galaxies at very high redshifts

Leitherer, C., Schaerer D., Goldader J. D., Delgado, R.M.G., Robert, C., Kune D. F., de Mello D. F., Devost D., & Heckman T.M. 1999, ApJS, 123, 3
Starburst99: Synthesis Models for Galaxies with Active Star Formation

Le Brun, V., Bergeron, J., Boissé, P., & Deharveng, J. M. 1997, A&A, 321, 733
The nature of intermediate-redshift damped Ly-alpha absorbers

Ledoux, C., Petitjean, P., Bergeron, J., Wampler, E. J., & Srianand, R. 1998, A&A, 337, 51
On the kinematics of damped Lyman-alpha systems

Ledoux, C., Theodore, B., Petitjean, P., Bremer, M. N., Lewis, G. F., Ibata, R. A., Irwin, M. J., & Totten, E. J. 1998, A&A, 339, 77L
Adaptive optics imaging and integral field spectroscopy of APM 08279+5255: Evidence for gravitational lensing

Ledoux, C., Petitjean, P., & Srianand, R. 2003, MNRAS, submitted
The VLT-UVES survey for H2 in DLA systems

Lewis, G. F., Chapman, S. C., Ibata, R. A., Irwin, M. J., & Totten, E. J. 1998, ApJ, 505, 1L
Submillimeter Observations of the Ultraluminous Broad Absorption Line Quasar APM 08279+5255

Lewis, G. F., Robb R. M., & Ibata R. A. 1999, PASP, 111, 1503
Photometric Monitoring of the Gravitationally Lensed Ultraluminous Broad Absorption Line Quasar APM 08279+5255

Lewis, G. F., Ibata, R. A., Ellison, S. L., Aracil, B., Petitjean P., Pettini, M. & Srianand, R. 2002, MNRAS, 334, L7
Spatially resolved STIS spectra of the gravitationally lensed broad absorption line quasar APM08279+5255: the nature of component C and evidence for microlensing

Lilly, S. J., Le Fevre, O., Crampton, D., Hammer, F., & Tresse, L. 1995, ApJ, 455, 50
The Canada-France Redshift Survey. I. Introduction to the Survey, Photometric Catalogs, and Surface Brightness Selection Effects

Longair, M. S. 1998, in "Galaxy formation", eds Springer-Verlag

Lowenthal, J. D., Hogan, C. J., Green, R. F., Caulet, A., Woodgate, B. E., Brown, L., & Foltz, C. B. 1991, ApJ, 377, 73L
Discovery of a Ly-alpha galaxy near a damped Ly-alpha absorber at Z = 2.3

Lowenthal, J. D., Hogan, C. J., Green, R. F., Woodgate, B. E., Caulet, A., Brown, L., & Bechtold, J. 1995, ApJ, 451, 484
Imaging and Spectroscopy of Damped LY alpha Quasi-stellar Object Absorption-Line Clouds

Lowenthal, J. D., Koo, D. C., Guzman, R., Gallego, J., Phillips, A. C., Faber, S. M., Vogt, N. P., Illingworth, G. D., & Gronwall, C. 1997, ApJ, 481, 673
Keck Spectroscopy of Redshift Z approximately 3 Galaxies in the Hubble Deep Field

Lucy, L. 1977, AJ, 82, 1013
A numerical approach to the testing of the fission hypothesis

Madau, P. 1995, ApJ, 441, 18
Radiative transfer in a clumpy universe: The colors of high-redshift galaxies

Maller, A. H., Prochaska, J. X., Sommerville, R. S., & Primack, J. R. 2000, in *Clustering at high redshift*, ASP vol. 200, eds A. Mazure
Optical Counterparts to Damped Lyman-alpha Systems

Maller, A. H., McIntosh, D. H., Katz, N., & Weinberg, M. 2003, astro-ph/0304005.
The galaxy angular correlation functions and power spectrum from the two micron all sky survey

Matteucci, F., Molaro, P., Vladilo, G. 1997, å, 321, 45
Chemical evolution of damped Lyα *systems*

Mannucci, F., Thompson, D., Beckwith, S. V. W., & Williger, G. M. 1998, ApJ, 501, 11L
Infrared Emission-Line Galaxies Associated with Dampened Ly alpha and Strong Metal Absorber Redshifts

Massarotti, M., Iovino, A., Buzzoni, A., & Valls-Gabaud, D. 2001, A&A, 380, 425
New insights on the accuracy of photometric redshift measurements

McDonald, P., & Miralda-Escudé, J. 1999, ApJ, 519, 486
Galaxy Formation and the Kinematics of Damped LYalpha Systems

McMahon, R. G., Irwin, M. J., Giovanelli, R., Haynes, M. P., Wolfe, A. M., & Hazard, C. 1990, ApJ, 359, 302
An optical counterpart of the giant H I *cloud H I 1225+0146*

McMahon, R. G., White, R. L., Helfand, D. J., Becker, R. H. 2002, ApJS, 143, 1
Optical Counterparts for 70,000 Radio Sources: APM Identifications for the FIRST Radio

McLeod, B. in "1997 HST Calibration Workshop", ed. S. Casertano et al., p. 281

Mo, H. J., Jing, Y. P., & Boerner, G. 1992, ApJ, 392, 452
On the error estimates of correlation functions

Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347
Clustering of dark matter haloes

Mo, H. J., Jing, Y. P., & Boerner, G. 1997, MNRAS, 286, 979
Analytical approximations to low order statistics of dark matter distributions

Mo, H. J., Mao, S., & White, S. D. M. 1999, MNRAS, 304, 175
The structure and clustering of Lyman-break galaxies

Møller, P., Warren, S. J., Fall, S. M., Fynbo, J. U., & Jakobsen, P. 2002, ApJ, 574, 51
Are High-Redshift Damped Ly-alpha; Galaxies Lyman Break Galaxies?

Morton, D. C. 1974, ApJ, 193, L35
Interstellar abundances toward zeta Ophiuchi

Nagamine, K., Springel, V., & Hernquist, L. 2003a, MNRAS, submitted (astro-ph/0302187)
Abundance of damped Lyman-alpha absorbers in cosmological SPH simulations

Nagamine, K., Springel, V., & Hernquist, L. 2003b, MNRAS, submitted (astro-ph/0305409)
Star formation rate and metallicity of damped Lyman-alpha absorbers in cosmological SPH simulations

Norberg, P., Cole S., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., and the 2dF collaboration 2001, MNRAS, 336, 907
The 2dF Galaxy Redshift Survey: the bJ-band galaxy luminosity function and survey selection function

Oke, J. B., & Sandage, A. 1968, ApJ, 154, 21
Energy Distributions, K Corrections, and the Stebbins-Whitford Effect for Giant Elliptical Galaxies

Outram, P.J., Smith, R.J., Shanks, T., Boyle, B.J., Croom, S.M., Loaring, N.S., & Miller, L. 2001
The 2dF QSO Redshift Survey - VIII. Absorption systems in the 10k catalogue

Padmanabhan, T., Cen, R., Ostriker, J. P., & Summers, F. J. 1996, ApJ, 466, 604
Patterns in Nonlinear Gravitational Clustering: A Numerical Investigation

Padmanabhan, T. 1998, in "After the first three minutes", Cambridge University Press.

Peacock, J. A., & Heavens, A. F. 1985, MNRAS, 217, 805
The statistics of maxima in primordial density perturbations

Peacock, J. A. 1999, in "Cosmological physics", eds Cambridge University Press.

Peebles, P. J. E. 1966, ApJ, 146, 542
Primordial Helium Abundance and the Primordial Fireball. II

Peebles, P. J. E. 1980, "The large scale structure of the Universe", Princeton University Press.

Pei, Y. C., Fall, S. M., & Hauser, Michael G. 1999, ApJ, 422, 604
Cosmic Histories of Stars, Gas, Heavy Elements, and Dust in Galaxies

Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419L
A Measurement of Excess Antenna Temperature at 4080 Mc/s

Péroux, C., Irwin, M. J., McMahon, R. G., & Storrie-Lombardi, L. J. 2001, ApSSS, 217, 551
The Evolution and Space Density of Damped Lyman-alpha Absorbers

Péroux, C., private communication

Petitjean, P., Aracil, B., Srianand, R., & Ibata, R. A. 2000, A&A, 359, 457
Structure of the MgII and damped Lyman-alpha systems along the line of sight to APM 08279+5255 and damped Lyman-alpha systems along the line of sight to APM 08279+5255

Petrosian, A. R., Boulesteix, J., Comte, G., Kunth, D., & Lecoarer, E. 1997, A&A, 318, 390
An interferometric study of the blue compact dwarf galaxy IZW 18.

Pettini, M., Smith, L. J., Hunstead, R. W., & King, D. L. 1994, ApJ, 426, 79
Metal enrichment, dust and star formation in galaxies at high redshifts. Zn and Cr abundances for 17 Damped Ly-alpha systems

Pettini, M., Smith, L. J., King, D. L., & Hunstead, R. W. 1997, ApJ, 486, 665
The metallicity of high-redshift galaxies: The abundance of Zinc in 34 DLAs from $z = 0.7$ to 3.4

Pettini, M., Ellison, S. L., Steidel, C.C., & Bowen, D. 1999, ApJ, 510, 576
Metal Abundances at zj1.5: Fresh Clues to the Chemical Enrichment History of Damped LYalpha Systems

Pettini, M., Ellison, S. L., Steidel, C.C., Shapley, A. E., & Bowen, D.V. 2000, ApJ, 532, 65
Si and Mn Abundances in Damped Ly-alpha: Systems with Low Dust Content

Pettini, M. 2003, XIII Canary Islands Winter Scool of Astrophysics, 'Cosmochemistry: The Melting Pot of Elements' (astro-ph/0303272)
Elements abundances through the ages

Porciani, C., & Giavalisco, M. 2002, ApJ, 565, 24
The Clustering Properties of Lyman Break Galaxies at Redshift $z\sim3$

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1992, in "Numerical Recipes in C", Cambridge: Cambridge University Press

Prochaska, J. X., & Wolfe, A. M. 1997, ApJ, 487, 73
On the Kinematics of the Damped Lyman-alpha Protogalaxie

Prochaska, J. X., & Wolfe, A. M. 2002, ApJ, 566, 68
The UCSD HIRES/Keck I Damped Ly Abundance Database. II. The Implications

Prochaska, J. X., & Wolfe, A. M. 2003, AJ, submitted (astro-ph/0305314)
The age-metallicity relation of the Universe in neutral gas: the first 100 DLAs

Rao, S. M., & Briggs, F. 1993, ApJ, 419, 515
Neutral Hydrogen in Galaxies at the Present Epoch

Rao, S. M. & Turnshek, D. A. 1998, ApJ, 500, L115
Discovery of z = 0.0912 and z = 0.2212 Damped Ly alpha Absorption-Line Systems toward the Quasar OI 363: Limits on the Nature of Damped LY alpha Galaxies

Rao, S. M., & Turnshek, D. A. 2000, ApJS, 130, 1
Discovery of Damped Ly-alpha; Systems at Redshifts Less than 1.65 and Results on Their Incidence and Cosmological Mass Density

Rosenberg, J. L. 2000, Ph. D. Thesis — University of Massachusetts at Amherst
Filling the gap in the extragalactic census : a study at 21 cm and in the near-infrared

Rosenberg, J. L., & Schneider, S. E. 2002a, ApJ, 567, 247
The Arecibo Dual-Beam Survey: The H I *Mass Function of Galaxies*

Rosenberg, J. L., & Schneider, S. E. 2002b, ApJ, 585, 256
The contribution of HI-rich galaxies to the damped Lyman-alpha absorber population at $z = 0$

Sagan, C. 1996, Excerpted from a commencement address delivered May 11, 1996.

Sargent, W. L. W., Young, P. J., Boksenberg, A. & Tytler, D. 1980, ApJS, 42, 41
The distribution of Lyman-alpha absorption lines in the spectra of six QSOs - Evidence for an intergalactic origin

Saslaw, W. C. 2000, in "The distribution of the galaxies. Gravitational clustering in cosmology", Cambridge University Press

Savaglio, S. 2000, Invited talk at the IAU Symposium 204 "The Extragalactic Infrared Background and its Cosmological Implications", Manchester, August 2000, eds. M. Harwit & M.G. Hauser (astro-ph/0011473)
The Metallicity Evolution of Damped Lyman-alpha Systems

Sawicki, M. J., Lin H. & Yee H. K. C. 1997, AJ, 113, 1
Evolution of the Galaxy Population Based on Photometric Redshifts in the Hubble Deep Field

Schaye, J. 2001, ApJ, 559, L1
On the Relation between High-Redshift Starburst Galaxies and Damped Ly-alpha; Systems

Shapley, E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., & Pettini, M. 2001, ApJ, 562, 95
The Rest-Frame Optical Properties of z=3 Galaxies

Schlegel, D., Finkbeiner, D., & Davis, M., ApJ, 1998, 500, 525
Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds

Schmidt, M., Schneider, D. P., & Gunn, J. E. 1995, AJ, 110, 1
Spectroscopic CCD surveys for Quasars at Large redshifts IV

Schneider D. P., Schmidt, M., & Gunn, J. E. 1991, AJ, 101, 2004
Moderate Resolution Spectrophotometry of High Redshift Quasars

Schneider D. P., and the SDSS collaboration, 2002, AJ, 123, 567
The Sloan Digital Sky Survey Quasar Catalog. I. Early Data Release

Scott, J., Bechtold, J., & Dobrzycki, A. 2000, ApJS, 130, 37
A Uniform Analysis of the Ly-alpha; Forest at $z = 0 - 5$. I. The Sample and Distribution of Clouds at $z > 1.7$

Shimasaku, K., & the Subaru collaboration 2003, ApJ, accepted (preprint astro-ph/0302466)
Subaru Deep Survey. IV. Discovery of a Large-Scale Structure at Redshift $\simeq 5$

Smail, I., Hogg, D. W., Yan, L., Cohen, J. G. 1995, ApJ, 449, 105
Deep Optical Galaxy Counts with the Keck Telescope

Smith, R. J., Boyle, B. J., & Maddox, S. J. 1995, MNRAS, 277, 270
The environment of $z < 0.3$ QSOs

Smith, D. A., Allen, R. J., Bohlin, R. C., Nicholson, N., Stecher, T. P. 2000, ApJ, 538, 608
A New Probe of the Molecular Gas in Galaxies: Application to M101

Somerville, R. S., Primack, J. R., & Faber, S. M. 2001, MNRAS, 320, 504
The nature of high-redshift galaxies

Spinrad, H., Stern, D., Bunker, A., et al. 1998, ApJ, 509, 103
The nature of high-redshift galaxies

Steidel, C. C., & Dickinson, M. 1992, ApJ, 394, 81
The unusual field of the quasar 3C 336 - Identification of three foreground MG II absorbing galaxies

Steidel, C. C., & Hamilton, D. 1992, ApJ, 104, 941
Deep imaging of high redshift QSO fields below the Lyman limit. I - The field of Q0000-263 and galaxies at $z = 3.4$

Steidel, C. C., & Hamilton, D. 1993, ApJ, 105, 2017
Deep imaging of high redshift QSO fields below the Lyman limit. II - Number counts and colors of field galaxies

Steidel, C. C., Pettini, M., Dickinson, M., & Persson, S. E. 1995, AJ. 108, 2046

Steidel, C. C., Dickinson. M., Meyer D., Adelberger, K. L., & Sembach, K. R. 1997. ApJ, 480, 568
Quasar Absorbing Galaxies at z <= 1. I. Deep Imaging and Spectroscopy in the Field of 3C 336

Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1
Lyman-Break Galaxies at z>4 and the Evolution of the Ultraviolet Luminosity Density at High Redshift

Steidel, C. C 2000, Proc. SPIE, 4005, 22

Storrie-Lombardi, L. J., McMahon, R. G., & Irwin, M. J. 1996, MNRAS, 283, 79L
Evolution of neutral gas at high redshift: implications for the epoch of galaxy formation

Storrie-Lombardi, L. J., & Wolfe, A. M. 2000, ApJ, 543, 522
Surveys for z ¿ 3 Damped Ly Absorption Systems: The Evolution of Neutral Gas

Teplitz, H. I., Malkan, M., & McLean, I. S. 1998, ApJ, 506, 519
An Infrared Search for Star-forming Galaxies at z > 2

Tyson, N.D. 1988, ApJ, 329, 59L
On the possibility of gas-rich dwarf galaxies in the Lyman-alpha forest

Tytler, D. 1987, ApJ, 321, 19
The distribution of QSO absorption system column densities - Evidence for a single population

Tzanavaris, P., & Carswell, R. F. 2003, MNRAS, submitted
Size estimates for intervening C IV *absoerbers from high resolution spectroscopy of APM 08279+5255*

van den Hoek, L. B., de Blok, W. J. G., van der Hulst, J. M., & de Jong, T. 2000, A&A, 357, 397
The evolution of the stellar populations in low surface brightness galaxies

van der Hulst, J. M., Skillman, E. D., Smith, T. R., Bothun, G. D., McGaugh, S. S., & de Blok, W. J. G. 1993, AJ, 106, 548
Star formation thresholds in Low Surface Brightness galaxies

Veron-Cetty, M.-P., & Veron, P. 2001, A&A, 374, 92
A catalogue of quasars and active nuclei: 10th edition

Viegas, S. M. 1995, MNRAS, 276, 268
Abundances at high redshift: ionization correction factors

Vladilo, G., Bonifacio, P., Centurión, M., & Molaro, P. 2000, ApJ, 543, 24
Zinc as a Tracer of Metallicity Evolution of Damped Ly Systems

Vladilo, G. 2002a, ApJ, 569, 295
A Scaling Law for Interstellar Depletions

Vladilo, G. 2002b, A&A, 391, 407
Chemical abundances of damped Ly alpha systems:. A new method for estimating dust depletion effects

von Zeipel, M. H. 1908, Ann. d'Obs. Paris, 25, 1

Wand, M. P., & Jones, M. C. 1995, "Kernel smoothing", eds Chapman & Hall.

Warren, S. J., Mller, P., Fall, S. M., & Jakobsen, P. 2001, MNRAS, 326, 759
NICMOS imaging search for high-redshift damped Ly-alpha galaxies

Weinberg, D. H., Hernquist, L., & Katz, N. 2003, submitted to A&A, (astro-ph/0212356)
Galaxy clustering and Galaxy Bias in ΛCDM Universe

White, S. D. M., & Rees, M. 1978, MNRAS, 183, 341
Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering

White, R. L., Kinney, A. L., & Becker, R. H. 1993, AJ, 407, 456
New High-Redshift Damped Lyman-alpha Absorption Systems and the Redshift Evolution of Damped Absorbers

Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, ApJS, 61, 249
Damped Lyman-alpha absorption by disk galaxies with large redshifts. I - The Lick survey

Wolfe, A. M. 1993, ApJ, 402, 411
Clustering of galaxies near damped Lyman-alpha systems with $z = 2.6$

Wolfe, A. M., Turnshek, D. A., Lanzetta, K. M., & Lu, L. 1993, ApJ, 404, 480
Damped Lyman-alpha absorption by disk galaxies with large redshifts. IV - More intermediate-resolution spectroscopy

Wolfe, A. M., Lanzetta, K. M., Foltz, C. B., & Chaffee, F.H. 1995, ApJ, 454, 698
The Large Bright QSO Survey for Damped LY alpha Absorption Systems

Wolfe, A. M., Prochaska, J. X., & Gawiser, E. 2003a, ApJ, submitted (astro-ph/0304040)
CII Absorption in Damped Lyman Alpha Systems: (I) Star Formation Rates in a Two-Phase Medium*

Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2003b, ApJ, submitted (astro-ph/0304042)
CII Absorption in Damped Lyman Alpha Systems: (II) A New Window on the Star Formation History of the Universe*

Zehavi, I., Blanton, M. R., Frieman, J. A., Weinberg, D. H., Mo, H. J., Strauss, M. A., Anderson, S. F., Annis, J., Bahcall, N. A., Bernardi, M., and the Sloan collaboration 2002, ApJ, 571, 171
Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data

Zuo, L., & Lu, L. 1993, ApJ, 418, 601
Measurements of D A for a Large QSO Sample and Determination of Evolution of Lyman-Alpha Clouds

Zwaan, M. A., Staveley-Smith, L., Koribalski, B. S., Henning, P. A. et al 2003, AJ, 125, 2842
The 1000 Brightest HIPASS Galaxies: The H I*-Mass Function and* $\Omega_{\rm H\,I}$