



Universiteit
Leiden
The Netherlands

Deep infrared studies of massive high redshift galaxies

Labbé, I.

Citation

Labbé, I. (2004, October 13). *Deep infrared studies of massive high redshift galaxies*. Retrieved from <https://hdl.handle.net/1887/578>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/578>

Note: To cite this publication please use the final published version (if applicable).

Deep Infrared Studies of Massive High Redshift Galaxies

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 13 oktober 2004
klokke 14.15 uur

door

Ivo Ferdinand Louis Labbé
geboren te Alphen a/d Rijn
in 1972

Promotiecommissie

Promotor: Prof. dr. M. Franx

Referent: Prof. dr. P.D. Barthel (Universiteit Groningen)

Overige leden: Prof. dr. P.G. van Dokkum (Yale, New Haven, USA)
Prof. dr. K.H. Kuijken
Prof. dr. H.-W. Rix (MPIA, Heidelberg, Germany)
Prof. dr. P.T. de Zeeuw

Table of contents		Page
Chapter 1. Introduction and Summary		1
1	Introduction	1
1.1	Observational Cosmology	1
1.2	Galaxy Formation	2
1.3	Massive High Redshift Galaxies	4
1.4	The Faint InfraRed Extragalactic Survey	6
2	Outline and Summary	8
3	Conclusions and Outlook	11
Chapter 2. Ultradeep NIR ISAAC observations of the HDF-South: observations, reduction, multicolor catalog, and photometric redshifts		15
1	Introduction	16
2	Observations	17
2.1	Field Selection and Observing Strategy	17
2.2	Observations	19
3	Data Reduction	20
3.1	Flatfields and Photometric Calibration	21
3.2	Sky Subtraction and Cosmic Ray Removal	22
3.3	First Version and Quality Verification	23
3.4	Additional Processing and Improvements	23
3.5	Final Version and Post Processing	24
4	Final Images	25
4.1	Properties	25
4.2	Image Quality	27
4.3	Astrometry	28
4.4	Backgrounds and Limiting Depths	28
5	Source Detection and Photometry	30
5.1	Detection	31
5.2	Optical and NIR Photometry	32
6	Photometric Redshifts	35
6.1	Photometric Templates	36
6.2	Z_{phot} Uncertainties	36
6.3	Stars	37
7	Catalog Parameters	38
8	Analysis	39
8.1	Completeness and Number Counts	39
8.2	Color-Magnitude Distributions	44

9	Summary and Conclusions	48
Chapter 3. Large disk-like galaxies at high redshift		53
1	Introduction	54
2	Observations	55
3	Rest-frame Optical versus UV Morphology	55
4	Profile fits and sizes	57
5	Spectral Energy Distribution	58
6	Discussion	59
Chapter 4. The rest-frame optical luminosity density, color, and stellar mass density of the universe from $z=0$ to $z=3$		63
1	Introduction	64
2	Data	66
3	Measuring Photometric Redshifts and Rest-Frame Luminosities . .	67
3.1	Photometric Redshift Technique	67
3.1.1	Star Identification	68
3.2	Rest-Frame Luminosities	68
3.2.1	Emission Lines	70
4	The Properties of the Massive Galaxy Population	73
4.1	The Luminosity Density	73
4.1.1	The Evolution of $j_{\lambda}^{\text{rest}}$	76
4.2	The Cosmic Color	78
4.3	Estimating M/L_V^* and The Stellar Mass Density	81
5	Discussion	88
5.1	Comparison with other Work	88
5.2	Comparison with $\text{SFR}(z)$	90
5.3	The Build-up of the Stellar Mass	91
6	Summary & Conclusions	92
A	Derivation of z_{phot} Uncertainty	97
B	Rest-Frame Photometric System	99
C	Estimating Rest-Frame Luminosities	99
Chapter 5. The color magnitude distribution of field galaxies at $1 < z < 3$: the evolution and modeling of the blue sequence		103
1	Introduction	104
2	The Data	106
2.1	The Observations and Sample Selection	106
2.2	Photometric Redshifts and Rest-Frame Colors	107
3	The rest-frame Color-Magnitude Distribution of Galaxies from $z \sim 1$ to $z \sim 3$	108
4	The Color-Magnitude Relation of Blue Field Galaxies	109
4.1	The Slope and its Evolution	109
4.2	The Zeropoint and its Evolution	112
4.3	The Origin of the Blue Sequence in the Local Universe	114

4.4	Comparison to $z \sim 3$ Galaxies	117
4.4.1	The Models	117
4.4.2	Results	118
4.4.3	The Blue CMR in Various Rest-Frame Colors	119
5	Constraints of the Color-Magnitude Relation on the Star Formation Histories of Blue Galaxies at $z \sim 3$	120
5.1	A library of Star Formation Histories	120
5.2	Fitting Method	121
5.2.1	Creating Mock Observations	121
5.2.2	The Fitting	122
5.3	Results	123
5.3.1	Constant Star Formation	123
5.3.2	Exponentially Declining Star Formation	125
5.3.3	Repeated Bursts	126
5.3.4	Episodic Star Formation: The Duty Cycle	128
5.4	Discussion	130
6	The Onset of the Red galaxies	132
7	Summary and Conclusions	134
Chapter 6. IRAC Mid-infrared imaging of red galaxies at $z > 2$: new constraints on age, dust, and mass		139
1	Introduction	140
2	The Observations, Photometry, and Sample Selection	141
3	Mid-Infrared Properties of Red Galaxies at $z > 2$	143
4	Comparison to Lyman Break Galaxies	147
5	The rest-frame K-band mass-to-light ratio	149
6	Discussion and Conclusions	150
Nederlandse samenvatting (Dutch summary)		155
Curriculum Vitae		161
Nawoord / Acknowledgments		163

Introduction and Summary

1 Introduction

1.1 Observational Cosmology

COSMOLOGICAL studies have resulted in a “standard” model: a flat, homogeneous, isotropic universe on large scales, composed of ordinary matter, non-baryonic cold dark matter, and dark energy (Spergel et al. 2003). In this model, the dozen or so of parameters that characterize the universe, most importantly the density of baryons, dark matter, and the expansion of the universe, successfully describe astronomical observations on scales from a few Mpc to several 1000 Mpc (Freedman et al. 2001; Efstathiou et al. 2002; Spergel et al. 2003). Large scale structure and galaxies grew gravitationally, from tiny, nearly scale-invariant adiabatic Gaussian fluctuations.

I remember vividly the first cosmology conference I attended at the start of this thesis work. I confess the discussions about cosmic microwave background fluctuations seemed endless, as were the discussions on the large scale structure seen in galaxy surveys. I realize now that I did not fully appreciate how much easier my life was going to be knowing that the universe was spatially flat and 13.7 Gyr old. I could focus for many years on questions that would have been much harder to address until these major issues were solved. Given the enormity of the subject in both distance and time, what cosmologists have learned in the last century, leading up to the arrival of this “standard model”, is nothing short of an extraordinary success.

Leaving the mother of all questions “What is the the origin of the universe?” to the realm of literature and natural philosophy, there exist two major, widely recognized, unresolved issues in contemporary cosmology.

The first one relates to the seminal revelation by Fritz Zwicky who noted the existence of dark matter in the Coma cluster (Zwicky 1937), or the problem of missing mass, as it was then known. Zwicky measured the velocity dispersion of

galaxies in the Coma cluster and used the virial theorem (Eddington 1916), which relates the total internal kinetic energy of the cluster to its gravitational potential, to argue that there was much more mass in the cluster than could be attributed to the stars in galaxies. In solar units, the ratio of mass to optical luminosity of a galaxy such as the Milky Way is $\sim 3M_{\odot}/L_{\odot}$, whereas for the Coma cluster the ratio was about 400. This implies that there must be about 100 times more hidden or “dark” matter as compared with matter in stars. The realization grew that all visible matter is only a minor constituent of the universe.

Dark matter has proved remarkably elusive and despite more than 70 years of observational astronomy and experimental physics – and a host of respectable and less-respectable candidates – there has been no confirmation of its nature, except that it is gravitating, does not emit or absorb light, is non-baryonic, and non-leptonic. Worse, since the rise of the standard cosmological model it has been joined by its even more elusive cousin, Vacuum or Dark Energy: the mysterious substance that apparently dominates the energy budget of the Universe, and is believed to fuel its accelerated expansion. Yet in contrast to the fundamentally impenetrable mist of the Big Bang, Dark Matter and Dark Energy are not so far removed that their effects cannot be measured with the present-day methods and technology. If anything, its presence challenges us towards new physics and promises to keep cosmology vibrant for quite some time.

The second great question in cosmology is the main inspiration of this thesis. How did galaxies, harboring most of the visible matter in the universe, form out of an almost perfectly smooth distribution of matter, 379,000 years after the Big Bang? In the standard model, the growth of structure with time on a range of scales, from the largest scale structure, to galaxies, and to perhaps even the first stars, is driven by the gradual hierarchical merging of Cold Dark Matter (CDM) Halos (White & Rees 1978). Cold intergalactic gas cooled onto these halos and was subsequently converted into the stars which shine as galaxies. Observation of galaxies can thus be used to trace the evolution of both normal and dark matter.

Given the theoretical foundation of structure formation and the general principles thought to govern the formation of galaxies, the formidable task is to understand the physics of star formation on galactic scales, the resultant feedback of energy and material into the interstellar and intergalactic medium, and the role of feedback from supermassive black holes that lurk in the center of galaxies. However formidable, the problem is likely one of complexity, but not in any sense a fundamental problem requiring new physics to be discovered.

1.2 Galaxy Formation

One of the most prominent and important clues to understanding galaxy evolution is the range and distribution of galaxy morphologies present in the local universe. These morphologies were first classified by Hubble (1930) and consist of four major classes: (i) elliptical galaxies (E), (ii) lenticular galaxies (S0), (iii) ‘spiral’ galaxies,

(iv) all remaining galaxies, considered ‘irregular’. Elucidating the origin of the “Hubble Sequence” is a crucial and necessary constraint on all models of galaxy formation.”

One of the strongest constraints comes from E+S0s galaxies, collectively called “early-type” galaxies, which are the most massive galaxies in the local universe. The standard CDM scenario prescribes that early-type galaxies formed by mergers of spiral galaxies at relatively recent times (Toomre & Toomre 1972; Toomre 1977; White & Rees 1978). In this scenario, the appearance of galaxies should reflect the growing and merging pattern of their dark matter halos, and as such galaxies should be growing through merging and/or star formation to the present day (e.g., White & Frenk 1991; Kauffmann & White 1993).

Strangely enough, extensive study of nearby early-type galaxies in clusters contradicts this picture. The massive early-type galaxies form an extremely homogeneous class, obeying tight relations in their properties, such as colors and luminosities (Sandage & Visvanathan 1978). Since the colors of stellar populations change with age, the small scatter in the color-magnitude relation constrains the scatter in age at a given magnitude. For ellipticals in the Coma Cluster, the estimated intrinsic scatter of 0.04 in $U - V$ colors implied a spread in age of less than 15% (Bower, Lucey, & Ellis 1992). The usual interpretation is that the stars in these galaxies formed at high redshift $z > 2$. Detailed modelling of the spectra showed a similar picture (Trager et al. 2000).

The formation scenario of nearby early-type galaxies in clusters is closer to the classical picture of the monolithic collapse, the antipole of the hierarchical galaxy formation. In the monolithic scenario, E+S0s assembled their mass and formed their stars in a rapid event, of much shorter duration than their average age (Eggen, Lynden-Bell, & Sandage 1962; Larson 1975; van Albada 1982). The formation process happened at high redshifts predicting that the progenitors of today’s massive early-type galaxies would be vigorously star-forming with star formation rates of much more than $100 M_{\odot} \text{yr}^{-1}$.

For decades, the chief way to make the distinction between these opposite formation scenarios was through minute observations of the colors and spectra of galaxies in the local universe. However, direct inference from the fossil evidence to explain its origin is impossible given the complexity of galaxy formation. Even with powerful computational solutions progressing at high pace, neither hydrodynamical simulations (e.g., Katz & Gunn 1991; Springel et al. 2001; Steinmetz & Navarro 2002) nor state-of-the-art “semi-analytic” models (e.g., Kauffmann, White, & Guiderdoni 1993; Somerville & Primack 1999; Cole, Lacey, Baugh, & Frenk 2000) provide accurate predictions for the formation history of galaxies. The major problem still to hurdle is the correct description of galactic scale baryonic processes involved in stellar birth, evolution, and death. In contrast to the cosmic microwave background, where we were confident of the physics, were able to construct elaborate deductive models, and test them observationally, the physics of galaxy formation is perhaps too complicated to consider from first principles.

Here, it is more robust to restrict our science to simple inferences from direct observations.

Clearly, observing galaxies *while they are forming* provides a powerful and complementary approach. Nature’s gift to cosmologists, the finite speed of light, allows us to look back in time as we observe galaxies at increasing distance, as measured by their redshifts. This has been the corner-stone of the “look-back” approach to studying galaxy evolution. Observations of extremely distant and extremely faint galaxies at $z = 2$ and higher allow us a direct look into the process of their formation.

1.3 Massive High Redshift Galaxies

Answers to many of the questions are now becoming available from large statistical surveys of galaxies, using new instruments on the largest ground-based and space-based telescopes. Multicolor imaging is our most powerful tool for understanding galaxy evolution, as broadband filter measurements can go very deep and cover large areas. However, even the most ambitious apparent magnitude-selected surveys showed that simply observing fainter galaxies is a relatively inefficient means of assembling significant samples of galaxies at high redshift (e.g., Lilly et al. 1996; Ellis et al. 1996 Songaila et al. 1994; Cohen et al. 1996; Cowie et al. 1996). Most faint galaxies on the sky are in fact nearby.

The study of distant galaxies was revolutionized in the past decade, when the increased surveying power was coupled with advances in photometric pre-selection techniques. Of particularly importance was the identification of a large sample of high redshift galaxies from a specific color signature – red in $U - B$ colors, blue in $B - V$ colors. This color signature is produced when neutral hydrogen, from the intergalactic medium and within the galaxies, absorbs the flux of young hot stars ($t \lesssim 10^7$ yr) shortward of Lyman 1216 Å and the 912 Å Lyman limit, which corresponds to the U -band at $z \sim 3$ (e.g., Guhathakurta, Tyson, & Majewski 1990; Songaila, Cowie, & Lilly 1990, Steidel & Hamilton (1992, Steidel, Pettini, & Hamilton 1995).

At the time this thesis work commenced, Lyman Break Galaxies (LBGs), as they are commonly called, were among the best studied classes of distant galaxies. Close to a 1000 had been spectroscopically confirmed (Steidel et al. 1996a,b, 2003) at $z \gtrsim 2$, and they were found to be a major constituent of the early universe, with space densities similar to local luminous galaxies. Could LBGs be the forerunners of today’s massive ellipticals?

The strong clustering of LBGs indicated that they were luminous tracers of the underlying massive dark matter halos (Adelberger et al. 1998; Giavalisco et al. 1998; Giavalisco & Dickinson 2001) suggesting that they evolve into massive galaxies. Initially it was thought that they were very massive (Steidel et al. 1996b), but direct internal kinematic measurements revealed velocity dispersions of 60–100 kms^{-1} (Pettini et al. 2001; Shapley et al. 2001), which combined with their

compact sizes of a few kiloparsecs (Giavalisco, Steidel, & Macchetto 1996) suggested virial masses of only $10^{10} M_{\odot}$, less than 1/10th of today's most massive galaxies. Their spectra were consistent with on-going star forming activity, resembling that of local star bursting galaxies, with only moderate extinction by dust (Adelberger & Steidel 2000). The estimated star formation rates were only $10 - 15 M_{\odot} \text{yr}^{-1}$, much lower than expected for the progenitors of local early-types in the monolithic scenario. Studies of the stellar composition, using optical and near-infrared broadband photometry, reinforced this picture of relatively low-mass, moderately star forming objects (Sawicki & Yee 1998; Shapley et al. 2001; Papovich, Dickinson, & Ferguson 2001). These properties are expected in the standard hierarchical scenario of galaxy formation, where LBGs are the low-mass building blocks that merge to become present-day massive galaxies in groups and clusters (e.g., Mo & Fukugita 1996; Baugh et al. 1998).

It was quickly realized that LBGs were selected in a very peculiar way. Galaxies that were dust-reddened by more than $E(B-V) \gtrsim 0.4$ were not selected (Adelberger & Steidel 2000). Could there still be a population of vigorously star forming galaxies that escaped detection because it was enshrouded in dust? We know that complementary samples of heavily obscured high-redshift galaxies are found in sub-mm, radio (Smail et al. 2000), and X-ray surveys (Cowie et al. 2001; Barger et al. 2001). Their number densities are much lower than LBGs, although their contribution to the total star formation rate can be significant (cf., Adelberger & Steidel 2000; Barger, Cowie, & Sanders 1999).

Even more important, have we overlooked a major population at $z > 2$ of evolved galaxies that resemble the present-day elliptical and spiral galaxies? None of the techniques described above would select those normal galaxies, whose light is dominated by evolved stars. Specifically, even the massive Milky Way would have never been selected with the Lyman Break technique if placed at $z \sim 3$, as it does not have the required high far-UV surface brightness.

It is much easier to detect evolved galaxies at $z > 2$ in the near-infrared (NIR), where one can access their rest-frame optical light, and where evolved stars emit the bulk of their light. The rest-frame optical light is also much less sensitive to the effects of dust obscuration and on-going star formation than the UV, and expected to be a better (yet imperfect) tracer of stellar mass.

Researchers had already begun using NIR data to look for evolved galaxies at $z > 1$. The technique relied on photometric pre-selection on the stellar Balmer and 4000 Å break. The Balmer discontinuity at 3650 Å is generally strong in stellar populations of age 10^8 to 10^9 yr, and is most pronounced in the photosphere of A-stars, while even older stellar populations show a characteristic 4000 Å break due to the sudden onset below 4000 Å of metallic and molecular absorption in cool stars.

Evolved galaxies were targeted by selecting on red optical-NIR colors $R-K > 5$ or $I-K > 4$, which can be produced as the Balmer/4000 Å break redshifts

out of the optical filters. Recent surveys yielded large samples of these so-called Extremely Red Objects at redshifts $0.8 < z < 1.8$ (EROs; e.g., Elston, Rieke, & Rieke 1988; Hu & Ridgway 1994; Thompson et al. 1999; Yan et al. 2000; Scodreggio & Silva 2000; Daddi et al. 2000; McCarthy et al. 2001; Smith et al. 2002, Moustakas et al. 2004), but their relation to the well-formed massive early-type galaxies at $z < 1$ has not been firmly established. Apparently, a minor fraction of 30% are strongly clustered and evolved galaxies (e.g., Cimatti et al 2002, Daddi et al. 2002), while the rest are believed to be dust reddened star-forming objects (e.g., Yan & Thompson 2003; Moustakas et al. 2004).

To observe evolved galaxies at even higher redshifts $z > 2$ one has to develop criteria to select the galaxies and at the same time go deep enough to overcome the cosmological effect of the $(1+z)^4$ surface brightness dimming. Normal evolved galaxies at $z \sim 3$ would be incredibly faint. In addition as the Balmer/4000Å break is less strong than the Lyman break, it requires a combination of extremely deep optical and NIR imaging to select them.

There were very few datasets that reached the required depths, most of them taken with the WFPC2 and NICMOS instruments on the Hubble Space Telescope (Thompson et al 2000; Dickinson et al 2000; Williams et al 2000). The largest survey to date is that of Dickinson et al. (2000), who imaged the Hubble Deep Field North (HDFN) WFPC2 field in J_{110} and H_{160} with NICMOS, finding that most high redshift galaxies would have been picked up by the Ly-break technique. Nevertheless, the total area studied is still very small, and the depth of the K -band data, which is important in constraining the contribution of evolved stars at $z \sim 3$, is not well matched to that of the rest of the NICMOS or WFPC2 imaging data.

To remedy this situation, we started the Faint InfraRed Extragalactic Survey: an ultradeep optical-to-infrared multicolor survey of high redshift galaxies.

1.4 The Faint InfraRed Extragalactic Survey

This dissertation is based on the Faint InfraRed Extragalactic Survey (FIRES; Franx et al. 2000) and its aims are intimately connected with the goals of this survey. FIRES is a large public program at the *Very Large Telescope* (VLT) consisting of very deep NIR imaging of two fields. The fields are the WFPC2-field of Hubble Deep Field South (HDFS), and the field around the $z = 0.83$ cluster MS1054-03, both selected for their exquisite, deep optical WFPC2 imaging with the HST.

The central question in this thesis is:

How did massive galaxies assemble over time?

This general question is addressed in the following chapters by analyzing the following specific issues:

1. **Have we overlooked a major population at high redshift perhaps resembling the present-day normal elliptical and spiral galaxies?**
2. **When did the Hubble Sequence of galaxies start manifesting?**
3. **How and when did galaxies assemble the bulk of their stellar mass?**
4. **Can we use the distribution of galaxy properties to constrain formation scenarios?**
5. **What is the detailed stellar composition of $z > 2$ galaxies: are there passively evolving galaxies at $z \sim 3$, and what is the role of dust?**

The body of this work deals with the properties of faint distant galaxies as observed in the HDFs, a small, otherwise undistinguished high-galactic latitude patch of sky. The HDFs and its counterpart in the north, the HDFN, constitute milestones in optical imaging, as the WFPC2 camera on the HST was pushed to its limits. The ultra-deep imaging in four optical bands ($U_{300}, B_{450}, V_{606}, I_{814}$) of these “empty” fields, named after the absence of any large foreground galaxies, allowed an unprecedented deep look of the distant universe, and opened the door to the study of normal galaxies at $z > 2$.

It is often wondered whether such a small field presents us a fair view of the universe. Luckily, the universe seems not to be fractal or hierarchically structured beyond a few hundred Mpc, and voids and superclusters like the ones near us simply repeat. Therefore at large distances, even a pinhole survey such as the HDFs may sample enough volume as to obtain a representative picture. We should always remember, however, that the field size and volume of the HDFs is small.

Capitalizing on the advances in NIR detector capabilities, we took to the Infrared Spectrometer and Array Camera (ISAAC, Moorwood 1997) on the VLT, and observed this tiny field in the NIR J_s, H , and K_s filters for more than 100 hours total, and only under the best seeing conditions. The second field, centered on the $z = 0.83$ cluster MS1054-03, was observed for 80 hours (Förster Schreiber et al. 2004a). While not as spectacularly deep, the area surveyed in MS1054-03 is nearly five times larger, and turned out to be a crucial element for its ability to reinforce our findings in the HDFs.

A special asset in the deep imaging set was the K_s -band, the reddest band from the ground where achievable sensitivity and resolution were still somewhat comparable to the space-based optical data. At $z \sim 3$ it probes rest-frame wavelengths $\lambda \sim 5400\text{\AA}$, comfortably redward of the Balmer/4000 $\text{\AA}$ break and therefore crucial to assess the build-up of evolved stars.

Aiming for evolved galaxies at $z > 2$, we experimented with simple color-

selection criteria, analogous to those applied to EROs at lower redshift $1 < z < 2$. We selected high-redshift galaxies with the simple color criterion $J_s - K_s > 2.3$, specifically designed to target the Balmer/4000 Å breaks at redshifts between $2 < z < 4$. While candidate high-redshift galaxies with even redder $J - K$ colours have been reported by other authors as well (e.g. Scodeggio & Silva 2000; Hall et al. 2001; Totani et al. 2001; Saracco et al. 2001, 2003), the focus was usually on the objects with the most extreme colors. The $J_s - K_s > 2.3$ criterion, in fact, is not extreme at all as it corresponds to a $U - V$ color of 0.1 at $z = 2.7$. Such a selection would include all but the bluest present-day Hubble Sequence galaxies.

2 Outline and Summary

We present in **Chapter 2** the FIRES observations of the HDFS, the data reduction, the assembly of the photometric source catalogs, and the photometric redshifts. These data constitute the deepest groundbased NIR images to date, and the deepest K_s -band in any field, even from space (Labbé et al. 2003). We released the reduced data, catalog of sources, and photometric redshifts to the community, and they are now in use by many researchers worldwide.

One immediate scientific breakthrough was the identification of a significant population of galaxies with red $J_s - K_s > 2.3$ colors at $z > 2$. We find these Distant Red Galaxies, or DRGs as we shall now call them, in substantial numbers, with space densities about half of that of LBGs selected from ground-based imaging (Franx et al. 2003). Our follow-up studies suggested that DRGs, at a given rest-frame optical luminosity, have higher ages, contain more dust, and are more massive than LBGs (Franx et al. 2003; van Dokkum et al. 2004; Förster Schreiber et al. 2004),

These galaxies had been previously missed because they are extremely faint in the optical (rest-frame UV) and emit most of their light in the NIR (rest-frame optical). Surprisingly, galaxies with comparable colors at $2 < z < 3.5$ were almost absent in HDFN (cf., Labbé et al. 2003, Papovich, Dickinson, & Ferguson 2001) even though it was surveyed over a similar area and depth (Dickinson 2000). Clearly, cosmic variance due to large scale structure in the universe plays a role here, and the possibility existed that neither of the two fields was representative. Later, our findings in the HDFS were confirmed by the discovery of DRGs at similar densities in the much larger MS1054-03 field (see, e.g., van Dokkum et al. 2003; Förster Schreiber et al. 2004).

Another galaxy population that believed to be absent at high redshift, were large disk galaxies. Disk galaxies are thought to undergo a relatively simple formation process in which gas cools and contracts in dark matter halos to form rotationally supported disks with exponential light profiles (Fall & Efstathiou 1980; Mo, Mao, & White 1998). In the standard hierarchical model, large disks form relatively late $z < 1$, and very few of them are expected as early as $z \sim 2$.

In **Chapter 3** we report the discovery of 6 galaxies in the HDF-S at $1.5 \lesssim z \lesssim 3$ with colors, morphologies and sizes comparable to local spiral galaxies (Labbé et al. 2003b). The irregular optical WFPC2 (rest-frame far-UV) morphologies galaxies had previously been misinterpreted, because they traced the sites of unobscured star-formation rather than the underlying evolved population. Here, the combination of bandpass shifting and surface brightness dimming had given an exaggerated impression of evolution towards high redshift. In the NIR, however (rest-frame optical), the morphologies were much more regular than in the rest-frame far-UV, with *well resolved* exponential profiles as expected for rotating stellar disks. Models of disk formation in the standard CDM scenario (Mo, Mao, & White 1999) currently underpredict the number of large disks at high redshift by a factor of two. Only larger samples and kinematical studies, to establish the presence of rotating disks, can tell how serious this discrepancy is, and when classical Hubble sequence spiral galaxies came into existence.

In **Chapter 4** we analyze the cosmological growth of the stellar mass density from redshift $z \sim 3$ to $z = 0$, as traced by optically luminous galaxies in the HDF-S. Measuring accurate stellar masses from broadband photometry is quite hard. We resorted to interpreting the mean cosmic colors $(U - B)_{rest}$ and $(B - V)_{rest}$ using stellar population synthesis models. We assumed an appropriate star formation history for the universe as a whole, and used the models to derive the global mass-to-light ratio M/L_V in the V -band.

We found that the universe at $z \sim 3$ had a ~ 10 times lower stellar mass density than it does today, and half of the stellar mass of the universe was formed by $z = 1 - 1.5$, broadly consistent with independent results obtained in the HDFN (Dickinson et al. 2003). Interestingly, the distant red galaxies discovered earlier in the survey may have contributed as much as $\sim 50\%$ to the cosmic stellar mass density at $z \sim 3$.

In **Chapter 5** we studied the rest-frame optical color-magnitude distribution of galaxies in the FIRES fields. We focused in particular on the redshift range $z \sim 3$, where observations in the HDFN showed that blue star-forming galaxies followed a clear trend, such that galaxies more luminous in the rest-frame V -band had slightly redder $U - V$ colors (Papovich, Dickinson, & Ferguson 2001; Papovich et al. 2004). The origin of this color-magnitude relation (CMR) for blue galaxies, or blue sequence, was not clearly understood, even at low redshift (cf., Peletier & de Grijs 1998; Tully et al. 1998; Zaritsky, Kennicutt, & Huchra 1994; Bell & De Jong 2001).

We analyzed spectra of nearby galaxies in the Nearby Field Galaxy Survey (Jansen, Fabricant, Franx, & Caldwell 2000; Jansen, Franx, Fabricant, & Caldwell 2000; Jansen, Franx, & Fabricant 2001), and found that the relation is mainly one of increasing dust-opacity with increasing luminosity. We also showed that the slope of this relation does not change with redshift, and may have a similar origin already at $z \sim 3$.

Similarly to studies of the red sequence cluster ellipticals (Bower, Lucey, & Ellis 1992; van Dokkum et al. 1998), we interpreted the scatter around the relation as the result of a spread in ages of the stellar populations. The blue-sequence scatter is fairly narrow, has a conspicuous blue envelope, and skew to red colors. After exploring a range of formation models for the galaxies, we concluded that models with episodic star formation explain most aspects of the $z = 3$ color-magnitude distribution. The episodic models cycle through periods of star formation and quiescence, rejuvenating the stellar population during each active episode. The result is that the luminosity weighted ages of the stars are smaller than the age of the galaxy, i.e., the time since the galaxy first started forming stars. They may provide a solution of the enigmatic observation that $z = 3$ galaxies are much bluer than expected if they were as old as the universe (e.g., Papovich, Dickinson, & Ferguson 2001).

Chapter 6 presents a study of the stellar composition of distant red galaxies and Lyman Break galaxies, using mid-IR imaging from IRAC on the Spitzer Space Telescope. Our previous studies indicated that DRGs have higher ages, contain more dust, and are more massive than LBGs at a given rest-frame optical luminosity (Franx et al. 2003; van Dokkum et al. 2004; Förster Schreiber et al. 2004), and may contribute comparably to the cosmic stellar mass density (Franx et al. 2003; Rudnick et al. 2003). Nevertheless, the nature of their red colors is still poorly understood, and the masses are somewhat uncertain. Are they all truly old, or are some also very young and very dusty? What is the fraction of passively, evolving “dead” systems? How much do the DRGs contribute to the stellar mass density in a mass-selected sample? And how do they relate to the blue Lyman break galaxies. Finally, what is their role in the formation and evolution of massive galaxies?

In this chapter, we present deep IRAC 3.6 – 8 micron imaging of the HDFs field. The new IRAC data reached rest-frame NIR wavelengths, which were crucial in determining the nature of DRGs in comparison to LBGs. We uniquely identified 3 out of 11 DRGs as old passively evolving systems at $z \sim 2.5$. Others were heavily reddened star-forming galaxies, for which we are now better able to distinguish between the effects of age and dust. Furthermore, the rest-frame NIR data allowed more robust estimates of the stellar mass and stellar mass-to-light ratios (M/L_K). We found that in a mass-selected sample DRGs contribute 1.5 – 2 times as much as the LBGs to the cosmic stellar mass density at $2 < z < 3.5$. Also, at a given rest-frame K luminosity the red galaxies are twice as massive with average stellar masses $\sim 10^{11} M_\odot$, and their M/L_K mass-to-light ratios exhibit only a third of the scatter compared to the LBGs. This is consistent with a picture where DRGs are more massive, more evolved, and have started forming at higher redshift than most LBGs.

3 Conclusions and Outlook

We have presented evidence in this thesis that previous imaging surveys gave a biased view of the early universe. The immediate conclusions of the Faint InfraRed Extragalactic Survey are that large numbers of evolved and dust-obscured galaxies at $z = 2 - 3$ have been overlooked, that up to half of the stellar mass in the universe at $z = 2 - 3$ was unaccounted for, and that the morphologies of galaxies were misinterpreted.

With the newest optical-to-MIR multiwavelength surveys, we are for the first time obtaining a better census of the massive galaxies in the early universe. We are one step closer to tracing the assembly of massive galaxies directly, and it seems warranted now to interpolate between the properties of galaxies populations at different epochs. Ultimately, that approach will help us to understand how galaxies evolved from the cradle to the present-day.

Even so, for a full comparison with local samples, our data sets are still much too small. If we would dissect the galaxy population by redshift, luminosity, color, morphology, or environment, to analyze galaxy formation in all its complexity, we would be left with few galaxies in every subsample. A straightforward extension of the current work is thus to obtain much larger samples.

Apart from the obvious enlargement of the samples, it is now crucial to follow-up the current samples with high-resolution NIR imaging, and optical and NIR spectroscopy. High resolution NIR imaging allows to unambiguously determine the rest-frame optical morphologies of galaxies to high redshift. Unfortunately, given the limited availability of the Hubble Space Telescope, and the slow survey speed of the NICMOS camera in particular, we must await the arrival of next-generations of space telescopes (e.g., JWST or JDEM) or the maturing of ground-based solutions, such as active optics with laser guide stars. A dearth of high-resolution imaging of distant galaxies will continue to exist for some time to come.

On the other hand, new NIR spectroscopic instruments promise spectacular advances. In the coming years a number of multi-object NIR spectrographs will arrive at 8–10 meter class telescopes. With these instruments we can obtain deep spectra for hundreds of distant galaxies at the same time. Spectroscopic redshifts constrain the space densities of the sources and allow to get accurate rest-frame colors. Kinematic measurements of nebular lines provide direct estimates of the dynamical masses involved, and reveal any ordered rotation. Finally, their emission line strengths provide independent constraints on star formation rates (SFRs) and dust extinction, helping to break the degeneracies inherent to modeling of the broadband SEDs (e.g., Papovich et al. 2001, Shapley et al 2001). Every topic discussed in this thesis, from the study of high-redshift disk galaxies, via the stellar populations of DRGs, to the origin of the high-redshift color-magnitude relation (and its scatter), benefits directly from deep spectra.

References

- Adelberger, K. L., Steidel, C. C., Giavalisco, M., Dickinson, M., Pettini, M., & Kellogg, M. 1998, *ApJ*, 505, 18
- Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 1998, *ApJ*, 498, 504
- Bower, R. G., Lucey, J. R., & Ellis, R. S. 1992, *MNRAS*, 254, 601
- Casertano, S. et al., *AJ*, 120, pp. 2747–2824, 2000
- Cimatti, A., et al. 2002, *A&A*, 391, L1
- Cohen, J. G., Cowie, L. L., Hogg, D. W., Songaila, A., Blandford, R., Hu, E. M., & Shopbell, P. 1996, *ApJ*, 471, L5
- Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, *AJ*, 112, 839
- Daddi, E., Cimatti, A., Pozzetti, L., Hoekstra, H., Röttgering, H. J. A., Renzini, A., Zamorani, G., & Mannucci, F. 2000, *A&A*, 361, 535
- Daddi, E., et al. 2003, *ApJ*, 588, 50
- Dickinson, M., *Philos. Trans. R. Soc. London A* **358**, p. 2001, 2000
- Dickinson, M., Papovich, C., Ferguson, H. C., & Budavári, T. 2003, *ApJ*, 587, 25
- Eddington, A. S. 1916, *MNRAS*, 76, 525
- Efstathiou, G., et al. 2002, *MNRAS*, 330, L29
- Eggen, Lynden-Bell, & Sandage 1962
- Ellis, R. S., Colless, M., Broadhurst, T., Heyl, J., & Glazebrook, K. 1996, *MNRAS*, 280, 235
- Elston, R., Rieke, G. H., & Rieke, M. J. 1988, *ApJ*, 331, L77
- Erb, D. K., Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., Hunt, M. P., Moorwood, A. F. M., & Cuby, J. 2003, *ApJ*, 591, 101
- Fontana, A., Poli, F., Menci, N., Nonino, M., Giallongo, E., Cristiani, S., & D’Odorico, S. 2003, *ApJ*, 587, 544
- Fontana, A., et al. 2003, *ApJ*, 594, L9
- Forster Schreiber et al. 2004, *ApJ*, accepted
- Forster Schreiber et al. 2004, in preparation
- Franx, M. et al., *The Messenger* **99**, pp. 20–22, 2000
- Franx, M. et al. 2003, *ApJ*, 587, L79
- Freedman, W. L., et al. 2001, *ApJ*, 553, 47
- Giavalisco, M. & Dickinson, M. 2001, *ApJ*, 550, 177
- Giavalisco, M., Steidel, C. C., Adelberger, K. L., Dickinson, M. E., Pettini, M., & Kellogg, M. 1998, *ApJ*, 503, 543
- Guhathakurta, P., Tyson, J. A., & Majewski, S. R. 1990, *ApJ*, 357, L9
- Hall, P. B., et al. 2001, *AJ*, 121, 1840
- Hu, E. M. & Ridgway, S. E. 1994, *AJ*, 107, 1303
- Hubble, E. P. 1930, *ApJ*, 71, 231
- Iverson, R. J., Smail, I., Barger, A. J., Kneib, J.-P., Blain, A. W., Owen, F. N., Kerr, T. H., & Cowie, L. L. 2000, *MNRAS*, 315, 209
- Katz, N. & Gunn, J. E. 1991, *ApJ*, 377, 365
- Kauffmann, G., Charlot, S., & White, S. D. M. 1996, *MNRAS*, 283,
- Kauffmann, G. & White, S. D. M. 1993, *MNRAS*, 261, 921
- Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, *MNRAS*, 264, 201
- Labbé, I., et al. 2003, *AJ*, 125, 1107
- Labbé et al 2004, in preparation
- Larson, R. B. 1975, *MNRAS*, 173, 671
- Lilly, S. J., Le Fevre, O., Hammer, F., & Crampton, D. 1996, *ApJ*, 460, L1

- Lowenthal, J. D. et al. 1997, *ApJ*, 481, 673
- Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A., *MNRAS* 283, pp. 1388–1404, 1996
- McCarthy, P. J., et al. 2001, *ApJ*, 560, L131
- Mo, H. J. & Fukugita, M. 1996, *ApJ*, 467, L9
- Moorwood, A. F. 1997, *Proc. SPIE*, 2871, 1146
- Moorwood, A. F. M., van der Werf, P. P., Cuby, J. G., & Oliva, E. 2000, *A&A*, 362, 9
- Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, *ApJ*, 559, 620
- Pettini, M., Kellogg, M., Steidel, C. C., Dickinson, M., Adelberger, K. L., & Giavalisco, M. 1998, *ApJ*, 508, 539
- Pettini, M., Shapley, A. E., Steidel, C. C., Cuby, J., Dickinson, M., Moorwood, A. F. M., Adelberger, K. L., & Giavalisco, M. 2001, *ApJ*, 554, 981
- Pettini, M., Shapley, A. E., Steidel, C. C., Cuby, J., Dickinson, M., Moorwood, A. F. M., Adelberger, K. L., & Giavalisco, M. 2001, *ApJ*, 554, 981
- Rudnick, G. et al. 2001, *AJ*, 122, 2205
- Rudnick, G., et al. 2003, *ApJ*, 599, 847
- Sandage, A. & Visvanathan, N. 1978, *ApJ*, 223, 707
- Saracco, P., D’Odorico, S., Moorwood, A., & Cuby, J. G. 2001, *Ap&SS*, 276, 991
- Saracco, P., et al. 2003, *A&A*, 398, 127
- Sawicki, M. & Yee, H. K. C. 1998, *AJ*, 115, 1329
- Scodreggio, M. & Silva, D. R. 2000, *A&A*, 359, 953
- Shapley, A. E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., & Pettini, M. 2001, *ApJ*, 562, 95
- Smail, I., Ivison, R. J., Owen, F. N., Blain, A. W., & Kneib, J.-P. 2000, *ApJ*, 528, 612
- Smith, N. 2002, *MNRAS*, 337, 1252
- Smith, V. V., et al. 2002, *AJ*, 124, 3241
- Somerville, R. S. & Primack, J. R. 1999, *MNRAS*, 310, 1087
- Songaila, A., Cowie, L. L., Hu, E. M., & Gardner, J. P. 1994, *ApJS*, 94, 461
- Songaila, A., Cowie, L. L., & Lilly, S. J. 1990, *ApJ*, 348, 371
- Spergel, D. N., et al. 2003, *ApJS*, 148, 175
- Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, *AJ*, 112, 352
- Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, *ApJ*, 462, L17
- Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, *ApJ*, 519, 1
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., Pettini, M., Dickinson, M., & Giavalisco, M. 2003, *ApJ*, 592, 72
- Steidel, C. C. & Hamilton, D. 1992, *AJ*, 104, 941
- Steidel, C. C., Pettini, M., & Hamilton, D. 1995, *AJ*, 110, 2519
- Steinmetz, M. & Navarro, J. F. 2002, *New Astronomy*, 7, 155
- Thompson, R. I., Storrie-Lombardi, L. J., Weymann, R. J., Rieke, M. J., Schneider, G., Stobie, E., & Lytle, D. 1999, *AJ*, 117, 17
- Totani, T., Yoshii, Y., Maihara, T., Iwamuro, F., & Motohara, K. 2001, *ApJ*, 559, 592
- Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, *AJ*, 120, 165
- van Albada, T. S. 1982, *MNRAS*, 201, 939
- van Dokkum, P. G., Franx, M., Kelson, D. D., Illingworth, G. D., Fisher, D., & Fabricant, D. 1998, *ApJ*, 500, 714
- van Dokkum, P. G., Franx, M., Fabricant, D., Illingworth, G. D., & Kelson, D. D. 2000, *ApJ*, 541, 95

- van Dokkum, P. G. et al. 2003, ApJ, 587, L83
van Dokkum, P. G., et al. 2004, ApJ, 611, 703
White, S. D. M. & Frenk, C. S. 1991, ApJ, 379, 52
White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341
Williams, R. E. et al. 1996, AJ, 112, 1335
Williams, R. E. et al. 2000, AJ, 120, 2735
Yan, L., McCarthy, P. J., Weymann, R. J., Malkan, M. A., Teplitz, H. I., Storrie-
Lombardi, L. J., Smith, M., & Dressler, A. 2000, AJ, 120, 575
Zwicky, F. 1937, ApJ, 86, 217

Ultradeep Near-Infrared ISAAC Observations of the Hubble Deep Field South observations, reduction, multicolor catalog and photometric redshifts

ABSTRACT

We present deep near-infrared (NIR) J_s , H , and K_s -band ISAAC imaging of the WFPC2 field of the Hubble Deep Field South (HDF-S). The $2.5' \times 2.5'$ high Galactic latitude field was observed with the VLT under the best seeing conditions with integration times amounting to 33.6 hours in J_s , 32.3 hours in H , and 35.6 hours in K_s . We reach total AB magnitudes for point sources of 26.8, 26.2, and 26.2 respectively (3σ), which make it the deepest ground-based NIR observations to date, and the deepest K_s -band data in any field. The effective seeing of the coadded images is $\approx 0''.45$ in J_s , $\approx 0''.48$ in H , and $\approx 0''.46$ in K_s . Using published WFPC2 optical data, we constructed a K_s -limited multicolor catalog containing 833 sources down to $K_{s,AB}^{tot} \lesssim 26$, of which 624 have seven-band optical-to-NIR photometry. These data allow us to select normal galaxies from their rest-frame optical properties to high redshift ($z \lesssim 4$). The observations, data reduction and properties of the final images are discussed, and we address the detection and photometry procedures that were used in making the catalog. In addition, we present deep number counts, color distributions and photometric redshifts of the HDF-S galaxies. We find that our faint K_s -band number counts are flatter than published counts in other deep fields, which might reflect cosmic variations or different analysis techniques. Compared to the HDF-N, we find many galaxies with very red $V - H$ colors at photometric redshifts $1.95 < z_{phot} < 3.5$. These galaxies are bright in K_s with infrared colors redder than $J_s - K_s > 2.3$ (in Johnson magnitudes). Because they are extremely faint in the observed optical, they would be missed by ultraviolet-optical selection techniques, such as the U-dropout method.

Ivo Labbé, Marijn Franx, Gregory Rudnick, Natascha M. Förster Schreiber,
 Hans-Walter Rix, Alan Moorwood, Pieter G. van Dokkum, Paul van der Werf,
 Huub Röttgering, Lottje van Starckenburg, Arjen van de Wel, Konrad Kuijken, &
 Emanuele Daddi

Astronomical Journal, **125**, 3, 1107 (2003)

1 Introduction

IN the past decade, our ability to routinely identify and systematically study distant galaxies has dramatically advanced our knowledge of the high-redshift universe. In particular, the efficient U-dropout technique (Steidel et al. 1996a,b) has enabled the selection of distant galaxies from optical imaging surveys using simple photometric criteria. Now more than 1000 of these Lyman break galaxies (LBGs) are spectroscopically confirmed at $z \gtrsim 2$, and have been subject to targeted studies on spatial clustering (Giavalisco & Dickinson 2001), internal kinematics (Pettini et al. 1998, 2001), dust properties (Adelberger & Steidel 2000), and stellar composition (Shapley et al. 2001; Papovich, Dickinson, & Ferguson 2001). Although LBGs are among the best studied classes of distant galaxies to date, many of their properties like their prior star formation history, stellar population ages, and masses are not well known.

More importantly, it is unclear if the ultraviolet-optical selection technique alone will give us a fair census of the galaxy population at $z \sim 3$ as it requires galaxies to have high far-ultraviolet surface brightnesses due to on-going spatially compact and relatively unobscured massive star formation. We know that there exist highly obscured galaxies, detected in sub-mm and radio surveys (Smail et al. 2000), and optically faint hard X-ray sources (Cowie et al. 2001; Barger et al. 2001) at high redshift that would not be selected as LBGs, but their number densities are low compared to LBGs and they might represent rare populations or transient evolutionary phases. In addition, the majority of present-day elliptical and spiral galaxies, when placed at $z \sim 3$, would not satisfy any of the current selection techniques for high-redshift galaxies. Specifically, they would not be selected as U-dropout galaxies because they are too faint in the rest-frame UV. It is much easier to detect such galaxies in the near-infrared (NIR), where one can access their rest-frame optical light.

Furthermore, observations in the near-infrared allow the comparison of galaxies of different epochs at fixed rest-frame wavelengths where long-lived stars may dominate the integrated light. Compared to the rest-frame far-UV, the rest-frame optical light is less sensitive to the effects of dust extinction and on-going star formation, and provides a better tracer of stellar mass. By selecting galaxies in the near-infrared K_s -band, we expect to obtain a more complete census of the galaxies that dominate the stellar mass density in the high-redshift universe, thus tracing the build-up of stellar mass directly.

In this context we initiated the Faint InfraRed Extragalactic Survey (FIRES; Franx et al. 2000), a large public program carried out at the *Very Large Telescope* (VLT) consisting of very deep NIR imaging of two selected fields. We observed fields with existing deep optical WFPC2 imaging from the *Hubble Space Telescope* (HST): the WFPC2-field of Hubble Deep Field South (HDF-S), and the field around the $z \approx 0.83$ cluster MS1054-03. The addition of NIR data to the optical photometry is required not only to access the rest-frame optical, but also to

determine the redshifts of faint galaxies from their broadband photometry alone. While it may be possible to go to even redder wavelengths from the ground, the gain in terms of effective wavelength leverage is less dramatic compared to the threefold increase going from the I to K -band. This is because the K_s -band is currently the reddest band where achievable sensitivity and resolution are reasonably comparable to deep space-based optical data. Preliminary results from this program were presented by Rudnick et al. (2001, hereafter R01).

Here we present the full NIR data set of the HDF-S, together with a K_s -selected multicolor catalog of sources in the HDF-S with seven-band optical-to-infrared photometry (covering $0.3 - 2.2\mu\text{m}$), unique in its image quality and depth. This paper focusses on the observations, data reduction and characteristic properties of the final images. We also describe the source detection and photometric measurement procedures and lay out the contents of the catalog, concluding with the NIR number counts, color distributions of sources, and their photometric redshifts. The results of the MS1054-03 field will be presented by Förster Schreiber et al. (2002) and a more detailed explanation of the photometric redshift technique can be found in Rudnick et al. (2001, 2002b). Throughout this paper, all magnitudes are expressed in the AB photometric system (Oke 1971) unless explicitly stated otherwise.

2 Observations

2.1 Field Selection and Observing Strategy

The high Galactic latitude field of the HDF-S is a natural choice for follow-up in the near-infrared given the existing ultradeep WFPC2 data in four optical filters (Williams et al. 1996, 2000; Casertano et al. 2000). The Hubble Deep Fields (North and South) are specifically aimed at constraining cosmology and galaxy evolution models, and in these studies it is crucial to access rest-frame optical wavelengths at high redshift through deep infrared observations. Available ground-based NIR data from SOFI on the NTT (da Costa et al. 1998) are not deep enough to match the space-based data. To fully take advantage of the deep optical data requires extremely deep wide-field imaging in the infrared at the best possible image quality; a combination that in the southern hemisphere can only be delivered by the Infrared Spectrometer And Array Camera (ISAAC; Moorwood 1997), mounted on the Nasmyth-B focus of the 8.2 meter VLT Antu telescope. The infrared camera has a $2.5' \times 2.5'$ field of view similar to that of the WFPC2 ($2.7' \times 2.7'$). ISAAC is equipped with a Rockwell Hawaii 1024 $\times$ 1024 HgCdTe array, offering imaging with a pixel scale of $0''.147 \text{ pix}^{-1}$ in various broad and narrow band filters.

Our NIR imaging consists of a single ISAAC pointing centered on the WFPC2 main-field of the HDF-S ($\alpha = 22^h 32^m 55.464$, $\delta = -60^\circ 33' 05.01''$, J2000) in the J_s , H and K_s filters, which gives good sampling of rest-frame optical wavelengths

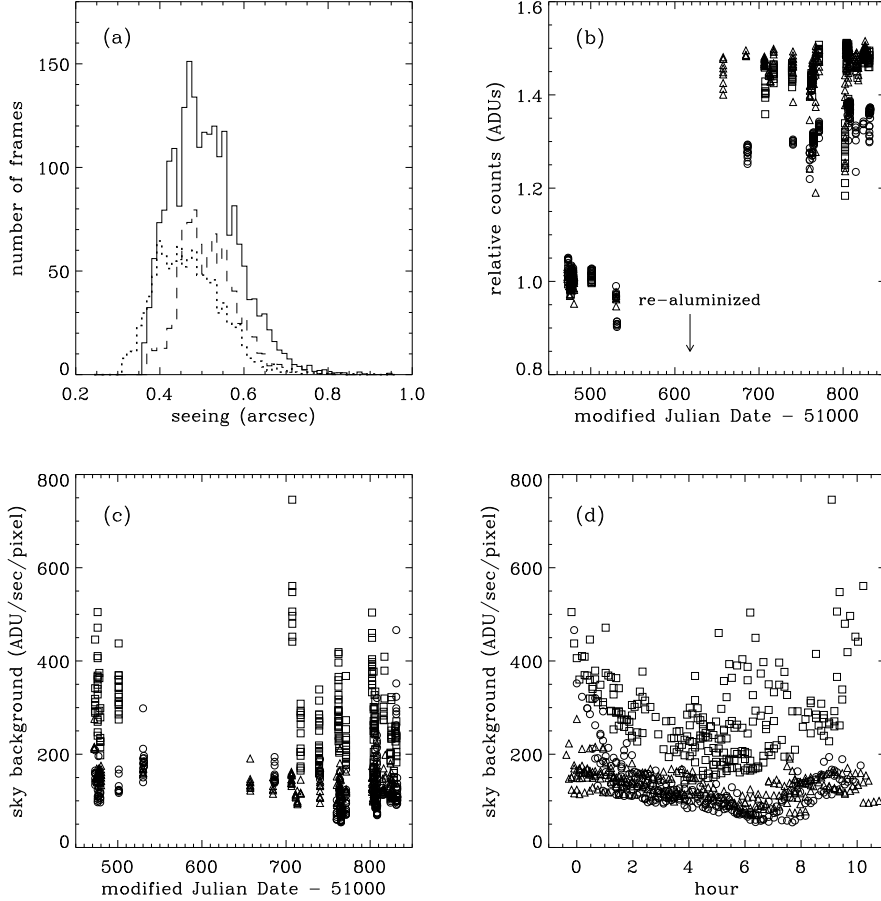


Figure 1 — Shown are the raw data in the filters J_s (dotted line or circles), H (dashed line or squares), and K_s (solid line or triangles). (a) Histogram of the median seeing in the raw ISAAC images weighted by the weight function of Eq. 2 used to combine the images. (b) Relative instrumental counts in a $\approx 3''$ radius aperture of four bright non-saturated stars in individual sky-subtracted exposures, plotted against Julian Date. The relative increase in counts, slightly dependent on wavelength, after cleaning and re-aluminization of the mirror directly reflects the increase in efficiency of the telescope, because the sky background levels (c) remained the same. Presumably, the photons were scattered by dirt rather than absorbed before cleaning. (d) Nightly sky variations are largest and most rapid in the H -band and mean sky levels are highest at the beginning and ending of the night. J_s -band varies less and peaks at the start of the night, whereas K -band levels are most stable.

over the redshift range $1 < z < 4$. The J_s filter is being established as the new standard broadband filter at $\approx 1.24\mu\text{m}$ by most major observatories (Keck, Gemini, Subaru, ESO), and is photometrically more accurate than the classical J because it is not cut off by atmospheric absorption. It is a top-hat filter with

sharp edges, practically the same effective wavelength as the normal J filter, and half-transmittance points at $1.17\mu\text{m}$ and $1.33\mu\text{m}$. We used the K_s filter which is bluer and narrower than standard K , but gives a better signal-to-noise ratio (SNR) for faint sources because it is less affected by the high thermal background of the atmosphere and the telescope. The ISAAC H and K_s filters are close to those used to establish the faint IR standard star system (Persson et al. 1998), while the J_s filter requires a small color correction. The WFPC2 filters that are used are $F300W$, $F450W$, $F606W$ and $F814W$ which we will call U_{300} , B_{450} , V_{606} and I_{814} , respectively, where the subscript indicates the central wavelength in nanometers.

The observing strategy for the HDF-S follows established procedures for ground-based NIR imaging. The dominance of the sky background and its rapid variability in the infrared requires dithering of many short exposures. We used a $20''$ jitter box in which the telescope is moved in a random pattern of Poissonian offsets between successive exposures. This jitter size is a trade-off between keeping a large area at maximum depth and ensuring that each pixel has sufficient exposures on sky. Individual exposures have integration times of 6×30 s in J_s , 6×20 s in H , and 6×10 s in K_s (subintegrations $\times$ detector integration times). We requested service mode observations amounting to 32 hours in each band with a seeing requirement of $\lesssim 0''.5$, seeing conditions that are only available 25% of the time at Paranal. The observations were grouped in 112 observation blocks (OBs), each of which uniquely defines a single observation of a target, including pointing, number of exposures in a sequence, and filter. The calibration plan for ISAAC provides the necessary calibration measurements for such blocks, including twilight flats, detector darks, and nightly zero points by observing LCO/Palomar NICMOS standard stars (Persson et al. 1998).

2.2 Observations

The HDF-S was observed from October to December 1999 and from April to October 2000 under ESO program identification 164.O-0612(A). A summary of the observations is shown in 1. We obtained a total of 33.6, 32.3 and 35.6 hours in J_s , H and K_s , distributed over 33, 34 and 55 OBs, or 1007, 968 and 2136 frames, respectively. This represents all usable data, including aborted and re-executed OBs that were outside weather specifications or seeing constraints. In the reduction process these data are included with appropriate weighting (see section 3.4). Sixty-eight percent of the data was obtained under photometric conditions and the average airmass of all data was 1.25. A detailed summary of observational parameters with pointing, observation date, image quality and photometric conditions can be found on the FIRES homepage on the World Wide Web (<http://www.strw.leidenuniv.nl/~fires>).

An analysis of various observational parameters reveals some surprising trends in the data, whereas other expected relations are less apparent. An overview is given in Figure 1. The median seeing on the raw images is better than $0''.5$ in all bands, with the seeing of 90% of the images in the range $0''.4 - 0''.65$, as can be

seen in Figure 1.

Seeing may vary strongly on short timescales but it is not related to any other parameter. The most drastic trend in the raw data is the change of sensitivity with date. Since the cleaning and re-alumization of the primary mirror in March 2000 the count rates of bright stars within a $\approx 3''$ aperture increased by +29% in J_s , +45% in H , and +45% K_s , which is reflected by a change in zero points before and after this date. Because the average NIR sky levels remained the same, this increase proportionally improved the achievable signal-to-noise for background-limited sources. The change in throughput was caused by light scattering, which explains why the sky level remained constant. Sky levels in J_s and H , dominated by airglow from OH-emission lines in the upper atmosphere (typically 90 km altitude), vary unpredictably on the timescale of minutes, but also systematically with observed hour. The average sky level is highest at the beginning and end of each night with peak-to-peak amplitudes of the variation being 50% relative to the average sky brightnesses over the night. The background in K_s is dominated by thermal emission of the telescope, instrument, and atmosphere and is mainly a function of temperature. The K_s background is the most stable of all NIR bands and only weakly correlated with airmass; our data do not show a strong thermal atmospheric contribution, which should be proportional to atmospheric path length. We take into account the variations of the background and seeing through weighting in the data coadding process.

3 Data Reduction

The reduction process included the following steps: quality verification, flat-fielding, bad pixel correction, sky subtraction, distortion correction, registration, photometric calibration and weighting of individual frames, and combination into a single frame. We used a modified version of the DIMSUM¹ package and standard routines in IRAF² for sky subtraction and coadding, and the ECLIPSE³ package for creating the flatfields and the initial bad pixel masks. We reduced the ISAAC observations several times with an increasing level of sophistication, applying corrections to remove instrumental features, scattered light, or clear artifacts when required. Here we describe the first version of the reduction (v1.0) and the last version (v3.0), leaving out the intermediate trial versions. The last version produced the final J_s , H and K_s images, on which the photometry (see section 5) and analysis (see section 8) is based.

¹DIMSUM is the Deep Infrared Mosaicing Software package developed by Peter Eisenhardt, Mark Dickinson, Adam Stanford, and John Ward, and is available via ftp to <ftp://iraf.noao.edu/contrib/dimsumV2/>

²IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the AURA, Inc., under cooperative agreement with the NSF.

³ECLIPSE is an image processing package written by N. Devillard, and is available at <ftp://ftp.hq.eso.org/pub/eclipse/>

3.1 Flatfields and Photometric Calibration

We constructed flatfields from images of the sky taken at dusk or dawn, grouped per night and in the relevant filters, using the **flat** routine in **ECLIPSE**, which also provided the bad pixel maps. We excluded a few flats of poor quality and flats that exhibited a large jump between the top row of the lower and bottom row of the upper half of the array, possibly caused by the varying bias levels of the Hawaii detector. We averaged the remaining nightly flats per month, and applied these to the individual frames of the OBs taken in the same month. If no flatfield was available for a given month we used an average flat of all months. The stability of these monthly flats is very good and the structure changes little and in a gradual way. We estimate the relative accuracy to be $0.2 - 0.4\%$ per pixel from the pixel-to-pixel rms variation between different monthly flats. Large scale gradients in the monthly flats do not exceed 2% . We checked that standard stars, which were observed at various locations on the detector, were consistent within the error after flatfielding.

Standard stars in the LCO/Palomar NICMOS list (Persson et al. 1998) were observed each night, in a wide five-point jitter pattern. For each star, on each night, and in each filter, we measured the instrumental counts in a circular aperture of radius 20 pixel ($2''.94$) and derived zero points per night from the magnitude of that star in the NICMOS list. We identify non-photometric nights after comparison with the median of the zero points over all nights before and after re-aluminization in March 2000 (see section 2.2). The photometric zero points exhibit a large increase after March 2000 but, apart from this, the night-to-night scatter is approximately 2% . We adopted the mean of the zero points after March 2000 as our reference value. See Table 2 for the list of the adopted zero points. By applying the nightly zero points to 4 bright unsaturated stars in the HDF-S, observed on the same night under photometric conditions, we obtain calibrated stellar magnitudes with a night-to-night rms variation of only $\approx 1 - 1.5\%$. No corrections for atmospheric absorption were required because the majority of the science data were obtained at similar airmass as the standard star observations. In addition, instrumental count rates of HDF-S stars in individual observation blocks reveal no correlation with airmass. We used the calibrated magnitudes of the 4 reference stars, averaged over all photometric nights, to calibrate every individual exposure of the photometric and non-photometric OBs. The detector non-linearity, as described by Amico et al. (2001), affects the photometric calibration by $\lesssim 1\%$ in the H -band, where the exposure levels are highest. Because the effect is so small, we do not correct for this. We did not account for color terms due to differences between the ISAAC and standard filter systems. Amico et al. (2001) report that the ISAAC H and K_s filters match very well those used to establish the faint IR standard star system of Persson et al. (1998). Only the ISAAC J_s filter is slightly redder than Persson's J and this may introduce a small color term, $\approx -0.04 \cdot (J - K)_{LCO}$. However, the theoretical transformation between ISAAC magnitudes and those of LCO/Palomar have never been experimentally verified.

Furthermore, the predicted color correction is small and could not be reproduced with our data. In the absence of a better calibration we chose not to apply any color correction. We did apply Galactic extinction correction when deriving the photometric redshifts, see section 6, but it is not applied to the catalog.

As a photometric sanity check, we compared $2''$ circular diameter aperture magnitudes of the brightest stars in the final (version 3.0, described below) images to magnitudes based on a small fraction of the data presented by R01. Each data set was independently reduced, the calibration based on different standard stars, and the shallower data were obtained before re-aluminization of the primary mirror. The magnitudes of the brightest sources in all bands agree within 1% between the versions, indicating that the internal photometric systematics are well under control. For the NIR data, the adopted transformations from the Johnson (1966) system to the AB system are taken from Bessell & Brett (1988) and we apply $J_{s,AB} = J_{s,Vega} + 0.90$, $H_{AB} = H_{Vega} + 1.38$ and $K_{s,AB} = K_{s,Vega} + 1.86$.

3.2 Sky Subtraction and Cosmic Ray Removal

The rapidly varying sky, typically 25 thousand times brighter than the sources we aim to detect, is the primary limiting factor in deep NIR imaging. In the longest integrations, small errors in sky subtraction can severely diminish the achievable depth and affect faint source photometry. The IRAF package DIMSUM provides a two-pass routine to optimally separate sky and astronomical signal in the dithered images. We modified it to enable handling of large amounts of data and replaced its co-adding subroutine, which assumes that the images are undersampled, by the standard IRAF task IMAGES.IMMATCH.IMCOMBINE. The following is a brief summary of the steps performed by the REDUCE task in DIMSUM.

For every science image in a given OB a sky image is constructed. After scaling the exposures to a common median, the sky is determined at each pixel position from a maximum of 8 and a minimum of 3 adjacent frames in time. The lowest and highest values are rejected and the average of the remainder is taken as the sky value. These values are subtracted from the scaled image to create a sky subtracted image. A set of stars is then used to compute relative shifts, and the images are integer registered and averaged to produce an intermediate image. All astronomical sources are identified and a corresponding object mask is created. This mask is used in a second pass of sky subtraction where pixels covered by objects are excluded from the estimate of the sky.

The images show low-level pattern due to bias variations. Because they generally reproduce they are automatically removed in the skysubtraction step. We find cosmic rays with DIMSUM, using a simple threshold algorithm and replacing them by the local median, unless a pixel is found to have cosmic rays in more than frames per OB. In this case the pixel is added to the bad pixel map for that OB.

3.3 First Version and Quality Verification

The goal of the first reduction of the data set is to provide a non-optimized image, which we use to validate and to assess the improvements from more sophisticated image processing. The first version consists of registration on integer pixels and combination of the sky subtracted exposures per OB. For each of the 122 OBs, we created an average and a median combined image to verify that cosmic rays and other outliers were removed correctly, and we visually inspected all 4149 individual sky subtracted frames as well, finding that many required further processing as described in the following section. Finally, we generated the version 1.0 images (the first reduction of the full data set) by integer pixel shifting all OBs to a common reference frame, and coaveraging them into the J_s , H and K_s images. While this first reduction is not optimal in terms of depth and image quality, it is robust owing to its straightforward reduction procedure.

3.4 Additional Processing and Improvements

The individual sky subtracted frames are affected by a number of problems or instrumental features, which we briefly describe below, together with the applied solutions and additional improvements that lead to the version 3.0 images. The most important problems are:

- Detector bias residuals, most pronounced at the rows where the read-out of the detector starts at the bottom (rows 1, 2, ...) and halfway (rows 513, 514, ...), caused by the complex bias behaviour of the Rockwell Hawaii array. These variations are uniform along rows, and we removed the residual bias by subtracting the median along rows in individual sky subtracted exposures, after masking all sources.
- Imperfect sky subtraction, caused by stray light or rapid background variations. Strong variations in the backgrounds, reflection from high cirrus, reflected moonlight in the ISAAC optics or patterns of less obvious origin can lead to large scale residuals in the sky subtraction, particularly in J_s and H . For some OBs, we successfully removed the residual patterns by splitting the sequence in two (in case of a sudden appearance of stray light), or subtracting a two-piece cubic spline fit along rows and columns to the background in individual frames, after masking all sources. We rejected a few frames, or masked the affected areas, if this simple solution did not work.
- Unidentified cosmic rays or bad pixels. A small amount of bad pixels were not detected by ECLIPSE or DIMSUM routines but need to be identified because we average the final images without additional clipping or rejection. By combining the sky subtracted frames in a given OB without shifts and with the sources masked, we identified remaining cosmic rays or outliers through sigma clipping. We added $\sim 60 - 100$ pixels per OB to the corresponding bad pixel map.

Several steps were taken to improve the quality and limiting depths of the version 1.0 images, the most important of which are:

- Distortion correction of the individual frames and direct registration to the 3×3 blocked I_{814} image ($0.119'' \text{pixel}^{-1}$), our preferred frame of reference. We obtained the geometrical distortion coefficients for the 3rd order polynomial solution from the ISAAC WWW-page⁴. The transformation procedure involves distortion correcting the ISAAC images, adjusting the frame-to-frame shifts, and finding the linear transformation to the WFPC2 I_{814} frame of reference. This linear transformation is the best fit mapping of source positions in the blocked WFPC2 I_{814} image to the corresponding positions in the corrected J_s -band image⁵. Compared to version 1.0 described in the previous section, this procedure increases registration accuracy and image quality, decreases image smearing at the edges introduced by the jittering and differential distortion. Given the small amplitude of the ISAAC field distortions, the effect on photometry is negligible. In the linear transformation and distortion correction step the image is resampled once using a third-order polynomial interpolation, with a minimal effect of the interpolant on the noise properties.
- Weighting of the images. We substantially improved the final image depth and quality by assigning weights to individual frames that take into account changes in seeing, sky transparency, and background noise. Two schemes were applied: one that optimizes the signal-to-noise ratio (SNR) within an aperture of the size of the seeing disk, and one that optimizes the SNR per pixel. The first improves the detection efficiency of point sources, the other optimizes the surface brightness photometry. The weights w_i of the frames are proportional to either the inverse scaled variance $zpscale_i \times var_i$ within a seeing disk of size s_i , or to the inverse scaled variance per pixel, where the scaling $zpscale_i$ is the flux calibration applied to bring the instrumental counts of our four reference stars in the HDF-S to the calibrated magnitude.

$$w_{i,point} \propto (zpscale_i \times var_i \times s_i^2)^{-1} \quad (1)$$

$$w_{i,extended} \propto (zpscale_i \times var_i)^{-1} \quad (2)$$

3.5 Final Version and Post Processing

The final combined J_s , H , and K_s images (version 3.0) were constructed from the individually registered, distortion corrected, weighted and unclipped average of the 1007, 968, and 2136 NIR frames respectively. Ultimately, less than 3% of individual frames were excluded in the final images because of poor quality. In this step we also generated the weight maps, which contain the weighted exposure time per pixel. We produced three versions of the images, one with optimized

⁴ISAAC home page: <http://www.eso.org/instruments/isaac>

⁵We have noticed that the mapping solution changed slightly after the remount of ISAAC in March 2000, implying a 0.1% scale difference.

weights for point sources, one with optimized weights for surface brightness, and one consisting of the best quartile seeing fraction of all exposures, also optimized for point sources. The weighting has improved the image quality by 10–15% and the background noise by 5–10%, and distortion correction resulted in subpixel registration accuracy between the NIR images and I_{814} -band image over the entire field of view.

The sky subtraction routine in DIMSUM and our additional fitting of rows and columns (see section 3.4) have introduced small negative biases in combined images, caused by systematic oversubtraction of the sky which was skewed by light of the faint extended PSF wings or very faint sources, undetectable in a single OB. Because of this, the flatness of the sky on large scales was limited to about 10^{-5} . The negative bias was visible as clearly defined orthogonal stripes at P.A. $\approx 6^\circ$, as well as dark areas around the brightest stars or in the crowded parts of the images. To solve this, we rotated a copy of the final images back to the orientation in which we performed sky subtraction, fitted a 3-piece cubic spline to the background along rows and columns (masking all sources), re-rotated the fit, and subtracted it. The sky in the final images is flat to a few $\times 10^{-6}$ on large ($> 20''$) scales.

4 Final Images

The reduced NIR J_s , H , and K_s images and weight maps can be obtained from the FIRES-WWW homepage (<http://www.strw.leidenuniv.nl/~fires>). Throughout the rest of the paper we will only consider the images optimized for point source detection which we will use to assemble the catalog of sources.

4.1 Properties

The pixel size in the NIR images equals that of the 3×3 blocked WFPC2 I_{814} -band image at $0.119'' \text{ pixel}^{-1}$. The combined ISAAC images are aligned with the HST version 2 images (Casertano et al. 2000) with North up, and are normalized to instrumental counts per second. The images are shallower near the edges of the covered area because they received less exposure time in the dithering process, which is reflected in the weight map containing the fraction of total exposure time per pixel. The area of the ISAAC K_s -band image with weight per pixel $w_K \geq 0.95, 0.2$, and 0.01 covers $4.5, 7.2$ and 8.3 arcmin^2 , while the area used for our preferred quality cut for photometry ($w \geq 0.2$ in all seven bands) is 4.7 arcmin^2 . The NIR images have been trimmed where the relative exposure time per pixel is less than 1%.

Figure 2 shows the noise-equalized K_s -detection image obtained by division with the square root of the exposure-time map. The richness in faint compact sources and the flatness of the background are readily visible. Figure 3 shows a RGB color composite image of the I_{814} , J_s , and K_s images. The PSF of the space-

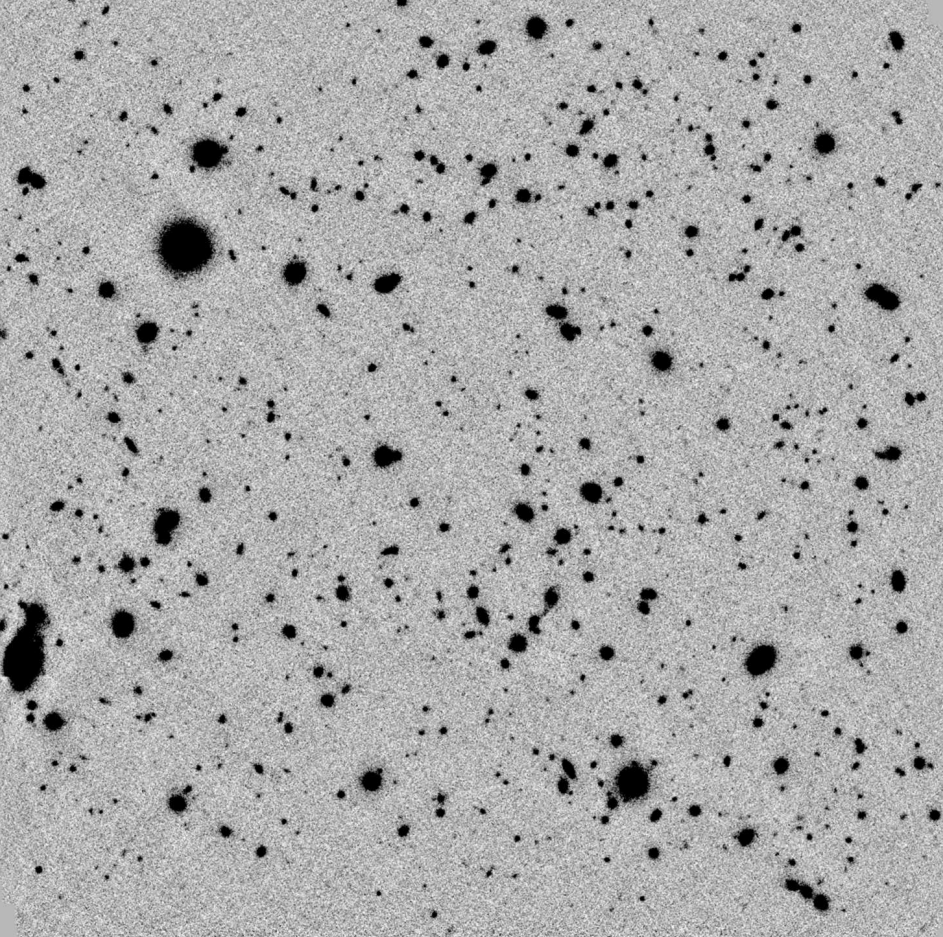


Figure 2 — The HDF-S field in the ISAAC K_s -band divided by the square root of the weight map (based on the fractional exposure time per pixel) and displayed at linear scaling. The total integration time is 35.6 hours, the stellar FWHM $\approx 0''.46$ and the total field size is $2.85' \times 2.85'$.

based I_{814} image has been matched to that of the NIR images at FWHM $\approx 0''.46$ (see section 4.2) and three adjacent WFPC2 I_{814} flanking fields have been included for visual purposes. We set the linear stretch of both images to favor faint objects. Immediately striking is the rich variety in optical-NIR colors, even for the faint objects, indicating that the NIR observations are very deep and that there is a wide range of observed spectral shapes, which can result from different types of galaxies over a broad redshift range.

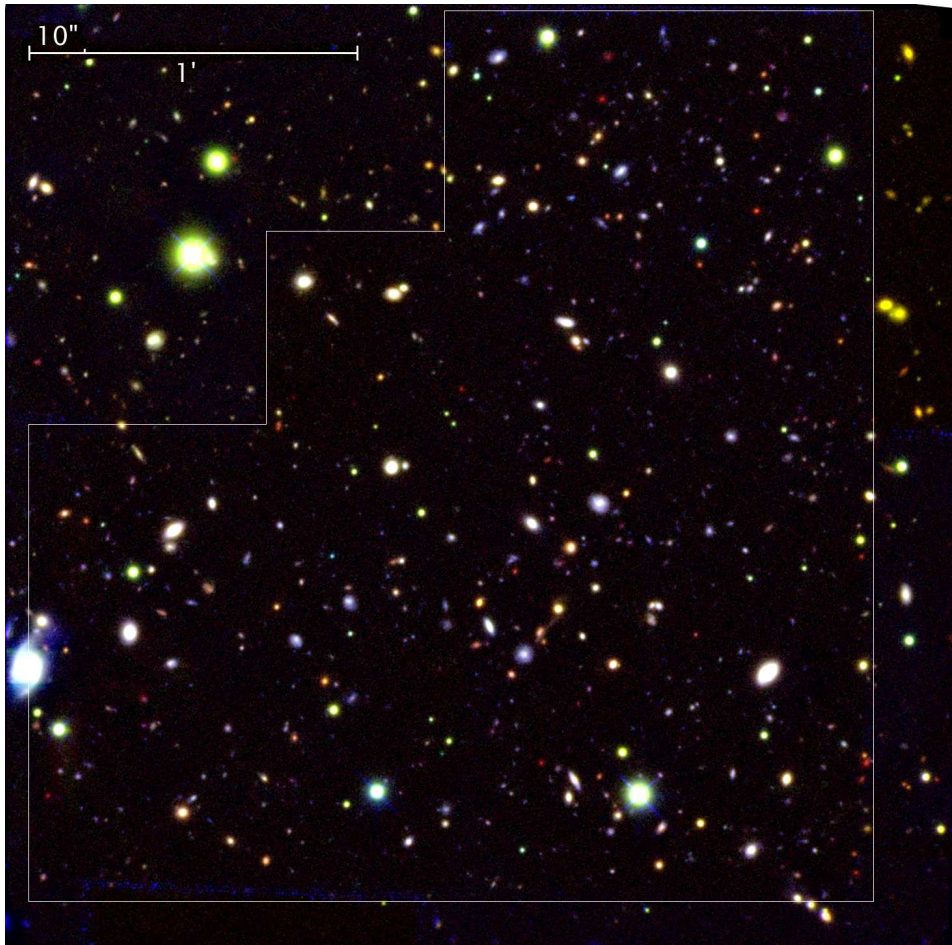


Figure 3 — Three-color composite image of the ISAAC field on top of the WFPC2 main-field and parts of three WFPC2 flanking fields. The main-field is outlined in white and North is up. The images are registered and smoothed to a common seeing of $\text{FWHM} \approx 0''.46$, coding WFPC2 I_{814} in blue, ISAAC J_s in green and ISAAC K_s in red. There is a striking variety in optical-to-infrared colors, especially for fainter objects. A number of sources with red colors have photometric redshifts $z > 2$ and they are candidates for relatively massive, evolved galaxies. These galaxies would not be selected by the U-dropout technique because they are too faint in the observer's optical.

4.2 Image Quality

The NIR PSF is stable and symmetric over the field with a gaussian core profile and an average ellipticity < 0.05 over the J_s , H , and K_s images. The median FWHM of the profiles of ten selected isolated bright stars is $0''.45$ in J_s , $0''.48$ H , and $0''.46$ K_s with $0''.04$ amplitude variation over the images.

For consistent photometry in all bands we convolved the measurement images to a common PSF, corresponding to that of the H -band which had worst effective seeing ($\text{FWHM} = 0''.48$). The similarity of PSF structure across the NIR images allowed simple gaussian smoothing for a near perfect match. The complex PSF structure of the WFPC2 requires convolving with a special kernel, which we constructed by deconvolving an average image of bright isolated non-saturated stars in the H -band with the average I_{814} -band image of the same stars. Division of the stellar growth curves of the convolved images by the H -band growth curve shows that the fractional enclosed flux agrees to within 3% at radii $r \geq 0''.35$.

4.3 Astrometry

The relative registration between ISAAC and WFPC2 images needs to be very precise, preferably a fraction of an original ISAAC pixel over the whole field of view, to allow correct cross-identification of sources, accurate color information and morphological comparison between different bands. To verify our mapping of ISAAC to WFPC2 coordinates, we measured the positions of the 20 brightest stars and compact sources in all registered ISAAC exposures, and we compared their positions with those in the I_{814} image. The rms variation in position of individual sources is about $0.2 - 0.3$ pixel at $0.119'' \text{ pixel}^{-1}$ ($25 - 35$ mas), but for some sources systematic offsets between the NIR and the optical up to 0.85 pixel (100 mas) remain. The origin of the residuals is unclear and we cannot fit them with low order polynomials. They could be real, intrinsic to the sources, or due to systematic errors in the field distortion correction of ISAAC or WFPC2⁶. However, for all our purposes, the effect of positional errors of this amplitude is unimportant. The error in absolute astrometry of the HST HDF-S coordinate system, estimated to be less than 40 mas, is dominated by the systematic uncertainty in the positions of four reference stars (Casertano et al. 2000; Williams et al. 2000).

4.4 Backgrounds and Limiting Depths

The noise properties of the raw individual ISAAC images are well described by the variance of the signal collected in each pixel since both Poisson and read noise are uncorrelated. However, image processing, registration and combination have introduced correlations between neighbouring pixels and small errors in the background subtraction may also contribute to the noise. Understanding the noise properties well is crucial because limiting depths and photometric uncertainties rely on them.

Instead of a formal description based on the analysis of the covariance of correlated pixel pairs, we followed an empirical approach where we fit the dependence

⁶The ISAAC field distortion might have changed over the years, but this cannot be checked because recent distortion measurements are unavailable. The worst case errors of relative positions across the four WFPC2 chips can be $0''.1$ (Vogt et al. 1997), but is expected to be smaller for the HDF-S images.

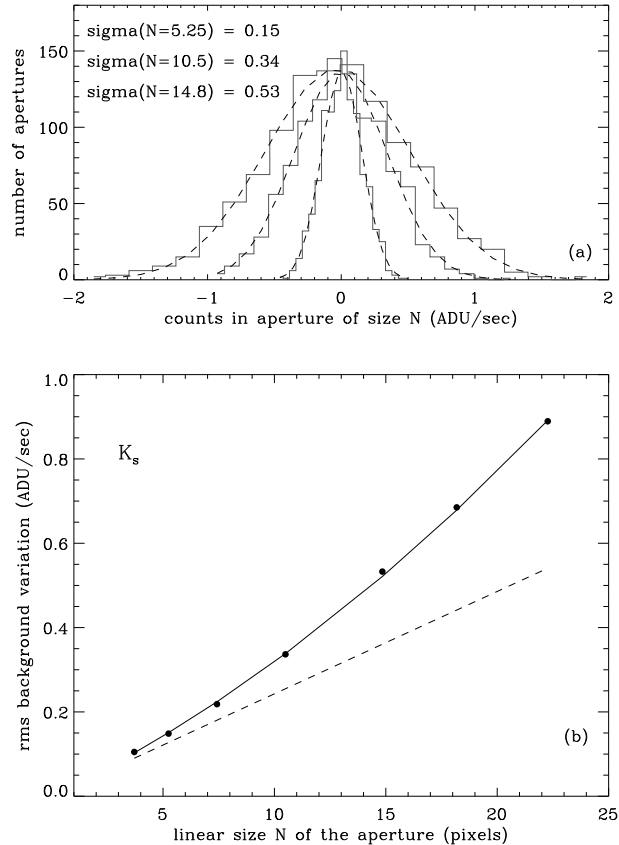


Figure 4 — Scaling relation of the measured background rms noise as a function of linear size $N = \sqrt{A}$ of apertures with area A . (a) Gaussians are fitted to histograms of K_s counts in randomly placed apertures of increasing size, excluding pixels belonging to sources. This correctly accounts for pixel-to-pixel correlations and other effects, allowing us to measure the true rms variation as a function of linear size of aperture. (b) The K_s -band results (solid points), together with the best-fit scaling relation of Eq. 3 (solid line), show that the measured variation in large apertures exceeds the variation expected from linear (Gaussian) scaling of the pixel-to-pixel noise (dashed line), likely due to large scale correlated fluctuations of the background.

of the rms background variation in the image as a function of linear size $N = \sqrt{A}$ of apertures with area A . Directly measuring the effective flux variations in apertures of different sizes provides a more realistic estimate of signal variations than formal Gaussian scaling $\sigma(N) = N\bar{\sigma}$ of the pixel-to-pixel noise $\bar{\sigma}$, as is often done.

We measured fluxes in 1200 non-overlapping circular apertures randomly placed on the registered convolved images, which were also used for photometry. We ex-

cluded all pixels belonging to sources detectable in K_s at the 5σ level (see section 5.1 for detection criteria). We used identical aperture positions for each band i and measured fluxes for circular aperture diameters ranging from $0''.5$ to $3''$. Then we obtained the flux dispersions by fitting a Gaussian distribution to the histogram of fluxes at each aperture size. Finally, we fitted a parameterized function of linear size to the different dispersions:

$$\sigma_i(N) = N\bar{\sigma}_i(a_i + b_iN)/\sqrt{w_i} \quad (3)$$

This equation describes the signal variation versus aperture size N over the entire image, taking into account spatial variations as a result of relative weight w_i for each passband i . As can be seen in Figure 4, it provides a good fit to the noise characteristics. The noise is significantly higher than expected from uncorrelated (Gaussian) noise, indicated by a dashed line in Figure 4b. Table 3 shows the best fit values in all bands and the corresponding limiting depths. The parameter a reflects the correlations of neighbouring pixels ($a > 1$), which is important in the WFPC2 images because of heavy smoothing, but also in the ISAAC images given the resampling from $0''.147$ to $0''.119 \text{ pixel}^{-1}$. The parameter b accounts for large scale correlated variations in the background ($b > 0$). This may be caused by the presence of sources at very faint flux levels (confusion noise) or instrumental features. Typically, the large scale correlated contribution per pixel is only 3-15% relative to the gaussian rms variation, but due to the N^2 proportionality the contribution to the variation in large apertures increases to significant levels. While the signal variations grow faster with area than expected from a Gaussian, at any specific scale the variation is consistent with a pure Gaussian.

From the analysis of the scaling relation of simulated colors we find that part of the large scale irregularities in the background are spatially correlated between bands. In particular, we measured the rms variation of the $I_{814} - V_{606}$ colors directly by subtracting in registered apertures the I_{814} -band fluxes from the V_{606} fluxes and fitting the dispersion of the difference at each linear size. On large scales rms variations are 30% smaller than predicted from Eq. 3 if the noise were uncorrelated. Yet, if we subtract the two fluxes in random apertures, the scaling of the background variation is consistent with the prediction. A similar effect is seen for the $I_{814} - J_s$ color, but at a smaller amplitude. The spatial coherence of the background variations between filters and across cameras suggests that part of the background fluctuations may be associated with sources at very faint flux levels. Other contributions are likely similar flatfielding or skysubtraction residuals from one band to another.

5 Source Detection and Photometry

The detection of sources at very faint magnitudes against a noisy background forces us to trade off completeness and reliability. A very low detection threshold may

generate the most complete catalog, but we must then apply additional criteria to assess the reliability of each detection given that such a catalog will contain many spurious sources. More conservatively, we choose the lowest possible threshold for which contamination by noise is unimportant. We aim to produce a catalog with reliable colors suitable for robustly modeling of the intrinsic spectral energy distribution. Using SExtractor version 2.2.2 (Bertin & Arnouts 1996) with a detection procedure that optimizes sensitivity for point-like sources, we construct a K_s -band selected catalog with seven band optical-to-infrared photometry.

5.1 Detection

To detect objects with SExtractor using a constant signal-to-noise criterion over the entire image, including the shallower outer parts, we divide the point source optimized K_s -image by the square root of the weight (exposure time) map to create a noise-equalized detection image. A source enters the catalog if, after low-pass filtering of the detection image, at least one pixel is above ≈ 5 times the standard deviation of the filtered background, corresponding to a total K_s -band magnitude limit for point sources of $K_s \approx 26.0$. This depth is reached for the central 4.5 arcmin². In total we have 833 detections in the entire survey area of 8.3 arcmin². Initially 820 sources are found but the detection software fails to detect sources lying in the extended wings of the brightest objects. To include these, we fit the surface brightness profiles of the brightest sources with the GALPHOT package (Franx et al. 1989) in IRAF, subtract the fit, and carry out a second detection pass with identical parameters. Thirteen new objects enter the catalog, and 9 sources detected in the first pass are replaced with improved photometry. The catalog identification numbers of all second-pass objects start at 10001, and the original entries of the updated sources are removed.

Filtering affects only the detection process and the isophotal parameters; other output parameters are affected only indirectly through barycenter and object extent. We chose a simple two-dimensional gaussian detection filter (FWHM= 0".46), approximating the core of the effective K_s -band PSF well. Hence, we optimize detectability for point-like sources, introducing a small bias against faint extended objects. In principle it is possible to combine multiple catalogs created with different filter sizes but merging these catalogs consistently is a complicated and subjective process yielding a modest gain only in sensitivity for larger objects. We prefer the small filter size equal to the PSF in the detection map because the majority of faint sources that we detect are compact or unresolved in the NIR and because we wish to minimize the blending effect of filtering on the isophotal parameters and on the confusion of sources. SExtractor applies a *multi-thresholding* technique to separate overlapping sources based on the distribution of the filtered K_s -band light. About 20% of the sources are blended because of the low value of the isophotal threshold; in SExtractor this must always equal the detection threshold. With the deblending parameters we used, the algorithm succeeds in splitting close groups of separate galaxies, without “oversplitting” galaxies with

rich internal structure.

We tested sensitivity to false detections by running SExtractor on a specially constructed K_s -band noise map created by subtracting, in pairs, individual K_s -band images of comparable seeing, after zero point scaling, and coaveraging the weighted difference images. This noise image has properties very similar to the noise in the original reduced image, including contributions from the detector and reduction process, but with no trace of astronomical sources. Our detection algorithm resulted in only 11 spurious sources over the full area.

5.2 Optical and NIR Photometry

We use SExtractor’s dual-image mode for spatially accurate and consistent photometry, where objects are detected and isophotal parameters are determined from the K_s -band detection image while the fluxes in all seven bands are measured in the registered and PSF matched images. We used fluxes measured in circular apertures $APER(D)$ with fixed-diameters D , isophotal apertures $APER(ISO)$ determined by the K_s -band detection isophote at the 5σ detection threshold, and $APER(AUTO)$ (autoscaling) apertures inspired by Kron (1980), which scales an elliptical aperture based on the first moments of the K_s -band light distribution. We select for each object the best aperture based on simple criteria to enable detailed control of photometry. We define two types of measurements:

- “color” flux, to obtain consistent and accurate colors. The optimal aperture is chosen based on the K_s flux distribution, and this aperture is used to measure the flux in all other bands.
- “total” flux, only in the K_s -band, which gives the best estimate of the total K_s flux.

For both measurements we treat blended sources differently from unblended ones, and consider a source blended when its `BLENDED` or `BIAS` flag is set by SExtractor, as described in Bertin & Arnouts (1996).

Our color aperture is chosen as follows, introducing the equivalent of a circular isophotal diameter $D_{iso} = 2\sqrt{A_{iso}/\pi}$ based on A_{iso} , the measured non-circular isophotal area within the detection-isophote:

if unblended

$$APER(COLOR) = \begin{cases} APER(ISO) & (0''.7 < D_{iso} < 2''.0) \\ APER(0''.7) & (D_{iso} \leq 0''.7) \\ APER(2''.0) & (D_{iso} \geq 2''.0) \end{cases}$$

if blended

$$APER(COLOR) = \begin{cases} APER(D_{iso}/s) & (0''.7 < D_{iso}/s < 2''.0) \\ APER(0''.7) & (D_{iso}/s \leq 0''.7) \\ APER(2''.0) & (D_{iso}/s \geq 2''.0) \end{cases} \quad (4)$$

The parameter s is the factor with which we shrink the circular apertures centered on blended sources, increasing the separation to the blended neighbour such that mutual flux contamination is minimal. This factor depends on the data set, and for our ISAAC K_s image we find that $s = 1.4$ is most successful. The smallest aperture considered, $APER(0''.7)$, ≈ 1.5 FWHM of the effective PSF, optimizes the S/N for photometry of point sources in unweighted apertures and prevents smaller more error-prone apertures. The largest allowed aperture, $APER(2''.0)$, prevents large and inaccurate isophotal apertures driven by the filtered K_s -light distribution. We continuously assessed the robustness and quality of color flux measurements by inspecting the fits of redshifted galaxy templates to the flux points, as described in detail in §6.

We calculate the total flux in the K_s -band from the flux measured in the AUTO aperture. We define a circularized AUTO diameter $D_{auto} = 2\sqrt{A_{auto}/\pi}$ with A_{auto} the area of the AUTO aperture, and define the total magnitude as:

if unblended

$$APER(TOTAL) = APER(AUTO)$$

if blended

$$APER(TOTAL) = APER(COLOR) \quad (5)$$

Finally, we apply an aperture correction using the growth curve of brighter stars to correct for the flux lost because it fell outside the “total” aperture. This aperture correction is necessary because it is substantial for our faintest sources, as shown in Figure 5 where we compare different methods to estimate magnitude. The aperture correction reaches 0.7 mag at the faint end, therefore magnitudes are seriously underestimated if the aperture correction is ignored.

We derive the 1σ photometric error for all measurements from Eq. 3 with the best-fit values shown in Table 3. These errors may overestimate the uncertainty in colors of adjacent bands (see section 4.4) but it should represent well the photo-

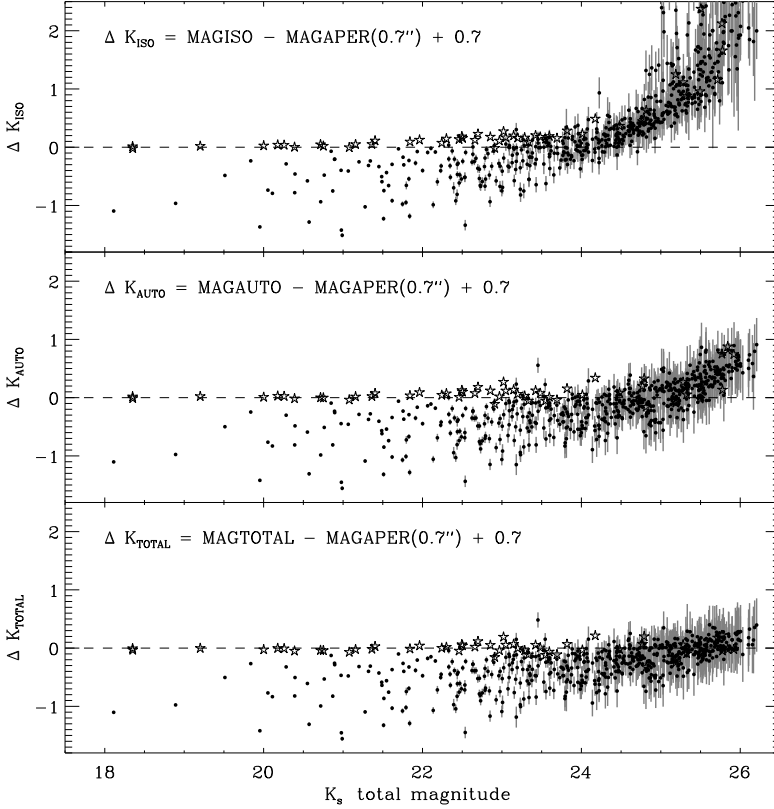


Figure 5 — Comparison of methods to estimate total K_s -band magnitude. Shown are isophotal (*top*), SExtractor’s auto-scaling AUTO (*middle*), and our “total” magnitudes (*bottom*) as defined in Eq. 5 and which are aperture corrected using stellar growth curve analysis. We subtracted the aperture corrected magnitude measured in an aperture of $0''.7$ ($MAGAPER(0''.7) - 0.7$), which produces the correct total magnitudes for stars and pointlike sources. Stars are marked by star symbols and fluxes are plotted with $\pm 1\sigma$ error bars. The turn-up at $K_s \approx 24$ of isophotal and at $K_s \approx 25$ of the AUTO magnitudes shows that these photometric schemes systematically underestimate the total flux at faint levels, due to the decreasing size of the used aperture with magnitude. This effect is nearly absent in the bottom panel, which shows the total magnitudes measured in this paper.

metric error over the entire $0.3\mu - 2.2\mu\text{m}$ wavelength range. The magnitudes may suffer from additional uncertainties because of surface brightness biases or possible biases in the sky subtraction procedure which could depend on object magnitude and size.

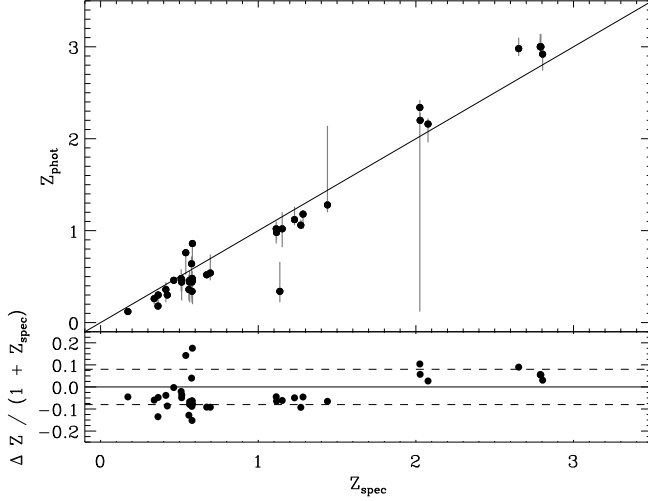


Figure 6 — Direct comparison of photometric redshifts to the 39 spectroscopic redshifts of objects in the HDF-S with good photometry in all bands. The 68% error bars are derived from our Monte Carlo simulations and the diagonal line corresponds to a one-to-one relation to guide the eye. While the agreement is excellent with no failures for this small sample and with mean $\Delta z/(1+z) = 0.08$, large asymmetric uncertainties remain for some objects indicating the presence of a second photometric redshift solution of comparable likelihood at a different redshift.

6 Photometric Redshifts

To physically interpret the seven-band photometry for our K_s -band selected sample, we use a photometric redshift (z_{phot}) technique explained in detail by R01. In summary, we correct the observed flux points for Galactic extinction (see Schlegel, Finkbeiner, & Davis 1998) and we model the rest-frame colors of the galaxies by fitting a linear combination of redshifted empirical galaxy templates. The redshift with the lowest χ^2 statistic, where

$$\chi^2(z) = \sum_{i=1}^{N_{filter}} \left[\frac{F_i^{data} - F_i^{model}}{\sigma_i^{data}} \right]^2 \quad (6)$$

is then chosen as the most likely z_{phot} . Using a linear combination of SEDs as F^{model} minimizes the a priori assumptions about the nature and stellar composition of the detected sources.

Our data set with three deep NIR bands samples the position of Balmer/4000Å break over $1 \lesssim z \lesssim 4$, allowing us to probe the redshift distribution of more evolved galaxy types that may have little rest-frame UV flux and hence a weak or virtually absent Lyman break.

6.1 Photometric Templates

We used the local Hubble type templates E, Sbc, Scd, and Im from Coleman, Wu, & Weedman (1980), the two starburst templates, SB1 and SB2, with low derived reddening from Kinney et al. (1996), and a 10 Myr old single age template model from Bruzual & Charlot (2002). The starburst templates are needed because many galaxies even in the nearby Universe have bluer colors than the bluest CWW templates. The observed templates are extended beyond their published wavelengths into the far-ultraviolet by power law extrapolation and into the NIR using stellar population synthesis models from Bruzual & Charlot (2002), with the initial mass functions and star-formation timescales for each template Hubble type from Pozzetti, Bruzual, & Zamorani (1996). We accounted for internal hydrogen absorption of each galaxy by setting the flux blueward of the 912Å Lyman limit to zero, and for the redshift-dependent cosmic mean opacity due to neutral intergalactic hydrogen by following the prescriptions of Madau (1995).

6.2 Z_{phot} Uncertainties

The best test of photometric redshifts is direct comparison to spectroscopic redshifts, but spectroscopic redshifts in the HDF-S are still scarce. We calculate the uncertainty in the photometric redshift due to the flux measurement errors using a Monte-Carlo (MC) technique derived from that used in R01 and fully explained in Rudnick et al. (2002b). At bright magnitudes template mismatch dominates the errors, something that is not modeled by the MC simulation. Hence, the MC error bars for bright galaxies are severe underestimates. At fainter magnitudes, the uncertainty is driven by errors in photometry (Fernández-Soto, Lanzetta, & Yahil 1999) and the MC technique should provide accurate z_{phot} uncertainties. Experience from R01 showed that two ways to correct for the template mismatch, setting a minimum fractional flux error or setting a minimum z_{phot} error based on the mean disagreement with z_{spec} , either degrade the accuracy of the z_{phot} measurement or reflect the systematic error only in the mean, while template mismatch can be a strong function of SED shape and redshift. A method based completely on Monte-Carlo techniques is preferable because it has a straightforwardly computable redshift probability function. This approach is desirable for estimating the rest-frame luminosities and colors (Rudnick et al. 2002b).

Therefore, we modify the MC errors directly using the FIRES photometry. In summary, we estimate the systematic component of the z_{phot} uncertainty by scaling up all the photometric errors for a given galaxy with a constant to bring residuals of the fit in agreement with the errors. This will not change the best fit redshift and SED and will not modify the MC errorbars of faint objects, but it will enlarge the redshift interval over which the templates can satisfactorily fit the bright objects. Only in case of widely different photometric errors between the visible and infrared might the modified MC uncertainties still underestimate the true z_{phot} uncertainty.

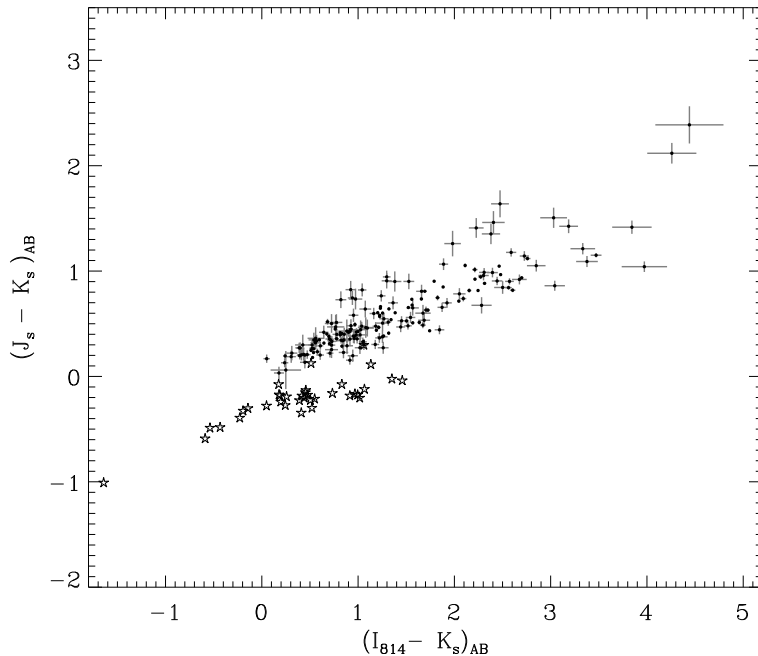


Figure 7 — $J_s - K_s$ versus $I_{814} - K_s$ color-color diagram (on the AB system) for the sources with $K_s < 24$ in the HDF-S with a minimum of 20% of the total exposure time in all bands. Identified stars are marked by a star symbol. The colors are plotted with $\pm 1\sigma$ error bars. There is a large variation in both $I - K_s$ and $J_s - K_s$ colors. Redshifted galaxies are well separated from the stellar locus in color-color space.

In Figure 6 we show a direct test of photometric redshifts of the 39 objects in the HDF-S with available spectroscopy and good photometry in all bands. The current set of spectroscopic redshifts in the HDF-S will appear in Rudnick et al. (2002a). For the small sample that we can directly compare, we find excellent agreement with no failures and with a mean $\Delta z / (1 + z_{spec}) \approx 0.08$ with $\Delta z = |z_{spec} - z_{phot}|$. It is encouraging to see that the modified 68% error bars that were derived from the Monte Carlo simulations are consistent with the measured disagreement between z_{phot} and the z_{spec} in the HDF-S. However, large asymmetric uncertainties remain for some objects, clearly showing the presence of a second photometric redshift solution of comparable likelihood at a vastly different redshift, revealing limits on the applicability of the photometric redshift technique.

6.3 Stars

In a pencil beam survey at high Galactic latitude such as the HDF-S, a limited number of foreground stars are expected. We identify stars as those objects which

have a better raw χ^2 for a single stellar template fit than the χ^2 for the galaxy template combination. The stellar templates are the NEXTGEN model atmospheres from Hauschildt et al. (1999) for main sequence stars with temperatures of 3000 to 10000K, assuming local thermodynamical equilibrium (LTE). Models of cooler and hotter stars cannot be included because non-LTE effects are important. We checked the resulting list of stars using the FWHM in the original B_{450} -band image and the $J_s - K_s$ color, excluding two objects (catalog IDs 207 and 296) that were obviously extended in B_{450} , and we find a total of 57 stars. As shown in Figure 7, most galaxies are clearly separated from the stellar locus in $I_{814} - K_s$ versus $J_s - K_s$ color-color space. Other cooler stars might still resemble SEDs of redshifted compact galaxies but the latter are generally redder in the infrared $J_s - K_s$ than most known M or methane dwarfs. Known cool L-dwarfs fall along a redder extension of the track traced by M-dwarfs in color-color space and have progressively redder $J_s - K_s$ colors for later spectral types. However measurements by Kirkpatrick et al. (2000) show the L-dwarf sequence abruptly stopping at $(J_s - K_s)_J \approx 2.1$ (the subscript noting Johnson magnitudes, see section 3.1 for the transformations to the AB system) whereas even cooler T-dwarfs have much bluer $(J_s - K_s)_J \approx 0$ colors than expected from their temperatures due to strong molecular absorption. This is important because if we would apply a $(J_s - K_s)_J > 2.3$ photometric criterion to select $z > 2$ galaxies (as discussed in section 8.2), then we should ensure that cool Galactic stars are not expected in such a sample. The published data on the lowest-mass stars suggest that they are too blue in infrared colors to be selected this way. Only heavily reddened stars with thick circumstellar dust shells, such as extreme carbon stars or Mira variables, or extremely metal-free stars having a hypothetical $\lesssim 1500$ K blackbody spectrum could also have red $(J_s - K_s)_J > 2.3$ colors but it seems unlikely that the tiny field of the HDF-S would contain such unusual sources.

7 Catalog Parameters

The K_s -selected catalog of sources is published electronically. We describe here a subset of the photometry containing the most important parameters. The catalog with full photometry and explanation can be obtained from the FIRES homepage ⁷.

- *ID*. — A running identification number in catalog order as reported by SExtractor. Sources added in the second detection pass have numbers higher than 10000.
- *x, y*. — The pixel positions of the objects corresponding to the coordinate system of the original (unblocked) WFPC2 version 2 images.
- *RA, DEC*. — The right ascension and declination in equinox J2000.0 coordinates of which only the minutes and seconds of right ascension, and negative arcminutes and arcseconds of declination are given. To these must be added

⁷<http://www.strw.leidenuniv.nl/~fires>

22^h(R.A.) and −60°(DEC).

- $f_{col,i} \pm \sigma_i$. — The sum of counts in the “color” aperture $f_{col,i}$ in band $i = \{U_{300}, B_{450}, V_{606}, I_{814}, J_s, H, K_s\}$ and its simulated uncertainty σ_i , as described in § 5.2. The fluxes are given in units of 10^{-31} ergs s^{−1} Hz^{−1} cm^{−2}.
- $K_{tot} \pm \sigma(K_{tot})$. — Estimate of the total K_s -band flux and its uncertainty. The sum of counts in the “total” aperture is corrected for missing flux assuming a PSF profile outside the aperture, as described in § 5.2.
- ap_{col} . — An integer encoding the aperture type that was used to measure $f_{col,i}$. This is either a (1) 0''7 diameter circular aperture, (2) 2''0 diameter circular aperture, (3) isophotal aperture determined by the detection-image isophote, or a (4) circular aperture with a reduced isophotal diameter $D = \sqrt{(A_{iso}/\pi)}/1.4$.
- ap_{tot} . — An integer encoding the aperture type that was used to measure K_{tot} . This is either a (1) automatic Kron-like aperture, or a (2) circular aperture within a reduced isophotal diameter.
- r_{col}, r_{tot} . — Circularized radii $r = \sqrt{A/\pi}$, corresponding to the area A of the specified “color” or “total” aperture.
- A_{iso}, A_{auto} . — Area of the detection isophote A_{iso} and area of the autoscaling elliptical aperture $A_{auto} = \pi * a * b$ with semi-major axis a and semi-minor axis b .
- $FWHM_K, FWHM_I$. — Full width at half maximum of a source in the K_s detection image $FWHM_K$, and that of the brightest I_{814} -band source that lies in its detection isophote $FWHM_I$. We obtained the latter by running SExtractor separately on the original I_{814} -image and cross correlating the I_{814} -selected catalog with the K_s -limited catalog.
- w_i . — The weight w_i represents, for each band i , the fraction of the total exposure time at the location of a source.
- $flags$. — Three binary flags are given. The *bias* flag indicates either that the AUTO aperture measurement is affected by nearby sources, or marks apertures containing more than 10% bad pixels. The *blended* flag indicates overlapping sources, while the *star* flag shows that the source SED is best fit with a stellar template (see section 6.3).

8 Analysis

8.1 Completeness and Number Counts

The completeness curves for point-sources in the J_s and K_s -band as a function of input magnitude are shown in Figure 8. Our 90% and 50% completeness levels on the AB magnitude system are 25.65 and 26.25, respectively, in K_s , and 26.30 and 26.90 in J_s .

We derived the limits from simulations where we extracted a bright non-saturated star from the survey image and add it back 30000 times at random

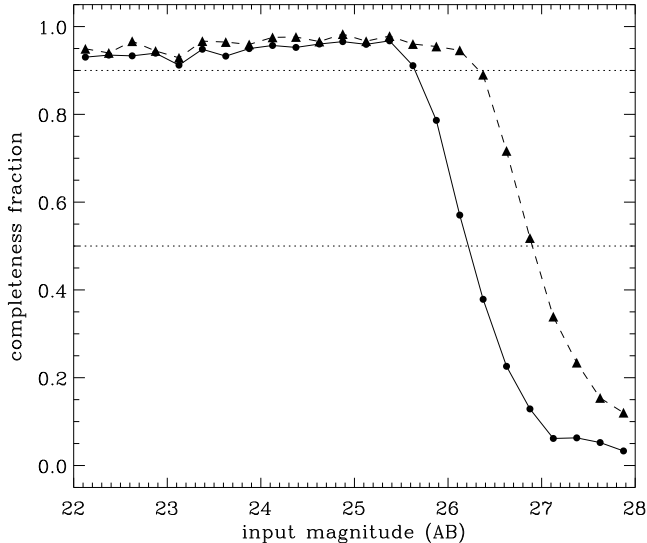


Figure 8 — Completeness curves (on the AB system) for the detectability of point sources in J_s (triangles) and K_s (points), based on simulations where we calculated the recovered fraction of stars that were dimmed to magnitudes between 22 and 28 and embedded in the survey images. The detection threshold of the source extraction software was set to 3.5σ of the filtered background rms. The dotted lines indicate the 50% and 90% completeness levels.

locations, applying a random flux scaling drawn from a rising count slope (or an increasing surface density of galaxies with magnitude) to bring it to magnitudes between $22 \leq K_{s,AB} \leq 28$. We added the dimmed stars back in series of 30 realizations so that they do not overlap each other. The rising count slope needs to be considered because the slope influences the number of recovered galaxies per apparent magnitude, as described below. The input count slope is based on the observed surface densities in the faint magnitude range where the signal-to-noise is $60 \lesssim SNR \lesssim 10$ (or $23 \lesssim K_{s,tot} \lesssim 25$) and where incompleteness does not yet play a role. We used only the deepest central 4.5 arcmin^2 of the J_s and K_s images ($w > 0.95$) with near uniform image quality and exclude four small regions around the brightest stars. In the simulation images we extract sources following the same procedures as described in §5.1, but applying a reduced ($\approx 3.5\sigma$) detection threshold. We measure the recovered fraction of input sources against apparent magnitude, and from this we estimate the detection efficiency of point-like sources which we use to correct the observed number counts. We executed this procedure in the J_s and K_s -band.

The resulting completeness curves assume that the true profile of the source is point-like and therefore they should be considered upper limits. An extended source would have brighter completeness limits depending on the true source size,

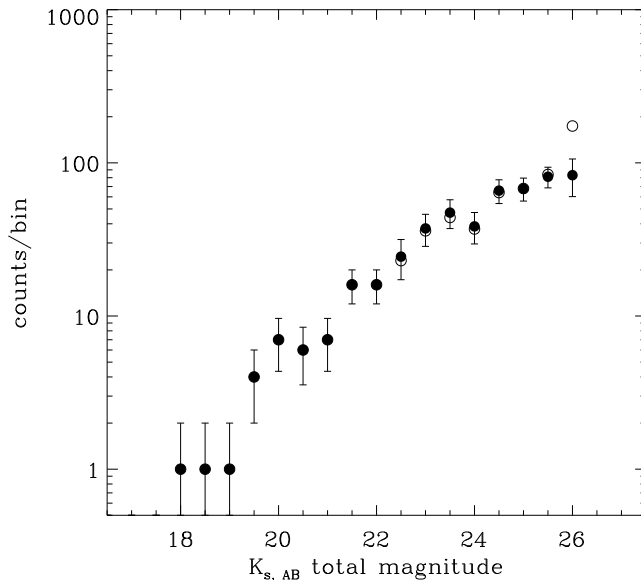


Figure 9 — Differential K_s -band counts (on the AB system) of galaxies in the HDF-S. The counts are based on auto-scaling apertures (Kron 1980) for isolated sources and adapted isophotal apertures for blended sources, both corrected to total magnitudes using stellar growth curve measurements. Raw counts (*open circles*) and counts corrected for incompleteness and false positive detections using point source simulations (*filled circles*) are shown. The small corrections at magnitudes $\gtrsim 23$ reflect missed sources due to confusion. Effective corrections at the faintest magnitudes $K_s \sim 25 - 25.5$ are very small because the loss of sources on negative noise regions (incompleteness) is compensated by the number of sources pushed above the detection limit by positive noise fluctuations. Only the faintest 0.5 mag bin centered on $K_s = 26.0$, bordering the 3.5σ detection limit ($K_s \approx 26.3$), is significantly corrected because of a contribution of false positive detections.

its flux profile, and the filter that is used in the detection process. However, the detailed treatment of detection efficiency as a function of source morphology and detection criteria is beyond the scope of this paper.

When using the completeness simulations to correct the number counts, we choose a simple approach and apply a single correction down to $\approx 50\%$ completeness, based on the ratio of the simulated counts per input magnitude bin to the total recovered counts per observed total magnitude bin. More sophisticated modeling is possible but requires detailed knowledge of the intrinsic size and shape distribution of faint NIR galaxies. The simple approach corrects for all effects resulting from detection criteria, photometric scheme, incompleteness, and noise peaks. We find it works well if the total magnitude of sources is measured correctly, with little systematic difference between the input and recovered magnitudes, which is the case for our photometric scheme (see section 5.2). It is

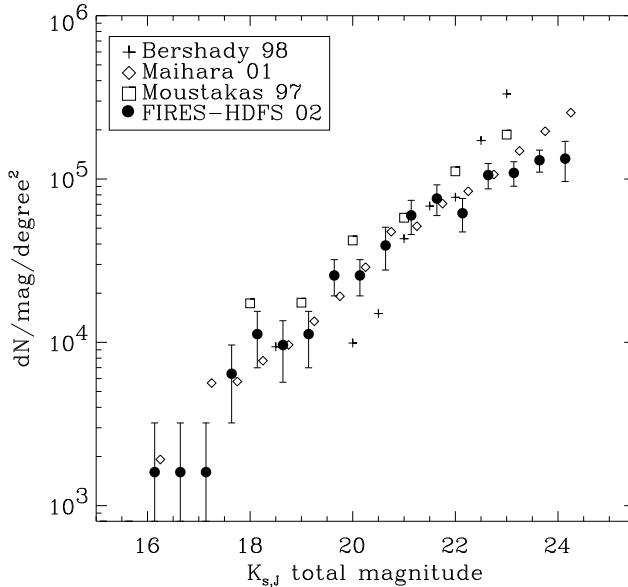


Figure 10 — FIRES K_s -band galaxy counts (on the conventional Johnson system) compared to published counts in deep K -band fields. The corrected counts (*filled circles*) are shown for FIRES data. The Maihara et al. (2001) counts have been plotted to their $S/N \sim 3$ limit. The slope at magnitudes $K_J > 21$ is flatter than reported in other surveys although straightforward comparisons are difficult, due to model-dependent correction factors of $\sim 2 - 3$ applied to the faintest data points in these surveys. The nature of the scatter in count slopes is unclear but field-to-field variations as well as different photometry and corrections procedures likely play a role. The FIRES counts need little correction for completeness effects or false positive detections, except for the $K_{s,J} = 24.25$ bin.

worth noticing that at $K_{s,AB} > 25$ we actually recovered slightly more counts in the *observed* magnitude bins than we put in. This is caused by the fact that, in the case of a rising count slope, there are more faint galaxies boosted by positive noise peaks than bright galaxies lost on negative noise peaks. This effect is strong at low signal-to-noise fluxes and results in a slight excess of recovered counts. This is the main reason that we required little correction up to the detection threshold, except for the faintest 0.5 mag bin centered on $K_s = 26.0$, which contained false positive detections due to noise. After removing stars (see section 6.3), we plot in Figure 9 the raw and corrected source counts against total magnitude.

Figure 10 presents a compilation of other deep K -band number counts from a number of published studies. The FIRES counts follow a $d\log(N)/dm$ relation with a logarithmic slope $\alpha \approx 0.25$ at $20 \lesssim K_{s,J} \lesssim 22$ (Johnson magnitudes) and decline at fainter magnitudes to $\alpha \approx 0.15$ at $22.0 \lesssim K_{s,J} \lesssim 24$. This flattening of the slope has not been seen in other deep NIR surveys, where we emphasize

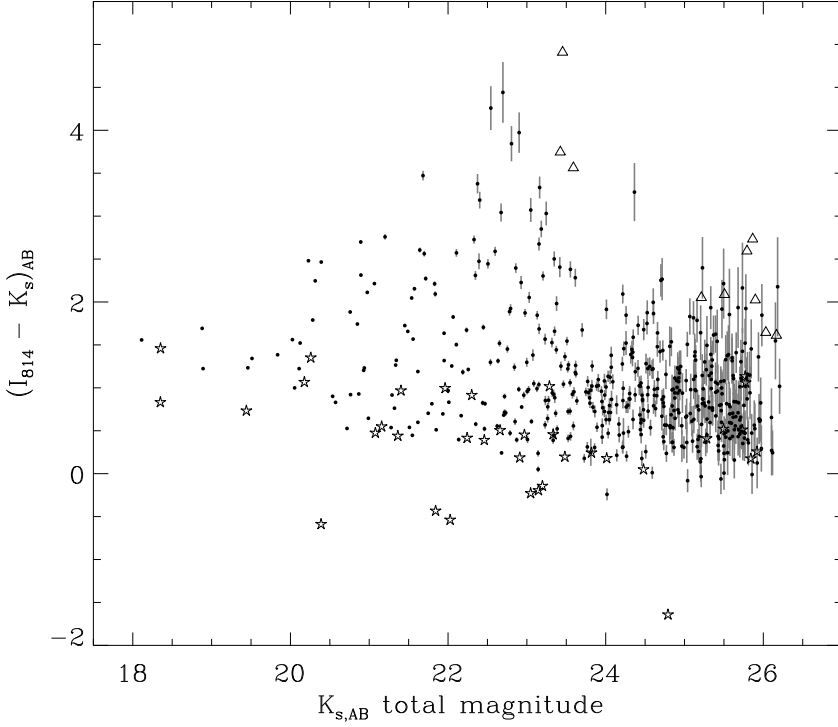


Figure 11 — $I_{814} - K_s$ versus K_s color-magnitude relation (on the AB system) for K_s -selected objects in the HDF-S. Only sources with a minimum of 20% of the total exposure time in all bands are included and identified stars are marked by a star symbol. Colors are plotted with $\pm 1\sigma$ error bars, and I_{814} measurements with $S/N < 2$ (*triangles*) are plotted at their 2σ confidence interval, indicating lower limits for the colors. There are more red sources with $I_{814} - K_s > 2.6$ at $K \sim 23$ than at $K \sim 24$ where the I_{814} is still sufficiently deep to select them. The transformation of the $I_{814} - K_s$ color from the AB system to the Johnson magnitude system is $(I_{814} - K_s)_J = (I_{814} - K_s)_{AB} + 1.43$.

that the FIRES HDFS field is the largest and the deepest amongst these surveys, and that only the counts in the last FIRES bin at $K_{s,J} = 24.25$ were substantially corrected. It is remarkable that the SUBARU Deep Field count slope $\alpha \approx 0.23$ of Maihara et al. (2001) looks smooth compared to the HDF-S although their survey area and the raw count statistics are slightly smaller.

Other authors (Djorgovski et al. 1995; Moustakas et al. 1997; Bershadsky, Lowenthal, & Koo 1998) find logarithmic counts slopes in K ranging from 0.23 to 0.36 over $20 \lesssim K_J \lesssim 23 - 24$, however the counts in the faintest bins in these surveys were boosted by factors of $\sim 2 - 3$, based on completeness simulations. The origin

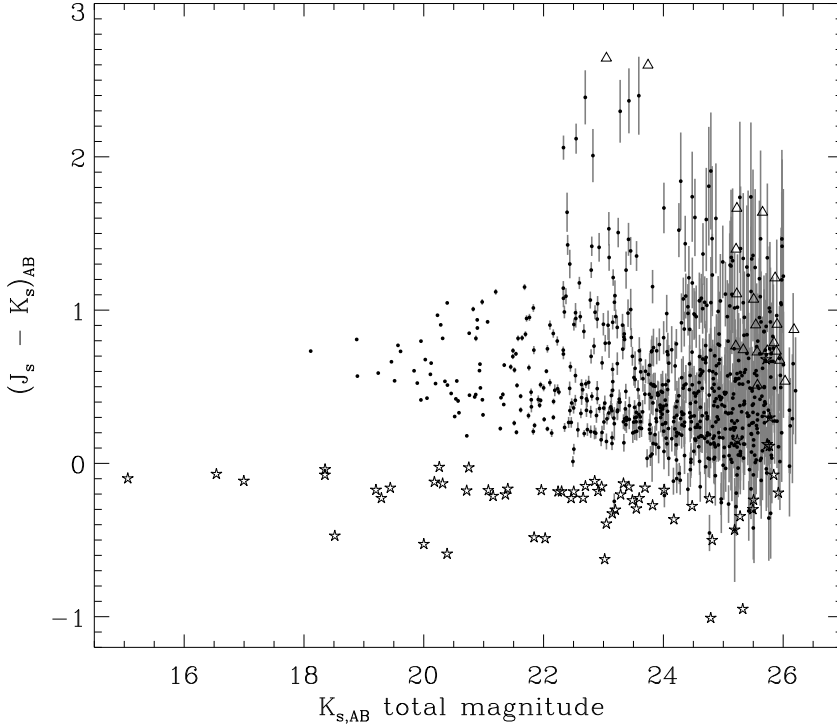


Figure 12 — Same as Figure 11 for the $J_s - K_s$ color. Striking is the number the galaxies with very red NIR colors $J_s - K_s \gtrsim 1.34$ (on the AB system) or $J_s - K_s \gtrsim 2.3$ (Johnson). These systems have photometric redshifts $z > 2$ and are extremely faint in the observer’s optical; as such they would not be selected with the U-dropout technique. Identified stars are well separated from redshifted galaxies and almost all have $J_s - K_s \leq 0$ colors. The transformation of the $J_s - K_s$ color from the AB system to the Johnson magnitude system is $(J_s - K_s)_J = (J_s - K_s)_{AB} + 0.96$.

of the faint-end discrepancies of the K counts is unclear. Cosmic variance can play a role, because the survey areas never exceed a few armin^2 , but also differences in the used filters (K_s, K_I, K) and differences in the techniques and assumptions used to estimate the total magnitude (see 5) or to correct the counts for incompleteness may be important. Further analysis is needed to ascertain whether size-dependent biases in the completeness correction play a role in the faint-end count slope.

8.2 Color-Magnitude Distributions

Figures 11 – 14 show color-magnitude diagrams of K_s -selected galaxies in the HDF-S. The $I_{814} - K_s$ versus K_s color-magnitude diagram in Figure 11 shows a

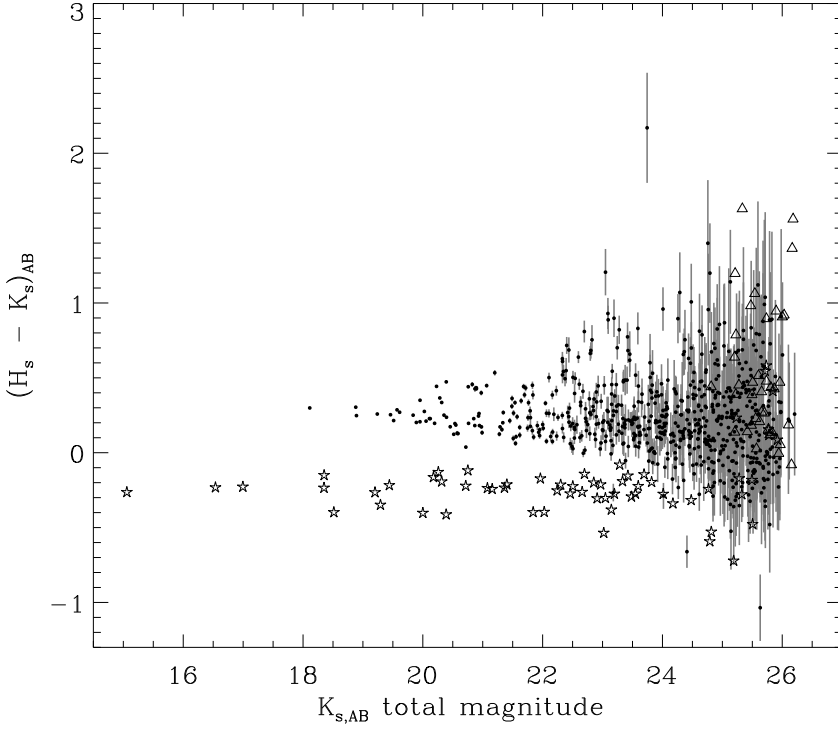


Figure 13 — Same as Figure 11 for the $H - K_s$ color. One of the galaxies is extremely red with $H - K_s \approx 2.2$ and is barely visible in J_s and H . The transformation of the $H - K_s$ color from the AB system to the Johnson magnitude system is $(H - K_s)_J = (H - K_s)_{AB} + 0.48$.

large number of extremely red objects (EROs) with $I_{814} - K_s \gtrsim 2.6$ (on the AB system) or $(I_{814} - K_s)_J \gtrsim 4$ (Johnson). There appears to be an excess of EROs at total magnitudes $K_{s,AB} \sim 23$ compared to magnitudes $K_{s,AB} \sim 24$. This is not caused by the insufficient signal-to-noise ratio in the I_{814} measurements. In a similar diagram for the $J_s - K_s$ color shown in Figure 12, there is a striking presence at the same K_s magnitudes of sources with very red $(J_s - K_s)_{AB} \gtrsim 1.34$ or $(J_s - K_s)_J \gtrsim 2.3$ colors. Such sources were also found by Saracco et al. (2001), using shallow NIR data, who suggested they might be dusty starbursts or ellipticals at $z > 2$. Interestingly, any evolved galaxy with a prominent Balmer/4000Å discontinuity in their spectrum, like most present-day Hubble Type galaxies, would have such very red observed NIR colors if placed at redshifts $z > 2$. While the $(J_s - K_s)_J \gtrsim 2.3$ sources we find are generally morphologically compact, with exceptions, we do not expect the sources with good photometry to be faint cool L-dwarf stars because known colors of such stars are $(J_s - K_s)_J \lesssim 2.1$ (see section

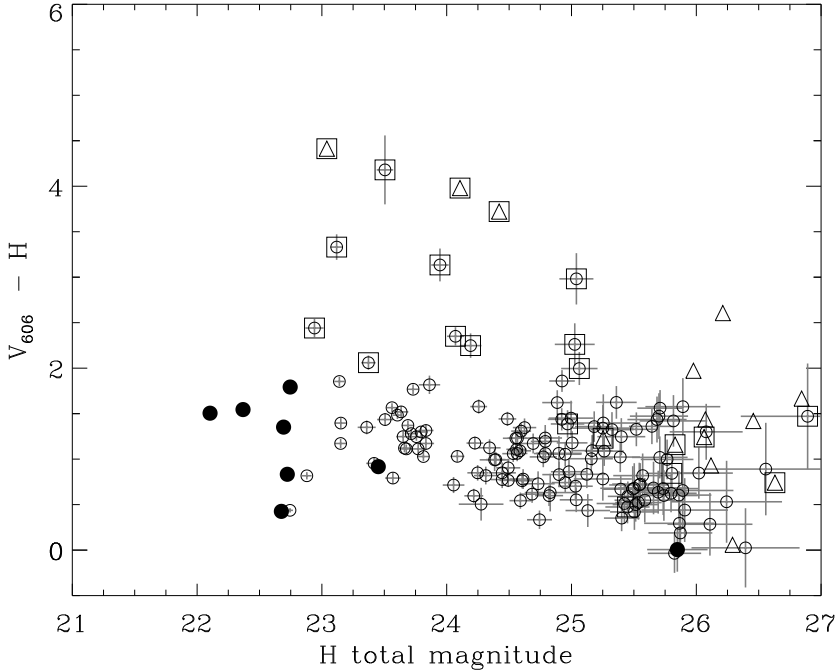


Figure 14 — $V_{606} - H$ versus H color-magnitude diagram (on the AB system) for galaxies in the HDF-S K_s -selected catalog with $1.95 < z_{phot} < 3.5$. Filled symbols indicate galaxies with spectroscopy. The number of candidates for red, evolved galaxies is much higher than in the HDF-N for a similar survey area, as shown in a identical plot in Fig. 1 of Papovich, Dickinson, & Ferguson (2001): we find 7 galaxies redder than $V_{606,AB} - H_{AB} \gtrsim 3$ and brighter than $H_{AB} \lesssim 25.5$, compared to only one in the HDF-N. Galaxies with $S/N < 2$ for the V_{606} measurement (*triangles*) are plotted at the 2σ confidence limit in V_{606} , indicating a lower limit on the $V_{606} - H$ color. The subsample of galaxies having red $(J_s - K_s)_J > 2.3$ colors (*open squares*) is also shown. The transformation of the $V_{606} - H_s$ color from the AB system to the Johnson magnitude system is $(V_{606} - H)_J = (V_{606} - H)_{AB} + 1.26$.

6.3). The photometric redshifts of all red NIR galaxies are $z_{phot} \gtrsim 2$, but they would be missed by ultraviolet-optical color selection techniques such as the U-dropout method, because most of them are barely detectable even in the deepest optical images. One bright NIR galaxy is completely undetected in the original WFPC2 images. The $(J_s - K_s)_J \gtrsim 2.3$ sources are studied in more detail by Franx et al. (2002) and the relative contributions of these galaxies and U-dropouts to the rest-frame optical luminosity density will be presented in Rudnick et al. (2002b).

If we select sources with $1.95 < z_{phot} < 3.5$, we find clear differences in the $V_{606} - H$ versus H color-magnitude diagram between our NIR-selected galaxies in

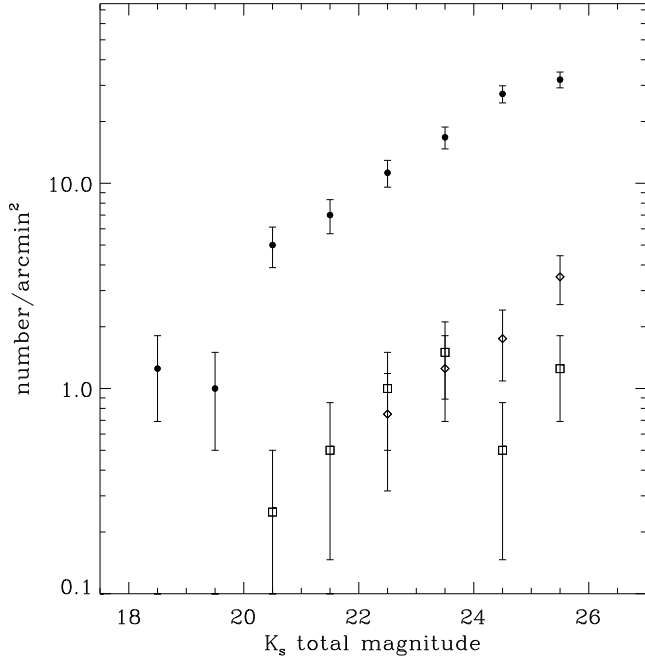


Figure 15 — The surface densities of galaxies selected by color in the 5σ catalog of sources of HDF-S. Presented are galaxies with $(J_s - K_s)_J > 2.3$ (in Johnson magnitudes) (*diamonds*), extremely red objects with $(I_{814} - K_s)_J > 4$ (*squares*), and all K_s -selected galaxies (*filled; points*) as a function total K_s -band AB magnitude. Only sources with a minimum exposure time of 40%, 40% and 90% of the total in I_{814} , J_s and K_s are plotted, so that the selection in K_s is uniform over the area, and the I_{814} and J_s observations are sufficiently deep to prevent a bias against objects with very red $I_{814} - K_s$ and $J_s - K_s$ colors. No corrections to the counts have been applied. The errorbars are poissonian and might underestimate the true uncertainty which would also contain contributions from large scale structure.

the HDF-S and those of the HDF-N (compare Figure 14 to Figure 1 of Papovich, Dickinson, & Ferguson 2001). Over a similar survey area and to similar limiting depths, we find 7 galaxies redder than $(V_{606} - H)_{AB} \gtrsim 3$ and brighter than total magnitude $H_{AB} \lesssim 25.5$, compared to only one in the HDF-N. While the surface density of such galaxies is not well known, it is clear that the HDF-N contains far fewer of them than the HDF-S. It remains to be seen if this is just field-to-field variation, or that one of the two fields is atypical. The results of the second much larger FIRES field centered on MS1054-03 (Förster Schreiber et al. 2002) should provide more insight into this issue. We note that all 7 $(V_{606} - H)_{AB} \gtrsim 3$ galaxies in the HDF-S are also amongst the brightest 16 $(J_s - K_s)_J > 2.3$ sources. Figure 15 shows the surface densities of EROs and galaxies with $(J_s - K_s)_J > 2.3$ colors as function of K_s -band total magnitude. The surface density of EROs peaks around $K_{s,AB} \approx 23$ and then drops or flattens at fainter magnitudes, contrary to

the number of $(J_s - K_s)_J > 2.3$ galaxies which keeps rising to the faintest K_s total magnitudes in our catalog.

9 Summary and Conclusions

We have presented the results of the FIRES deep NIR imaging of the WFPC2-field of the HDF-S obtained with ISAAC at the VLT: the deepest ground-based NIR data available, and the deepest K_s -band of any field. We constructed a K_s -selected multicolor catalog of galaxies, consisting of 833 objects with $K_{s,AB} \lesssim 26$ and photometry in seven-bands from 0.3 to 2.2μ for 624 of them. These data are available electronically together with photometric redshifts for 567 galaxies. Our unique combination of deep optical space-based data from the HST together with deep ground-based NIR data from the VLT allows us to sample light redder than the rest-frame V-band in galaxies with $z \lesssim 3$ and to select galaxies from their rest-frame optical properties, obtaining a more complete census of the stellar mass in the high-redshift universe. We summarize our main findings below:

- The K_s -band galaxy counts in HDF-S turn over at the faintest magnitudes and flatten from $\alpha \approx 0.25$ at total AB magnitudes $22 \lesssim K_s \lesssim 24$ to $\alpha \approx 0.15$ at $24 \lesssim K_s \lesssim 26$; this is flatter than counts in previously published deep NIR surveys, where the FIRES HDF-S field is largest and deepest amongst these surveys. The nature of the scatter in the faint-end counts is yet unclear but field-to-field variations as well as different analysis techniques likely play a role.
- The HDF-S contains 7 sources redder than $(V_{606} - H)_{AB} \gtrsim 3$ and brighter than total magnitude $H_{AB} \lesssim 25.5$ at photometric redshifts $1.95 < z_{phot} < 3.5$, while such galaxies were virtually absent in the HDF-N. They are much redder than regular U-dropout galaxies in the same field and are candidates for relatively massive, evolved systems at high redshift. The difference with the HDF-N might just reflect field-to-field variance, calling for more observations to similar limits with full optical-to-infrared coverage. Results from the second and larger FIRES field centered on MS1054-03 (Förster Schreiber et al. 2002) should provide more insight into this issue.
- We find substantial numbers of red galaxies with $(J_s - K_s)_J > 2.3$ that have photometric redshifts $z_{phot} > 2$. These galaxies would be missed by ultraviolet-optical color selection techniques such as the U-dropout method because most of them are barely detectable even in the deepest optical images. The surface densities of these sources in our field keeps rising down to the detection limit in K_s , in contrast to the number counts of EROs which peak at $K_{s,AB} \sim 23$ and then drop or flatten at fainter magnitudes.

The results of the HDF-S presented in this paper demonstrate the necessity of extending optical observations to near-IR wavelengths for a more complete census of the early universe. Our deep K_s -band data prove invaluable for they probe well into the rest-frame optical at $2 < z < 4$, where long-lived stars may dominate the

light of galaxies. We are pursuing follow-up programs to obtain more spectroscopic redshifts needed to confirm the above results. Updates on the FIRES programme and access to the reduced images and catalogues can be found at our website <http://www.strw.leidenuniv.nl/~fires>.

Acknowledgments

The data here presented have been obtained as part of an ESO Service Mode program. We would very much like to thank the ESO staff for their kind assistance and enormous efforts in taking these data and making them available to us. This project would not be possible without their dedicated work. This work was supported by a grant from the Netherlands Organization for Scientific Research. We would like to thank the Lorentz Center of Leiden University for its hospitality during several workshops. The comments of the referee helped to improve the paper.

References

- Adelberger, K. L. & Steidel, C. C. 2000, *ApJ*, 544, 218
P. Amico et al. eds, 2001, *ISAAC-SW Data Reduction Guide* (version 1.5)
Barger, A. J., Cowie, L. L., Bautz, M. W., Brandt, W. N., Garmire, G. P., Hornschemeier, A. E., Ivison, R. J., & Owen, F. N. 2001, *AJ*, 122, 2177
Bershady, M. A., Lowenthal, J. D., & Koo, D. C. 1998, *ApJ*, 505, 50
Bertin, E. & Arnouts, S. 1996, *A&AS*, 117, 393
Bessell, M. S. & Brett, J. M. 1988, *PASP*, 100, 1134 (JHKLMN photometric system)
Bruzual A., G., & Charlot, S. 2001, in preparation
Casertano, S. et al. 2000, *AJ*, 120, 2747
Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, *ApJS*, 43, 393 (CWW80)
Cowie, L. L. et al. 2001, *ApJ*, 551, L9
da Costa, L. et al. 1998 (astro-ph/9812105)
Dickinson, M., 2001a, in preparation
Djorgovski, S. et al. 1995, *ApJ*, 438, L13
Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, *ApJ*, 513, 34 (FLY99)
Forster Schreiber, N.M. et al. 2002, in preparation
Franx, M., Illingworth, G. D., Heckman, T. M., 1989, *A.J.*, 98 538-576
Franx, M. et al. 2000, *The Messenger*, 99, 20
Franx, M. et al. 2002, in preparation
Gialalisco, M. & Dickinson, M. 2001, *ApJ*, 550, 177
Hauschildt, P. H., Allard, F., Ferguson, J., Baron, E., & Alexander, D. R. 1999, *ApJ*, 525, 871
Johnson, H. L. 1966, *ARA&A*, 4, 193
Kinney, A. L., Calzetti, D., Bohlin, R. C., McQuade, K., Storchi-Bergmann, T., & Schmitt, H. R. 1996, *ApJ*, 467, 38
Kirkpatrick, J. D. et al. 2000, *AJ*, 120, 447
Kron, R. G. 1980, *ApJS*, 43, 305
Madau, P. 1995, *ApJ*, 441, 18
Maihara, T. et al. 2001, *PASJ*, 53, 25

- Moorwood, A. F. 1997, *Proc. SPIE*, 2871, 1146
- Moustakas, L. A., Davis, M., Graham, J. R., Silk, J., Peterson, B. A., & Yoshii, Y. 1997, *ApJ*, 475, 445
- Oke, J. B. 1971, *ApJ*, 170, 193
- Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, *ApJ*, 559, 620 (P01)
- Persson et al. 1998 *AJ*, 116, 2475
- Pettini, M., Kellogg, M., Steidel, C. C., Dickinson, M., Adelberger, K. L., & Giavalisco, M. 1998, *ApJ*, 508, 539
- Pettini, M., Shapley, A. E., Steidel, C. C., Cuby, J., Dickinson, M., Moorwood, A. F. M., Adelberger, K. L., & Giavalisco, M. 2001, *ApJ*, 554, 981
- Pozzetti, L., Bruzual A., G., & Zamorani, G. 1996, *MNRAS*, 281, 953
- Rudnick, G. et al. 2001, *AJ*, 122, 2205 (R01)
- Rudnick, G. et al. 2002, in preparation (R02a)
- Rudnick, G. et al. 2002, in preparation (R02b)
- Saracco, P., Giallongo, E., Cristiani, S., D’Odorico, S., Fontana, A., Iovino, A., Poli, F., & Vanzella, E. 2001, *A&A*, 375, 1
- Shapley, A. E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., & Pettini, M. 2001, *ApJ*, 562, 95
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, *AJ*, 112, 352
- Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, *ApJ*, 462, L17
- Smail, I., Ivison, R. J., Owen, F. N., Blain, A. W., & Kneib, J.-P. 2000, *ApJ*, 528, 612
- Voit, M., ed. 1997, *HST Data Handbook*, Vol. 1 (version 3.1; Baltimore: STScI)
- Williams, R. E. et al. 1996, *AJ*, 112, 1335
- Williams, R. E. et al. 2000, *AJ*, 120, 2735

Table 1 — Summary of the HDF-S Observations

Camera	Filter	Number of Exposures	Integration Time (h)	FWHM ¹ (arcsec)
WFPC2	<i>F300W</i>	102	36.8	0''16
WFPC2	<i>F450W</i>	51	28.3	0''14
WFPC2	<i>F606W</i>	49	27.0	0''13
WFPC2	<i>F814W</i>	56	31.2	0''14
ISAAC	<i>J_s</i>	1007	33.6	0''45
ISAAC	<i>H</i>	968	32.3	0''48
ISAAC	<i>K_s</i>	2136	35.6	0''46

¹ The full width at half-maximum of the best-fitting Gaussian.

Table 2 — Zero Points for the HDF-S

Data Set	Zero Point	
	Johnson mag	AB mag
<i>U</i> ₃₀₀	19.43	20.77
<i>B</i> ₄₅₀	22.01	21.93
<i>V</i> ₆₀₆	22.90	23.02
<i>I</i> ₈₁₄	21.66	22.09
<i>J_s</i>	24.70	25.60
<i>H</i>	24.60	25.98
<i>K_s</i>	24.12	25.98

Table 3 — Background noise in the HDF-S images

Data Set ¹	rms background ²	a ³	b ³	1 σ sky noise limit ⁴
U_{300}	1.34e-05	2.51	0.38	29.5
B_{450}	1.88e-05	2.65	0.40	30.3
V_{606}	3.80e-05	2.49	0.39	30.6
I_{814}	2.79e-05	2.59	0.39	30.0
J_s	0.0069	1.46	0.047	28.6
H	0.0165	1.43	0.044	28.1
K_s	0.0163	1.49	0.038	28.1

¹ The images are all at the $0''.119$ pixel⁻¹ scale. The WFPC2 are 3x3 block summed, and all images are smoothed to match the image quality of the H -band.

² Pixel-to-pixel rms variations (in instrumental counts per second) as measured directly in empty parts of the registered convolved images which were used for photometry.

³ Best-fit parameters of Eq. 3 which gives the effective rms variation of the background as a function of linear size of the aperture.

⁴ The 1σ sky noise limit in a $0''.7$ circular diameter aperture (≈ 0.4 arcsec²) in AB magnitudes using Eq. 3.

Large Disk-Like Galaxies at High Redshift

ABSTRACT

Using deep near-infrared imaging of the Hubble Deep Field South with ISAAC on the *Very Large Telescope* we find 6 large disk-like galaxies at redshifts $z = 1.4 - 3.0$. The galaxies, selected in K_s ($2.2\mu m$), are regular and surprisingly large in the near-infrared (rest-frame optical), with face-on effective radii $r_e = 0''.65 - 0''.9$ or $5.0 - 7.5 h_{70}^{-1}$ kpc in a Λ CDM cosmology, comparable to the Milky Way. The surface brightness profiles are consistent with an exponential law over $2 - 3$ effective radii. The WFPC2 morphologies in *Hubble Space Telescope* imaging (rest-frame UV) are irregular and show complex aggregates of star-forming regions $\sim 2''$ ($\sim 15 h_{70}^{-1}$ kpc) across, symmetrically distributed around the K_s -band centers. The spectral energy distributions show clear breaks in the rest-frame optical. The breaks are strongest in the central regions of the galaxies, and can be identified as the age-sensitive Balmer/4000 Å break. The most straightforward interpretation is that these galaxies are large disk galaxies; deep NIR data are indispensable for this classification. The candidate disks constitute 50% of galaxies with $L_V \gtrsim 6 \times 10^{10} h_{70}^{-2} L_\odot$ at $z = 1.4 - 3.0$. This discovery was not expected on the basis of previously studied samples. In particular, the Hubble Deep Field North is deficient in large galaxies with the morphologies and profiles we report here.

Ivo Labbé, Gregory Rudnick, Marijn Franx, Emanuele Daddi, Pieter G. van Dokkum, Natascha M. Förster Schreiber, Konrad Kuijken, Alan Moorwood, Hans-Walter Rix, Huub Röttgering, Ignacio Trujillo, Arjen van de Wel, Paul van der Werf, & Lottje van Starkenburg,
Astrophysical Journal Letters, **591**, L95

1 Introduction

DISK galaxies are believed to undergo a relatively simple formation process in which gas cools and contracts in dark matter halos to form rotationally supported disks with exponential light profiles (Fall & Efstathiou 1980; Mo, Mao, & White 1998). A critical test of any theory of galaxy formation is to reproduce the observed properties and evolution of galaxy disks.

Previous optical spectroscopy and HST imaging have yielded a wealth of data on disk galaxies at $z \lesssim 1$. (e.g., Vogt et al. 1996, 1997; Lilly et al. 1998; Barden et al. 2003), although contradictory claims have been made regarding the implications for the size and luminosity evolution with redshift (see Lilly et al. 1998; Mao, Mo, & White 1998; Barden et al. 2003), and the importance of surface brightness selection effects (see Simard et al. 1999; Bouwens & Silk 2002).

It is still unknown what the space density and properties are of disk galaxies at substantially higher redshift. Many galaxies at $z \sim 3$ have been identified using the efficient U-dropout technique (Steidel et al. 1996a,b). Most of these objects are compact with radii $\sim 1 - 2 h_{70}^{-1}$ kpc, while some are large and irregular (Giavalisco, Steidel, & Macchetto 1996; Lowenthal et al. 1997). However, the U-drop selection requires high far-UV surface brightness due to active, spatially compact, and unobscured star formation. As a result, large and UV-faint disk galaxies may have been overlooked and, additionally, the morphologies of LBGs could just reveal the unobscured star-forming regions rather than the more evolved underlying population which forms the disk.

The most direct evidence for the existence of large disks at high redshift has come from observations in the NIR, which provide access to the rest-frame optical. Here the continuum light is more indicative of the distribution of stellar mass than in the UV and nebular lines are accessible for kinematic measurements. van Dokkum & Stanford (2001) discuss a K -selected galaxy at $z = 1.34$ with a rotation velocity of $\sim 290 \text{ km s}^{-1}$. Erb et al. (2003) detect $\sim 150 \text{ km s}^{-1}$ rotation at $\sim 6 h_{70}^{-1}$ kpc radii in the $H\alpha$ emission line of galaxies at $z \sim 2.3$ and Moorwood et al. (2003) find $\gtrsim 100 \text{ km s}^{-1}$ rotation at $\sim 6 h_{70}^{-1}$ kpc from the center of a galaxy at $z = 3.2$, seen in the NIR spectrum of the $[\text{O III}]\lambda 5007 \text{ \AA}$ emission line.

The imaging data in these studies, however, are of limited depth and resolution, making it difficult to determine morphological properties. In this Letter, we present an analysis of the rest-frame ultraviolet-to-optical morphologies and spectral energy distributions (SEDs) of 6 large candidate disk galaxies at $z \sim 1.4 - 3$ using the deepest groundbased NIR dataset currently available (Labbé et al. 2003). Throughout, we adopt a flat Λ -dominated cosmology ($\Omega_M = 0.3, \Lambda = 0.7, H_0 = 70 h_{70} \text{ km s}^{-1} \text{ Mpc}^{-1}$). All magnitudes are expressed in the Johnson photometric system.

2 Observations

We obtained 102 hours of NIR J_s , H , and K_s imaging in the HDF-S ($2'5 \times 2'5$) under excellent seeing ($\text{FWHM} \approx 0''.46$), using ISAAC (Moorwood 1997) on the VLT. The observations were taken as part of the Faint InfraRed Extragalactic Survey (FIRES; Franx et al. 2000). We combined our data with existing deep optical HST/WFPC2 imaging (version 2; Casertano et al. 2000), in the U_{300} , B_{450} , V_{606} and I_{814} bands, and we assembled a K_s -selected catalog of sources with SExtractor (Bertin & Arnouts 1996). Photometric redshifts and rest-frame luminosities were derived by fitting a linear combination of empirical galaxy spectra and stellar population models to the observed flux points (Rudnick et al. 2001, 2003a). The reduced images, photometric catalog, and redshifts are presented in Labbé et al. (2003) and are all available on-line at the FIRES homepage¹. Furthermore, we obtained optical spectroscopy with FORS1 on the VLT for some of the sources (Rudnick et al. 2003b). Additional redshifts were obtained from Vanzella et al. (2002). As discussed in Rudnick et al. (2001, 2003a), our photometric redshifts yield good agreement with the spectroscopic redshifts, with rms $|z_{\text{spec}} - z_{\text{phot}}|/(1 + z_{\text{spec}}) \approx 0.05$ for $z_{\text{spec}} > 1.4$.

Large disk galaxies in the HDF-South were identified by fitting exponential profiles convolved with the Point Spread Function (PSF) to the K_s -band images. Six objects at $z > 1.4$ have effective radii $r_e > 3.6 h_{70}^{-1}$ kpc, three of which have spectroscopic redshifts. The mean redshift of the sample is 2.4. We will focus on these large galaxies in the remainder of the Letter. The structural properties of the full K_s -selected sample will be discussed in Trujillo et al. (2003).

3 Rest-frame Optical versus UV Morphology

The large galaxies are shown in Figure 1. They have a regular morphology in the ISAAC K_s -band ($2.2\mu\text{m}$), which probes rest-frame optical wavelengths between 5400 and 9000 Å. In contrast, the WFPC2 V_{606} and I_{814} -band morphologies, which map the unobscured star-forming regions at rest-frame UV wavelengths between 1500 and 3300 Å, are irregular with several knots up to $\sim 2''$ ($\sim 15 h_{70}^{-1}$ kpc) apart, symmetrically distributed around the K_s -band centers. In a few cases the observed optical light is spatially almost distinct from the NIR.

As a result of the structure in the WFPC2 imaging, 4 of the large objects have been split up into two sources by Casertano et al (2003). Fig. 1 shows the corresponding “segmentation” map by “SExtractor” which illustrates how the pixels in each image are allocated to different sources. The galaxies were not split up when the K_s band image was used to detect objects. However, the broader PSF in the K_s -band image can play a role: if we smooth the I_{814} -band data to the same resolution, we find that SExtractor only splits up 1 galaxy.

¹<http://www.strw.leidenuniv.nl/~fires>

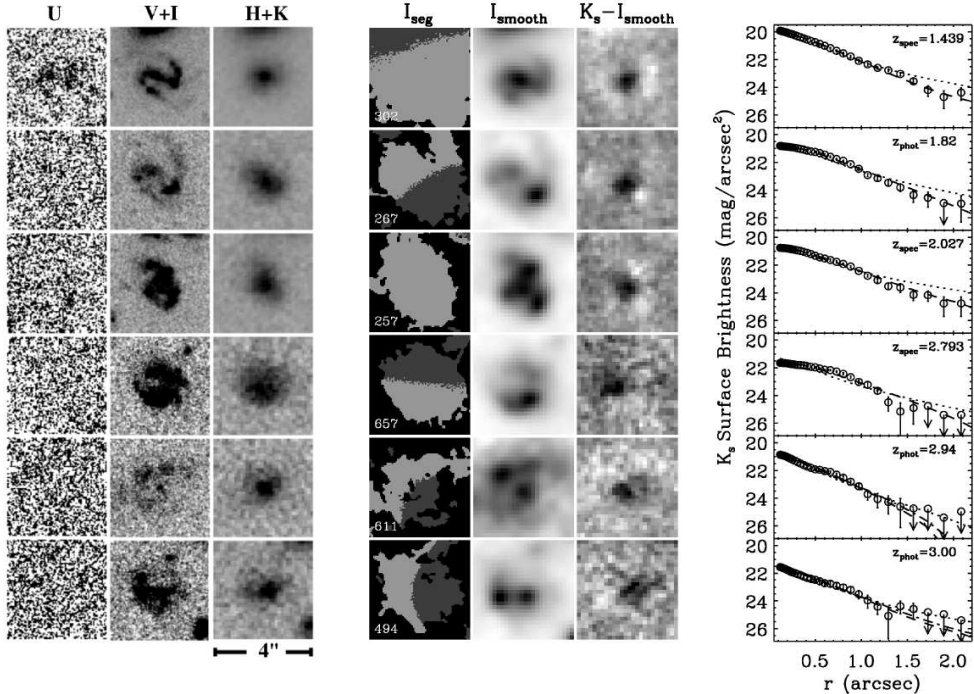


Figure 1 — *Left panels:* The WFPC2 U_{300} , averaged $V_{606} + I_{814}$ (rest-frame UV), and our averaged ISAAC $H + K_s$ images (rest-frame optical), scaled proportional to F_λ with arbitrary normalization per galaxy. The rest-frame UV morphologies are complex and symmetric with respect to the center of the smoother optical distribution. *Middle panels:* SExtractor’s I_{814} -band segmentation map, the smoothed I_{814} images, and the K_s -band images after subtracting the scaled, smoothed I_{814} -band. From the segmentation map it follows that detection in I_{814} would likely split up most of the sources. The central residuals in $K_s - I_{\text{smooth}}$ demonstrate that optical and NIR light are distributed differently and that all galaxies have a “red” nucleus. *Right panels:* The radial profiles in the K_s -band. The abscissa and the ordinate are respectively the mean geometric radial distance and the surface brightness along elliptical isophotes. The arrows mark 1σ confidence intervals for measurements with signal-to-noise less than 1. Overplotted are the best-fit exponential law (*dashed*), $r^{1/4}$ law (*dotted*), and point+exponential (*dash-dot*) for galaxy 494 and 611.

Hence, the question remains whether these 4 objects split up in I_{814} are superpositions or whether they are part of larger systems. We tested this by subtracting the PSF-matched I_{814} -band images from the K_s -band images. The I_{814} -band images were scaled to the K_s -band images to minimize the residuals. The residuals are shown in Fig. 1. In all cases, we find strong positive residuals close to the centers of the objects as defined in the K_s -band, whereas residuals at any of the

Table 1 — Properties of high redshift disk galaxies in the HDF-S

Galaxy ^a	$K_{s,tot}^b$	z	$M_{B,rest}^c$	$\mu_{0,B,rest}^d$	$r_{e,K}^e$	$r_{1/2,K}^f$	$r_{1/2,I}^g$	ϵ^h
302	19.70	1.439 ⁱ	-22.70	19.70	0.89	0.70	0.86	0.46
267	19.98	1.82	-22.88	19.92	0.75	0.74	0.88	0.37
257	20.25	2.027 ⁱ	-23.08	19.53	0.74	0.74	0.84	0.36
657	20.68	2.793 ⁱ	-23.56	19.33	0.76	0.70	0.74	0.18
611	20.53	2.94	-23.59	18.51	0.65 ^j	0.52	0.97	0.27
494	21.14	3.00	-23.31	18.84	0.75 ^j	0.56	0.86	0.47

^a Catalog identification (Labbé et al. 2003)^b K_s -band total magnitudes^c Rest-frame absolute B -band magnitudes^d Face-on rest-frame B -band surface brightnesses^e Face-on best-fit effective radii (arcsec)^f K_s half-light radii (arcsec)^g I_{814} half-light radii, matched to K_s ^h Ellipticityⁱ Spectroscopic redshifts^j Two-component models
(point + exponential)

I_{814} -band peaks might be expected in case of a chance superposition. Furthermore, we performed photometric redshift analyses for subsections of the images and found no evidence for components at different redshifts.

4 Profile fits and sizes

Next, we fitted simple models convolved with the PSF (FWHM $\approx 0''.46$) to the two-dimensional surface brightness distributions in the K_s -band. The images are well-described by a simple exponential law over 2 – 3 effective radii (galaxy 302, 267, 257 and 657) or by a point source plus exponential (galaxy 611 and 494), where the point source presumably represents the light emitted by a compact bulge contributing about 40% of the light. We also derived intensity profiles by ellipse fitting. As can be seen in Figure 1, most galaxies are well described by an exponential. The central surface brightnesses and effective radii, enclosing half of the flux of the model profile, are corrected to face-on and shown in Table 1. The central surface brightness is multiplied with $\sqrt{1 - \epsilon}$, as an intermediate case between optically thin and optically thick, and corrected for cosmological dimming.

The effective radii (semi-major axes) are surprisingly large, $r_e = 0''.65 - 0''.9$ ($5.0 - 7.5 h_{70}^{-1}$ kpc in a Λ CDM cosmology), comparable to the Milky Way and much larger than typical sizes of “normal” Ly-break galaxies (Giavalisco, Steidel, & Macchetto 1996; Lowenthal et al. 1997). As might be expected from the previous section, the I-band images have even larger effective radii. All galaxies have a “red” nucleus, and the colors become bluer in the outer parts.

Overall, the optical-to-infrared morphologies and sizes are strikingly similar to L^* disk galaxies in the local universe, with red bulges, more diffuse bluer ex-

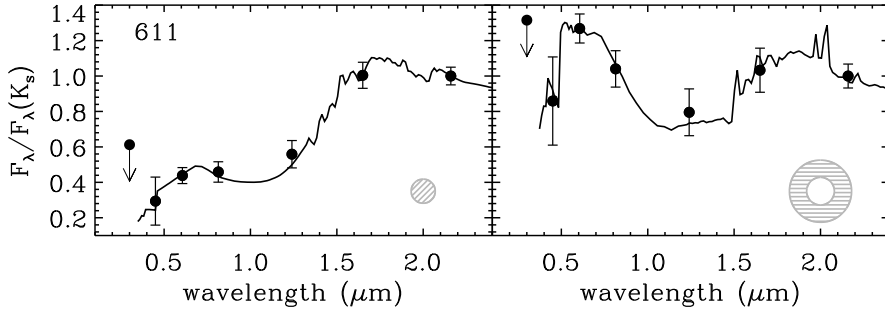


Figure 2 — The spectral energy distribution of source 611 in a $0.7''$ circular diameter aperture (*left*) and in a concentric $1'' - 2''$ diameter ring (*right*), normalized to the K_s -band flux. Overplotted are independent model fits from Rudnick et al. (2003a).

ponential disks and scattered, UV-bright star forming regions. Some even show evidence of well-developed grand-design spiral structure. However, the mean central surface brightness of the disks is $1 - 2$ mag higher than that of nearby disk galaxies and the mean rest-frame color ($U - V$) ≈ 0 is ~ 1 mag bluer (c.f. Lilly et al. 1998). Passive evolution can lead to disks with normal surface brightnesses at low redshift. Alternatively, the disks are disrupted later by interactions or evolve into S0's, which have higher surface brightnesses (e.g., Burstein 1979).

5 Spectral Energy Distribution

The overall SEDs of the galaxies show a large variety. Four of the galaxies (257, 267, 494 and 657) satisfy conventional U-dropout criteria (Madau et al. 1996; Giavalisco & Dickinson 2001). One galaxy is at too low redshift (302) to be classified as a U-dropout, and one other galaxy is too faint in the rest-frame UV (611). It has $J_s - K_s > 2.3$, and is part of the population of evolved galaxies identified by Franx et al (2003) and van Dokkum et al (2003).

The red colors of the central components can be due to either dust, higher age, emission line contamination, or a combination of effects. The SEDs show stronger Balmer/4000 Å breaks in the inner parts than in the outer parts (see Fig. 2 for an example). We derived colors inside and outside of a 0.7 arcsec radius centered on the K_s -band center. The mean differences are ~ 0.5 mag in observed $I_{814} - K_s$ and ~ 0.2 mag in rest-frame $U - V$. Higher resolution NICMOS data are required to address the population differences in more detail.

6 Discussion

We have found 6 large galaxies with characteristics similar to those of nearby disk galaxies: exponential profiles with large scale lengths, more regular and centrally concentrated morphologies in the restframe optical than in the rest-frame UV, and, as a result, red nuclei. It is very tempting to classify these galaxies as disk galaxies, given the similarities with low redshift disk galaxies. However, kinematic studies are necessary to confirm that the material is in a rotating disk. Photometric studies of larger samples are needed to constrain the thickness of the disks. We note that simulations of Steinmetz & Navarro (2002) can show extended structures during a merging or accretion event. The expect duration of this phase is short, however, while the disk galaxies comprise a high fraction of the bright objects. It is therefore unlikely that a significant fraction of the galaxies presented here are undergoing such an event.

The density of these large disk galaxies is fairly low: over a survey area of 4.7 arcmin² and to a magnitude limit of $K_{s,tot} = 22$ they make up 6 out of 52 galaxies at $1.4 \lesssim z \lesssim 3.0$. However, they do constitute 6 out of the 12 most rest-frame luminous galaxies $L_V \gtrsim 6 \times 10^{10} h_{70}^{-2} L_{\odot}$ in the same redshift range. The comoving volume density is $\sim 3 \times 10^{-4} h_{70}^3 \text{ Mpc}^{-3}$ at a mean redshift $z \approx 2.3$. Obviously, larger area surveys are needed to establish the true density. We note that three of the galaxies only have photometric redshifts, one of which (a U-drop galaxy at $z_{phot} = 1.82$) is in the poorly tested range $1.4 < z < 2.0$. The volume density of disk galaxies with $r_e > 3.6 h_{70}^{-1} \text{ kpc}$ in the local universe is much higher at $\sim 3 \times 10^{-3} h_{70}^3 \text{ Mpc}^{-3}$ (de Jong 1996), although many nearby disks would not be present in our high redshift sample because their surface brightness would be too low.

We note that similar galaxies are absent in the very deep Near-IR imaging data on the HDF-N (Williams et al. 1996; Dickinson 2000). Although notable differences between optical and NIR morphologies were reported for two of the largest LBGs in the HDF-N, no large galaxies were reported to have red nuclei and exponential profiles as in the HDF-S. The two fields are different in other aspects as well. We found earlier that the HDF-N is deficient in red sources (e.g., Labbé et al. 2003 versus Papovich, Dickinson, & Ferguson 2001) and the disk galaxies are more luminous in the H -band than most of the high-redshift galaxies found in the HDF-N (Papovich, Dickinson, & Ferguson 2001). The larger number of red galaxies in the HDF-S (e.g., Labbé et al. 2003, Franx et al. 2003) and their strong clustering (Daddi et al. 2003) may indicate that the red galaxies and large disks are both part of the same structures with high overdensities, and evolve into the highest overdensities at low redshift, i.e. clusters. If this is the case, the large disk galaxies may be the progenitor of large S0 galaxies in the nearby clusters, which have very similar colors as elliptical galaxies (e.g., Bower, Lucey, & Ellis 1992).

Finally, we can compare the observed disk sizes to theoretical predictions. It

is often assumed (Fall & Efstathiou 1980; Mo, Mao, & White 1998) that the disk scale length is determined by the spin parameter λ and the circular velocity of the virialized dark matter halo (Fall & Efstathiou 1980; Mo, Mao, & White 1998). For a Λ CDM cosmology, Mo, Mao, & White (1999) predict that the $z \sim 3$ space density of large ($r_e \gtrsim 3.6 h_{70}^{-1}$ kpc) bright U-dropouts is $1.1 \times 10^{-4} h_{70}^3 \text{ Mpc}^{-3}$, whereas we find $\sim 2 \times 10^{-4} h_{70}^3 \text{ Mpc}^{-3}$ for our 4 U-drops at a mean $\langle z \rangle \sim 2.4$. This difference of a factor of two is not very significant, given our low number statistics and small survey volume. The combination of sizes and rotation velocities will give much stronger constraints on these models; it may be possible to measure the kinematics of some of these large galaxies using NIR spectrographs on large telescopes.

Acknowledgments

We thank the staff at ESO for their hard work in taking these data and making them available. This research was supported by grants from the Netherlands Foundation for Research (NWO), the Leids Kerkhoven-Bosscha Fonds, and the Lorentz Center. GR thanks Frank van den Bosch and Jarle Brinchmann for useful discussions.

References

- Barden et al., 2003, ApJ, submitted (astro-ph/0302392)
 Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
 Bouwens, R. & Silk, J. 2002, ApJ, 568, 522
 Bower, R. G., Lucey, J. R., & Ellis, R. S. 1992, MNRAS, 254, 601
 Burstein, D. 1979, ApJ, 234, 435
 Casertano, S. et al., AJ, 120, pp. 2747–2824, 2000
 Daddi, E., et al., ApJ, in press
 de Jong, R. S. 1996, A&A, 313, 45
 Dickinson, M., *Philos. Trans. R. Soc. London* **A 358**, p. 2001, 2000
 Erb et al., 2003, ApJ, in press
 Fall, S. M. & Efstathiou, G. 1980, MNRAS, 193, 189
 Förster Schreiber, N.M. et al. 2003, in preparation
 Franx, M. et al., *The Messenger* **99**, pp. 20–22, 2000
 Franx, M. et al. 2003, ApJ, 587, L79
 Giavalisco, M., Steidel, C. C., & Macchetto, F. D. 1996, ApJ, 470, 189
 Giavalisco, M. & Dickinson, M. 2001, ApJ, 550, 177
 Labbé, I. et al. 2003, AJ, 125, 1107
 Lilly, S. et al. 1998, ApJ, 500, 75
 Lowenthal, J. D. et al. 1997, ApJ, 481, 673
 Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A., MNRAS283, pp. 1388–1404, 1996
 Mao, S., Mo, H. J., & White, S. D. M. 1998, MNRAS, 297, L71
 Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319
 Mo, H. J., Mao, S., & White, S. D. M. 1999, MNRAS, 304, 175

- Moorwood, A. F. 1997, *Proc. SPIE*, 2871, 1146
- Moorwood, A., van Der Werf, P., Cuby, J.-G., Oliva, E. 2003, *Proceedings of the Workshop on The Mass of Galaxies at Low and High Redshift*, eds. R. Bender and A. Renzini, Springer-Verlag, p302.
- Steinmetz, M. & Navarro, J. F. 2002, *New Astronomy*, 7, 155
- Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, *ApJ*, 559, 620
- Rudnick, G. et al. 2001, *AJ*, 122, 2205
- Rudnick, G. et al. 2003a, in preparation
- Rudnick, G. et al. 2003b, in preparation
- Schade, D., Lilly, S. J., Le Fevre, O., Hammer, F., & Crampton, D. 1996, *ApJ*, 464, 79
- Simard, L. et al. 1999, *ApJ*, 519, 563
- Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, *AJ*, 112, 352
- Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, *ApJ*, 462, L17
- Trujillo et al. 2003, in preparation
- van Dokkum, P. G. & Stanford, S. A. 2001, *ApJ*, 562, L35
- van Dokkum, P. G. et al. 2003, *ApJ*, 587, L83
- Vanzella, E. et al. 2002, *A&A*, 396, 847
- Vogt, N. P., Forbes, D. A., Phillips, A. C., Gronwall, C., Faber, S. M., Illingworth, G. D., & Koo, D. C. 1996, *ApJ*, 465, L15
- Vogt, N. P. et al. 1997, *ApJ*, 479, L121
- Williams, R. E. et al. 1996, *AJ*, 112, 1335

The Rest-Frame Optical Luminosity Density Color, and Stellar Mass Density of the Universe from $z=0$ to $z=3$

ABSTRACT

We present the evolution of the rest-frame optical luminosity density, $j_{\lambda}^{\text{rest}}$, of the integrated rest-frame optical color, and of the stellar mass density, ρ_* , for a sample of K_s -band selected galaxies in the HDF-S. We derived $j_{\lambda}^{\text{rest}}$ in the rest-frame U , B , and V -bands and found that $j_{\lambda}^{\text{rest}}$ increases by a factor of 1.9 ± 0.4 , 2.9 ± 0.6 , and 4.9 ± 1.0 in the V , B , and U rest-frame bands respectively between a redshift of 0.1 and 3.2. We derived the luminosity weighted mean cosmic $(U - B)_{\text{rest}}$ and $(B - V)_{\text{rest}}$ colors as a function of redshift. The colors bluen almost monotonically with increasing redshift; at $z = 0.1$, the $(U - B)_{\text{rest}}$ and $(B - V)_{\text{rest}}$ colors are 0.16 and 0.75 respectively, while at $z = 2.8$ they are -0.39 and 0.29 respectively. We derived the luminosity weighted mean M/L_V^* using the correlation between $(U - V)_{\text{rest}}$ and $\log_{10} M/L_V^*$ which exists for a range in smooth SFHs and moderate extinctions. We have shown that the mean of individual M/L_V^* estimates can overpredict the true value by $\sim 70\%$ while our method overpredicts the true values by only $\sim 35\%$. We find that the universe at $z \sim 3$ had ~ 10 times lower stellar mass density than it does today in galaxies with $L_V^{\text{rest}} > 1.4 \times 10^{10} h_{70}^{-2} L_{\odot}$. 50% of the stellar mass of the universe was formed by $z \sim 1 - 1.5$. The rate of increase in ρ_* with decreasing redshift is similar to but above that for independent estimates from the HDF-N, but is slightly less than that predicted by the integral of the SFR(z) curve.

Gregory Rudnick, Hans-Walter Rix, Marijn Franx, Ivo Labbé, Michael Blanton,
Emanuele Daddi, Natascha M. Förster Schreiber, Alan Moorwood, Huub
Röttgering, Ignacio Trujillo, Arjen van de Wel, Paul van der Werf, Pieter G. van
Dokkum, & Lottie van Starckenburg,
Astrophysical Journal, **599**, 847 (2003)

1 Introduction

A primary goal of galaxy evolution studies is to elucidate how the stellar content of the present universe was assembled over time. Enormous progress has been made in this field over the past decade, driven by advances over three different redshift ranges. Large scale redshift surveys with median redshifts of $z \sim 0.1$ such as the Sloan Digital Sky Survey (SDSS; York et al. 2000) and the 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2001), coupled with the near infrared (NIR) photometry from the 2 Micron All Sky Survey (2MASS; Skrutskie et al. 1997), have recently been able to assemble the complete samples, with significant co-moving volumes, necessary to establish crucial local reference points for the local luminosity function (e.g. Folkes et al. 1999; Blanton et al. 2001; Norberg et al. 2002; Blanton et al. 2003c) and the local stellar mass function of galaxies (Cole et al. 2001; Bell et al. 2003a).

At $z \lesssim 1$, the pioneering study of galaxy evolution was the Canada France Redshift Survey (CFRS; Lilly et al. 1996). The strength of this survey lay not only in the large numbers of galaxies with confirmed spectroscopic redshifts, but also in the I -band selection, which enabled galaxies at $z \lesssim 1$ to be selected in the rest-frame optical, the same way in which galaxies are selected in the local universe.

At high redshifts the field was revolutionized by the identification, and subsequent detailed follow-up, of a large population of star-forming galaxies at $z > 2$ (Steidel et al. 1996). These Lyman Break Galaxies (LBGs) are identified by the signature of the redshifted break in the far UV continuum caused by intervening and intrinsic neutral hydrogen absorption. There are over 1000 spectroscopically confirmed LBGs at $z > 2$, together with the analogous U-dropout galaxies identified using Hubble Space Telescope (HST) filters. The individual properties of LBGs have been studied in great detail. Estimates for their star formation rates (SFRs), extinctions, ages, and stellar masses have been estimated by modeling the broad band fluxes (Sawicki & Yee 1998; hereafter SY98; Papovich et al. 2001; hereafter P01; Shapley et al. 2001). Independent measures of their kinematic masses, metallicities, SFRs, and initial mass functions (IMFs) have been determined using rest-frame UV and optical spectroscopy (Pettini et al. 2000; Shapley et al. 2001; Pettini et al. 2001, 2002; Erb et al. 2003).

Despite these advances, it has proven difficult to reconcile the ages, SFRs, and stellar masses of individual galaxies at different redshifts within a single galaxy formation scenario. Low redshift studies of the fundamental plane indicate that the stars in elliptical galaxies must have been formed by $z > 2$ (e.g., van Dokkum et al. 2001) and observations of evolved galaxies at $1 < z < 2$ indicate that the present population of elliptical galaxies was already in place at $z \gtrsim 2.5$ (e.g., Benítez et al. 1999; Cimatti et al. 2002; but see Zepf 1997). In contrast, studies of star forming Lyman Break Galaxies spectroscopically confirmed to lie at $z > 2$ (LBGs; Steidel et al. 1996, 1999) claim that LBGs are uniformly very young and a factor of

10 less massive than present day L_* galaxies (e.g., SY98; P01; Shapley et al. 2001).

An alternative method of tracking the build-up of the cosmic stellar mass is to measure the total emissivity of all relatively unobscured stars in the universe, thus effectively making a luminosity weighted mean of the galaxy population. This can be partly accomplished by measuring the evolution in the global luminosity density $j(z)$ from galaxy redshift surveys. Early studies at intermediate redshift have shown that the rest-frame UV and B -band $j(z)$ are steeply increasing out to $z \sim 1$ (e.g., Lilly et al. 1996; Fried et al. 2001). Wolf et al. (2003) has recently measured $j(z)$ at $0 < z < 1.2$ from the COMBO-17 survey using $\sim 25,000$ galaxies with redshifts accurate to ~ 0.03 and a total area of 0.78 degrees. At rest-frame $2800\text{\AA}$ these measurements confirm those of Lilly et al. (1996) but do not support claims for a shallower increase with redshift which goes like $(1+z)^{1.5}$ as claimed by Cowie, Songaila, & Barger (1999) and Wilson et al. (2002). On the other hand, the B -band evolution from Wolf et al. (2003) is only a factor of ~ 1.6 between $0 < z < 1$, considerably shallower than the factor of ~ 3.75 increase seen by Lilly et al. (1996). At $z > 2$ measurements of the rest-frame UV $j(z)$ have been made using the optically selected LBG samples (e.g., Madau et al. 1996; Sawicki, Lin, & Yee 1997; Steidel et al. 1999; Poli et al. 2001) and NIR selected samples (Kashikawa et al. 2003; Poli et al. 2003; Thompson 2003) and, with modest extinction corrections, the most recent estimates generically yield rest-frame UV $j(z)$ curves which, at $z > 2$, are approximately flat out to $z \sim 6$ (cf. Lanzetta et al. 2002). Dickinson et al. (2003; hereafter D03) have used deep NIR data from NICMOS in the HDF-N to measure the rest-frame B -band luminosity density out to $z \sim 3$, finding that it remained constant to within a factor of ~ 3 . By combining $j(z)$ measurements at different rest-frame wavelengths and redshifts, Madau, Pozzetti, & Dickinson (1998) and Pei, Fall, & Hauser (1999) modeled the emission in all bands using an assumed global SFH and used it to constrain the mean extinction, metallicity, and IMF. Bolzonella, Pelló, & Maccagni (2002) measured NIR luminosity functions in the HDF-N and HDF-S and find little evolution in the bright end of the galaxy population and no decline in the rest-frame NIR luminosity density out to $z \sim 2$. In addition, Baldry et al. (2002) and Glazebrook et al. (2003a) have used the mean optical cosmic spectrum at $z \sim 0$ from the 2dFGRS and the SDSS respectively to constrain the cosmic star formation history.

Despite the wealth of information obtained from studies of the integrated galaxy population, there are major difficulties in using these many disparate measurements to re-construct the evolution in the stellar mass density. First, and perhaps most important, the selection criteria for the low and high redshift surveys are usually vastly different. At $z < 1$ galaxies are selected by their rest-frame optical light. At $z > 2$, however, the dearth of deep, wide-field NIR imaging has forced galaxy selection by the rest-frame UV light. Observations in the rest-frame UV are much more sensitive to the presence of young stars and extinction than observations in the rest-frame optical. Second, state-of-the-art deep surveys have only been performed in small fields and the effects of field-to-field variance at faint magnitudes, and in the rest-frame optical, are not well understood.

In the face of field-to-field variance, the globally averaged rest-frame color may be a more robust characterization of the galaxy population than either the luminosity density or the mass density because it is, to the first order, insensitive to the exact density normalization. At the same time, it encodes information about the dust obscuration, metallicity, and SFH of the cosmic stellar population. It therefore provides an important constraint on galaxy formation models which may be reliably determined from relatively small fields.

To track consistently the globally averaged evolution of the galaxies which dominate the stellar mass budget of the universe – as opposed to the UV luminosity budget – over a large redshift range a different strategy than UV selection must be adopted. It is not only desirable to measure $j(z)$ in a constant rest-frame optical bandpass, but it is also necessary that galaxies be selected by light redward of the Balmer/4000Å break, where the light from older stars contributes significantly to the SED. To accomplish this, we obtained ultra-deep NIR imaging of the WFPC2 field of the HDF-S (Casertano et al. 2000) with the Infrared Spectrograph And Array Camera (ISAAC; Moorwood et al. 1997) at the Very Large Telescope (VLT) as part of the Faint Infrared Extragalactic Survey (FIRES; Franx et al. 2000). The FIRES data on the HDF-S, detailed in Labbé et al. (2003; hereafter L03), provide us with the deepest ground-based J_s and H data and the overall deepest K_s -band data in any field allowing us to reach rest-frame optical luminosities in the V -band of $\sim 0.6 L_{*}^{\text{local}}$ at $z \sim 3$. First results using a smaller set of the data were presented in Rudnick et al. (2001; hereafter R01). The second FIRES field, centered on the $z = 0.83$ cluster MS1054-03, has ~ 1 magnitude less depth but ~ 5 times greater area (Förster Schreiber et al. 2003).

In the present work we will draw on photometric redshift estimates, z_{phot} for the K_s -band selected sample in the HDF-S (R01; L03), and on the observed SEDs, to derived rest-frame optical luminosities $L_{\lambda}^{\text{rest}}$ for a sample of galaxies selected by light redder than the rest-frame optical out to $z \sim 3$. In § 2 we describe the observations, data reduction, and the construction of a K_s -band selected catalog with $0.3 - 2.2\mu\text{m}$ photometry, which selects galaxies at $z < 4$ by light redward of the 4000Å break. In § 3 we describe our photometric redshift technique, how we estimate the associated uncertainties in z_{phot} , and how we measure $L_{\lambda}^{\text{rest}}$ for our galaxies. In § 4 we use our measures of $L_{\lambda}^{\text{rest}}$ for the individual galaxies to derive the mean cosmic luminosity density, $j_{\lambda}^{\text{rest}}$ and the cosmic color and then use these to measure the stellar mass density ρ_{*} as a function of cosmic time. We discuss our results in § 5 and summarize in § 6. Throughout this paper we assume $\Omega_{\text{M}} = 0.3$, $\Omega_{\Lambda} = 0.7$, and $H_0 = 70 \text{ h}_{70} \text{ km s}^{-1} \text{ Mpc}^{-1}$ unless explicitly stated otherwise.

2 Data

A complete description of the FIRES observations, reduction procedures, and the construction of photometric catalogs is presented in detail in L03; we outline the

important steps below.

Objects were detected in the K_s -band image with version 2.2.2 of the SExtractor software (Bertin & Arnouts 1996). For consistent photometry between the space and ground-based data, all images were then convolved to $0''.48$, the seeing in our worst NIR band. Photometry was then performed in the U_{300} , B_{450} , V_{606} , I_{814} , J_s , H , and K_s -band images using specially tailored isophotal apertures defined from the detection image. In addition, a measurement of the total flux in the K_s -band, $K_{s,AB}^{\text{tot}}$, was obtained using an aperture based on the SExtractor *AUTO* aperture¹. Our effective area is 4.74 square arcminutes, including only areas of the chip which were well exposed. All magnitudes are quoted in the Vega system unless specifically noted otherwise. Our adopted conversions from Vega system to the AB system are $J_{s,vega} = J_{s,AB} - 0.90$, $H_{vega} = H_{AB} - 1.38$, and $K_{s,vega} = K_{s,AB} - 1.86$ (Bessell & Brett 1988).

3 Measuring Photometric Redshifts and Rest-Frame Luminosities

3.1 Photometric Redshift Technique

We estimated z_{phot} from the broad-band SED using the method described in R01, which attempts to fit the observed SED with a linear combination of redshifted galaxy templates. We made two modifications to the R01 method. First, we added an additional template constructed from a 10 Myr old, single age, solar metallicity population with a Salpeter (1955) initial mass function (IMF) based on empirical stellar spectra from the 1999 version of the Bruzual A. & Charlot (1993) stellar population synthesis code. Second, a 5% minimum flux error was adopted for all bands to account for the night-to-night uncertainty in the derived zeropoints and for template mismatch effects, although in reality both of these errors are non-gaussian.

Using 39 galaxies with reliable FIRES photometry and spectroscopy available from Cristiani et al. (2000), Rigopoulou et al. (2000), Glazebrook (2003b)², Vanzella et al. (2002), and Rudnick et al. (2003) we measured the redshift accuracy of our technique to be $\langle |z_{spec} - z_{phot}| / (1 + z_{spec}) \rangle = 0.09$ for $z < 3$. There is one galaxy at $z_{spec} = 2.025$ with $z_{phot} = 0.12$ but with a very large internal z_{phot} uncertainty. When this object is removed, $\langle |z_{spec} - z_{phot}| / (1 + z_{spec}) \rangle = 0.05$ at $z_{spec} > 1.3$.

For a given galaxy, the photometric redshift probability distribution can be highly non-Gaussian and contain multiple χ^2 minima at vastly different redshifts. An accurate estimate of the error in z_{phot} must therefore not only contain the two-

¹The reduced images, photometric catalogs, photometric redshift estimates, and rest-frame luminosities are available online through the FIRES homepage at <http://www.strw.leidenuniv.nl/~fires>.

²available at <http://www.aao.gov.au/hdfs/>

sided confidence interval in the local χ^2 minimum, but also reflect the presence of alternate redshift solutions. The difficulties of measuring the uncertainty in z_{phot} were discussed in R01 and will not be repeated in detail here. To improve on R01, however, we have developed a Monte Carlo method which takes into account, on a galaxy-by-galaxy basis, flux errors and template mismatch. These uncertainty estimates are called δz_{phot} . For a full discussion of this method see Appendix Appendix A..

Galaxies with $K_{s,AB}^{tot} \geq 25$ have such high photometric errors that the z_{phot} estimates can be very uncertain. At $K_{s,AB}^{tot} < 25$, however, objects are detected at better than the 10-sigma level and have well measured NIR SEDs, important for locating redshifted optical breaks. For this reason, we limited our catalog to the 329 objects that have $K_{s,AB}^{tot} < 25$, lie on well exposed sections of the chip, and are not identified as stars (see §3.1.1).

3.1.1 Star Identification

To identify probable stars in our catalog we did not use the profiles measured from the WFPC2-imaging because it is difficult to determine the size at faint levels. At the same time, we verified that the stellar template fitting technique identified all bright unsaturated stars in the image. Instead, we compared the observed SEDs with those from 135 NextGen version 5.0 stellar atmosphere models described in Hauschildt, Allard, & Baron (1999)³. We used models with $\log(g)$ of 5.5 and 6, effective temperatures ranging from 1600 K to 10,000 K, and metallicities of solar and 1/10th solar. We identified an object as a stellar candidate if the raw χ^2 of the stellar fit was lower than that of the best-fit galaxy template combination. Four of the stellar candidates from this technique (objects 155, 230, 296, and 323) are obviously extended and were excluded from the list of stellar candidates. Two bright stars (objects 39 and 51) were not identified by this technique because they are saturated in the HST images and were added to the list by hand. We ended up with a list of 29 stars that had $K_{s,AB}^{tot} < 25$ and lie on well exposed sections of the chip. These were excluded from all further analysis.

3.2 Rest-Frame Luminosities

To measure the L_{λ}^{rest} of a galaxy one must combine its redshift with the observed SED to estimate the intrinsic SED. In practice, this requires some assumptions about the intrinsic SED.

In R01 we derived rest-frame luminosities from the best-fit combination of spectral templates at z_{phot} , which assumes that the intrinsic SED is well modeled by our template set. We know that for many galaxies the best-fit template matches the position and strength of the spectral breaks and the general shape of the SED. There are, however, galaxies in our sample which show clear residuals from the

³available at <http://dilbert.physast.uga.edu/~yeti/mdwarfs.html>

best fit template combination. Even for the qualitatively good fits, the model and observed flux points can differ by $\sim 10\%$, corresponding to a $\sim 15\%$ error in the derived rest-frame color. As we will see in §4.2, such color errors can cause errors of up to a factor of 1.5 in the V -band stellar mass-to-light ratio, M/L_V^* .

Here we used a method of estimating L_λ^{rest} which does not depend directly on template fits to the data but, rather, interpolates directly between the observed bands using the templates as a guide. We define our rest-frame photometric system in Appendix Appendix B. and explain our method for estimating L_λ^{rest} in Appendix Appendix C..

We plot in Figure 1 the rest-frame luminosities vs. redshift and enclosed volume for the $K_{s,AB}^{\text{tot}} < 25$ galaxies in the FIRES sample. The different symbols represent different δz_{phot} values and since the derived luminosity is tightly coupled to the redshift, we do not independently plot L_λ^{rest} errorbars. The tracks indicate the L_λ^{rest} for different SED types normalized to $K_{s,AB}^{\text{tot}} = 25$, while the intersection of the tracks in each panel indicates the redshift at which the rest-frame filter passes through our K_s -band detection filter. There is a wide range in L_λ^{rest} at all redshifts and there are galaxies at $z > 2$ with L_λ^{rest} much in excess of the local L_* values. Using the full FIRES dataset, we are much more sensitive than in R01; objects at $z \approx 3$ with $K_{s,AB}^{\text{tot}} = 25$ have $L_V^{\text{rest}} \approx 0.6 * L_{*,V}^{\text{local}}$, as defined from the $z=0.1$ sample of Blanton et al. (2003c; hereafter B03). As seen in R01 there are many galaxies at $z > 2$, in all bands, with $L_\lambda^{\text{rest}} \geq L_*^{\text{local}}$. R01 found 10 galaxies at $2 \leq z \leq 3.5$ with $L_B^{\text{rest}} > 10^{11} h_{70}^{-2} L_\odot$ and inferred a brightening in the luminosity function of $\sim 1 - 1.3$ magnitudes. We confirm their result when using the same local luminosity function (Blanton et al. 2001). Although this brightening is biased upwards by photometric redshift errors, we find a similar brightening of approximately ~ 1 magnitude after correction for this effect. As also noticed in R01, we found a deficit of luminous galaxies at $1.5 \lesssim z \lesssim 2$ although this deficit is not as pronounced at lower values of L_λ^{rest} . The photometric redshifts in the HDF-S, however, are not well tested in this regime. To help judge the reality of this deficit we compared our photometric redshifts on an object-by-object basis to those of the Rome group (Fontana et al. 2000)⁴ who derived z_{phot} estimates for galaxies in the HDF-S using much shallower NIR data. We find generally good agreement in the z_{phot} estimates, although there is a large scatter at $1.5 < z < 2.0$. Both sets of photometric redshifts show a deficit in the z_{phot} distribution, although the Rome group's gap is less pronounced than ours and is at a slightly lower redshift. In addition, we examined the photometric redshift distribution of the NIR selected galaxies of D03 in the HDF-N, which have very deep NIR data. These galaxies also showed a gap in the z_{phot} distribution at $z \sim 1.6$. Together these results indicate that systematic effects in the z_{phot} determinations may be significant at $1.5 < z < 2.0$. On the other hand, we also derived photometric redshifts for a preliminary set of data in the MS1054-03 field of the FIRES survey, whose filter set is similar, but which has a U instead of U_{300} filter. In this field,

⁴available at <http://www.mporzio.astro.it/HIGHZ/HDF.html>

no systematic depletion of $1.5 < z < 2$ galaxies was found. It is therefore not clear what role systematic effects play in comparison to field-to-field variations in the true redshift distribution over this redshift range. Obtaining spectroscopic redshifts at $1.5 < z < 2$ is the only way to judge the accuracy of the z_{phot} estimates in this regime.

We have also split the points up according to whether or not they satisfied the U-dropout criteria of Giavalisco & Dickinson (2001) which were designed to pick unobscured star-forming galaxies at $z \gtrsim 2$. As expected from the high efficiency of the U-dropout technique, we find that only 15% of the 57 classified U-dropouts have $z_{phot} < 2$. As we will discuss in §4.1 we measured the luminosity density for objects with $L_V^{rest} > 1.4 \times 10^{10} h_{70}^{-2} L_\odot$. Above this threshold, there are 62 galaxies with $2 < z < 3.2$, of which 26 are not classified as U-dropouts. These non U-dropouts number among the most rest-frame optically luminous galaxies in our sample. In fact, the most rest-frame optically luminous object at $z < 3.2$ (object 611) is a galaxy which fails the U-dropout criteria. 10 of these 26 objects, including object 611, also have $J - K > 2.3$, a color threshold which has been shown by Franx et al. (2003) and van Dokkum et al. (2003) to efficiently select galaxies at $z > 2$. These galaxies are not only luminous but also have red rest-frame optical colors, implying high M/L^* values. Franx et al. (2003) showed that they likely contribute significantly ($\sim 43\%$) to the stellar mass budget at high redshifts.

3.2.1 Emission Lines

There will be emission line contamination of the rest-frame broad-band luminosities when rest-frame optical emission lines contribute significantly to the flux in our observed filters. P01 estimated the effect of emission lines in the NICMOS F160W filter and the K_s filter and found that redshifted, rest-frame optical emission lines, whose equivalent widths are at the maximum end of those observed for starburst galaxies (rest-frame equivalent width $\sim 200\text{\AA}$), can contribute up to 0.2 magnitudes in the NIR filters. In addition, models of emission lines from Charlot & Longhetti (2001) show that emission lines will tend to drive the $(U - B)_{rest}$ color to the blue more easily than the $(B - V)_{rest}$ color for a large range of models. Using the UBV photometry and spectra of nearby galaxies from the Nearby Field Galaxy Survey (NFGS; Jansen et al. 2000a; Jansen et al. 2000b) we computed the actual correction to the $(U - B)_{rest}$ and $(B - V)_{rest}$ colors as a function of $(B - V)_{rest}$. For the bluest galaxies in $(B - V)_{rest}$, emission lines bluen the $(U - B)_{rest}$ colors by ~ 0.05 and the $(B - V)_{rest}$ colors only by < 0.01 . Without knowing beforehand the strength of emission lines in any of our galaxies, we corrected our rest-frame colors based on the results from Jansen et al. We ignored the very small correction to the $(B - V)_{rest}$ colors and corrected the $(U - B)_{rest}$ colors using the equation:

$$(U - B)_{corrected} = (U - B) - 0.0658 \times (B - V) + 0.0656 \quad (1)$$

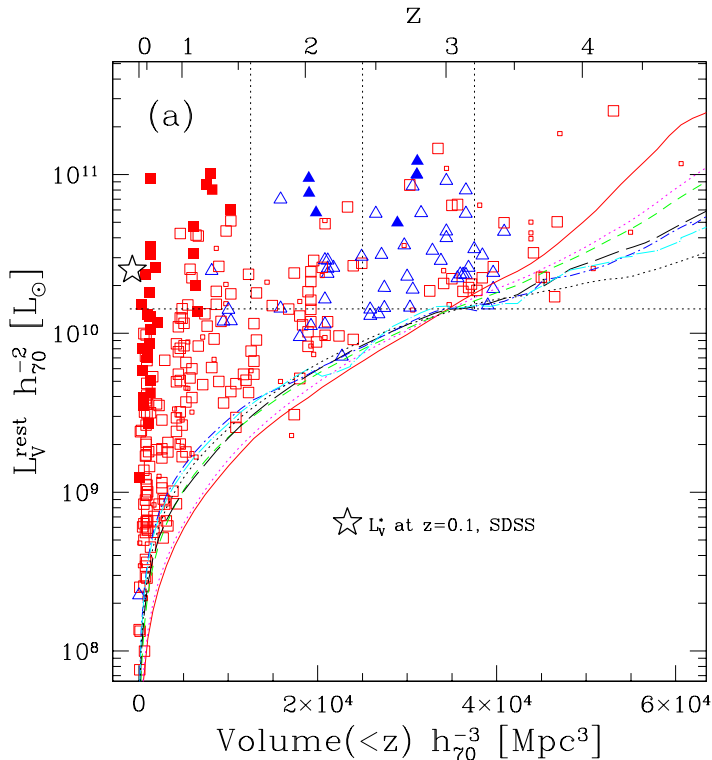
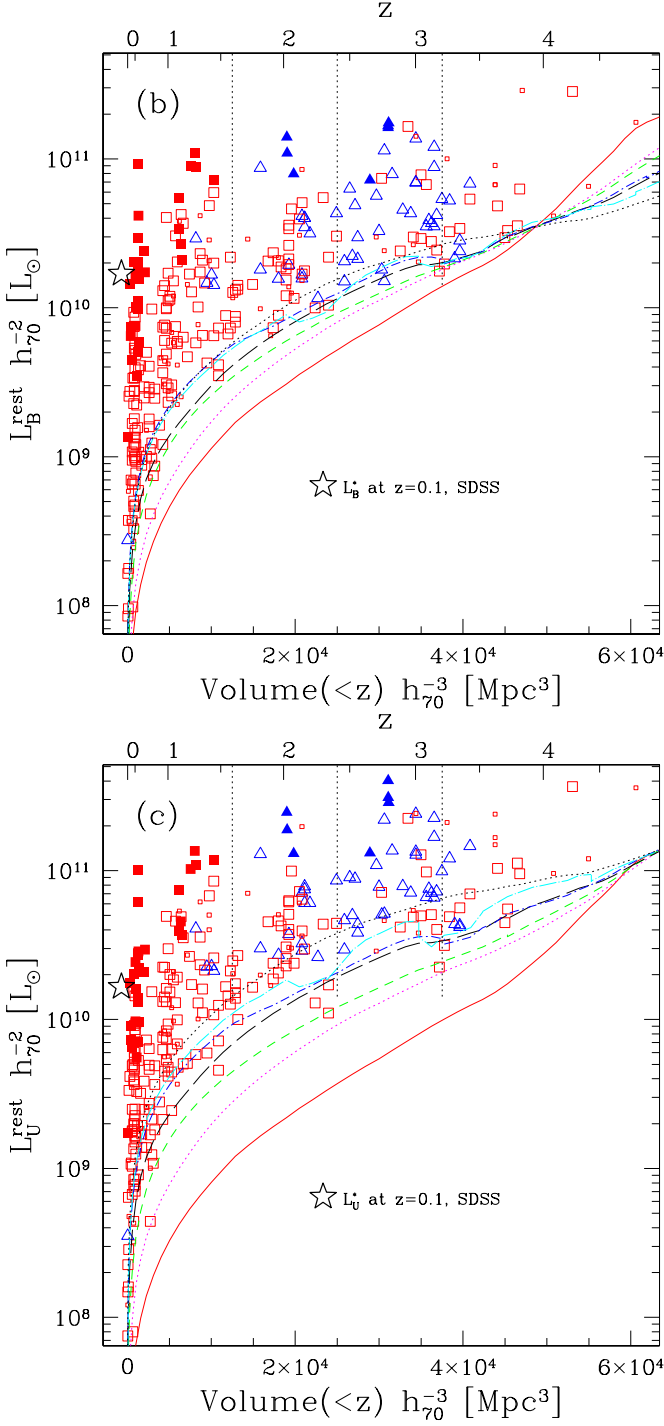


Figure 1 — The distribution of rest-frame V , B , and U -band luminosities as a function of enclosed co-moving volume and z_{phot} is shown in figures (a), (b), and (c) respectively for galaxies with $K_{s,AB}^{tot} < 25$. Galaxies which have spectroscopic redshifts are represented by solid points and for these objects L_{λ}^{rest} is measured at z_{spec} . Large symbols have $\delta z_{phot}/(1 + z_{phot}) < 0.16$ and small symbols have $\delta z_{phot}/(1 + z_{phot}) \geq 0.16$. Triangle points would be classified as U-dropouts according to the selection of Giavalisco & Dickinson (2001). As is expected, most of the galaxies selected as U-dropouts have $z_{phot} \gtrsim 2$ which would not be selected as U-dropouts. The large stars in each panel indicate the value of L_{*}^{local} from Blanton et al. (2003c). In the V -band we are sensitive to galaxies at 60% of L_{*}^{local} even at $z \sim 3$ and there are galaxies at $z_{phot} \geq 2$ with $L_{\lambda}^{rest} \geq 10^{11} h_{70}^{-2} L_{\odot}$. The tracks represent the values of L_{λ}^{rest} for our seven template spectra normalized at each redshift to $K_{s,AB}^{tot} = 25$. The specific tracks correspond to the E (solid), Sbc (dot), Scd (short dash), Im (long dash), SB1 (dot–short dash), SB2 (dot–long dash), and 10my (dot) templates. The horizontal dotted line in (a) indicates the luminosity threshold L_V^{thresh} above which we measure the rest-frame luminosity density j_{λ}^{rest} and the vertical dotted lines in each panel mark the redshift boundaries of the regions for which we measure j_{λ}^{rest} .

which corresponds to a linear fit to the NFGS data. These effects might be greater for objects with strong AGN contribution to their fluxes.



4 The Properties of the Massive Galaxy Population

In this section we discuss the use of the z_{phot} and $L_{\lambda}^{\text{rest}}$ estimates to derive the integrated properties of the population, namely the luminosity density, the mean cosmic rest-frame color, the stellar mass-to-light ratio M/L^* , and the stellar mass density ρ_* . As will be described below, addressing the integrated properties of the population reduces many of the uncertainties associated with modeling individual galaxies and, in the case of the cosmic color, is less sensitive to field-to-field variations.

4.1 The Luminosity Density

Using our $L_{\lambda}^{\text{rest}}$ estimates from the $K_{s,AB}^{\text{tot}} < 25$ galaxies (see §3.2), we traced the redshift evolution of the rest-frame optically most luminous, and therefore presumably most massive, galaxies by measuring the rest-frame luminosity density $j_{\lambda}^{\text{rest}}$ of the visible stars associated with them. The results are presented in Table 1 and plotted against redshift and elapsed cosmic time in Figure 2. As our best alternative to a selection by galaxy mass, we selected our galaxies in our reddest rest-frame band available at $z \sim 3$, i.e. the V -band. In choosing the z and $L_{\lambda}^{\text{rest}}$ regime over which we measured $j_{\lambda}^{\text{rest}}$ we wanted to push to as high of a redshift as possible with the double constraint that the redshifted rest-frame filter still overlapped with the K_s filter and that we were equally complete at all considered redshifts. By choosing an L_V^{rest} threshold, $L_V^{\text{thresh}} = 1.4 \times 10^{10} h_{70}^{-2} L_{\odot}$, and a maximum redshift of $z = 3.2$, we could select galaxies down to $0.6 L_{*,V}^{\text{local}}$ with constant efficiency regardless of SED type. We then divided the range out to $z = 3.2$ into three bins of equal co-moving volume which correspond to the redshift intervals 0–1.6, 1.6–2.41, and 2.41–3.2.

In a given redshift interval, we estimated $j_{\lambda}^{\text{rest}}$ directly from the data in two steps. We first added up all the luminosities of galaxies which satisfied our L_V^{thresh} criteria defined above and which had $\delta z_{phot} / (1 + z_{phot}) \leq 0.16$, roughly twice the mean disagreement between z_{phot} and z_{spec} (see §3.1). Galaxies rejected by our δz_{phot} cut but with $L_V^{\text{rest}} > L_V^{\text{thresh}}$, however, contribute to the total luminosity although they are not included in this first estimate. Under the assumption that these galaxies are drawn from the same luminosity function as those which passed the δz_{phot} cut, we computed the total luminosity, including the light from the n_{acc} accepted galaxies and the light lost from the n_{rej} rejected galaxies as

$$L_{\text{tot}} = L_{\text{meas}} \times \left(1 + \frac{n_{rej}}{n_{acc}}\right). \quad (2)$$

As a test of the underlying assumption of this correction we performed a K-S test on the distributions of K_s magnitudes for the rejected and accepted galaxies in each of our three volume bins. In all three redshift bins, the rejected galaxies have $K_{s,AB}^{\text{tot}}$ distributions which are consistent at the $> 90\%$ level with being drawn from the same magnitude distribution as the accepted galaxies. The total correction

per volume bin ranges from 5 – 10% in every bin. Our results do not change if we omit those galaxies whose photometric redshift probability distribution indicates a secondary minima in chisquared.

Uncertainties in the luminosity density were computed by bootstrapping from the $K_{s,AB}^{\text{tot}} < 25$ subsample. This method does not take cosmic variance into account and the errors may therefore underestimate the true error, which includes field-to-field variance.

Redshift errors might effect the luminosity density in a systematic way, as they produce a large error in the measured luminosity. This, combined with a steep luminosity function can bias the observed luminosities upwards, especially at the bright end. This effect can be corrected for in a full determination of the luminosity function (e.g., Chen et al. 2003), but for our application we estimated the strength of this effect by Monte-Carlo simulations. When we used the formal redshift errors we obtained a very small bias (6%), if we increase the photometric redshift errors in the simulation to be minimally as large as $0.08 * (1 + z)$, we still obtain a bias in the luminosity density on the order of 10% or less.

Because we exclude galaxies with faint rest-frame luminosities or low apparent magnitudes, and do not correct for this incompleteness, our estimates should be regarded as lower limits on the total luminosity density. One possibility for estimating the total luminosity density would be to fit a luminosity function as a function of redshift and then integrate it over the whole luminosity range. We don't go faint enough at high redshift, however, to tightly constrain the faint-end slope α . Because extrapolation of $j_{\lambda}^{\text{rest}}$ to arbitrarily low luminosities is very dependent on the value of α , we choose to use this simple and direct method instead. Including all galaxies with $K_{s,AB}^{\text{tot}} < 25$ raises the $j_{\lambda}^{\text{rest}}$ values in the $z = 0 - 1.6$ redshift bin by 86%, 74%, and 66% in the U , B , and V -bands respectively. Likewise, the $j_{\lambda}^{\text{rest}}$ values would increase by 38%, 35%, and 44% for the $z = 1.6 - 2.41$ bin and would increase by 5%, 5%, and 2% for the $z = 2.41 - 3.2$ bin, again in the U , B , and V -bands respectively.

The dip in the luminosity density in the second lowest redshift bin of the (a) and (b) panels of Figure 2 can be traced to the lack of intrinsically luminous galaxies at $z \sim 1.5 - 2$ (§3.2; R01). The dip is not noticeable in the U -band because the galaxies at $z \sim 2$ are brighter with respect to the $z < 1.6$ galaxies in the U -band than in the V or B -band, i.e. they have bluer $(U - B)_{\text{rest}}$ colors and $(U - V)_{\text{rest}}$ colors than galaxies at $z < 1.6$. This lack of rest-frame optically bright galaxies at $z \sim 1.5 - 2$ may result from systematics in the z_{phot} estimates, which are poorly tested in this regime and where the Lyman break has not yet entered the U_{300} filter, or may reflect a true deficit in the redshift distribution of K_s -band luminous galaxies (see §3.2).

At $z \lesssim 1$ our survey is limited by its small volume. For this reason, we supplement our data with other estimates of $j_{\lambda}^{\text{rest}}$ at $z \lesssim 1$.

We compared our results with those of the SDSS as follows. First, we selected

SDSS Main sample galaxies (Strauss et al. 2002) with redshifts in the SDSS Early Data Release (Stoughton et al. 2002) in the EDR sample provided by and described by Blanton et al. (2003a). Using the product `kcorrect v1_16` (Blanton et al. 2003b), for each galaxy we fit an optical SED to the $^{0.1}u^{0.1}g^{0.1}r^{0.1}i^{0.1}z$ magnitudes, after correcting the magnitudes to $z = 0.1$ for evolution using the results of Blanton et al. (2003c). We projected this SED onto the *UBV* filters as described by Bessell (1990) to obtain absolute magnitudes in the *UBV* Vega-relative system. Using the method described in Blanton et al. (2003a) we calculated the maximum volume $V_{\max}$ within the EDR over which each galaxy could have been observed, accounting for the survey completeness map and the flux limit as a function of position. $1/V_{\max}$ then represents the number density contribution of each galaxy. From these results we constructed the number density distribution of galaxies as a function of color and absolute magnitude and the contribution to the uncertainties in those densities from Poisson statistics. While the Poisson errors in the SDSS are negligible, cosmic variance does contribute to the uncertainties. For a more realistic error estimate, we use the fractional errors on the luminosity density from Blanton et al. (2003c). For the SDSS luminosity function, our L_V^{thresh} encompasses 54% of the total light.

In Figure 2 we also show the $j_{\lambda}^{\text{rest}}$ measurements from the COMBO-17 survey (Wolf et al. 2003). We used a catalog with updated redshifts and 29471 galaxies at $z < 0.9$, of which 7441 had $L_V^{\text{rest}} > L_V^{\text{thresh}}$ (the J2003 catalog; Wolf, C. private communication). Using this catalog we calculated $j_{\lambda}^{\text{rest}}$ in an identical way to how it was calculated for the FIRES data. We divided the data into redshift bins of $\Delta z = 0.2$ and counted the light from all galaxies contained within each bin which had $L_V^{\text{rest}} > L_V^{\text{thresh}}$. The formal 68% confidence limits were calculated via bootstrapping. In addition, in Figure 2 we indicate the rms field-to-field variations between the three spatially distinct COMBO-17 fields. As also pointed out in Wolf et al. (2003), the field-to-field variations dominate the error in the COMBO-17 $j_{\lambda}^{\text{rest}}$ determinations.

Bell et al. (2003b) point out that uncertainties in the absolute calibration and relative calibration of the SDSS and Johnson zeropoints can lead to $\lesssim 10\%$ errors in the derived rest-frame magnitudes and colors of galaxies. To account for this, we add a 10% error in quadrature with the formal errors for both the COMBO-17 and SDSS luminosity densities. These are the errors presented in Table 1 and Figure 2.

In Figure 2a we also plot the j_V^{rest} value of luminous LBGs determined by integrating the luminosity function of Shapley et al. (2001) to L_V^{thresh} . A direct comparison between our sample and theirs is not entirely straightforward because the LBGs represent a specific class of non-obscured, star forming galaxies at high redshift, selected by their rest-frame far UV light. Nonetheless, our $j_{\lambda}^{\text{rest}}$ determination at $z = 2.8$ is slightly higher than their determination at $z = 3$, indicating either that the HDF-S is overdense with respect to the area surveyed by Shapley et al. or that we may have galaxies in our sample which are not present in the

ground-based LBG sample.

D03 have also measured the luminosity density in the rest-frame B -band but, because they do not give their luminosity function parameters except for their lowest redshift bin, it is not possible to overplot their luminosity density integrated down to our L_V^{rest} limit.

4.1.1 The Evolution of j_λ^{rest}

We find progressively stronger luminosity evolution from the V to the U -band: whereas the evolution is quite weak in V , it is very strong in U . The j_λ^{rest} in our highest redshift bin is a factor of 1.9 ± 0.4 , 2.9 ± 0.6 , and 4.9 ± 1.0 higher than the $z = 0.1$ value in V , B , and U respectively. To address the effect of cosmic variance on the measured evolution in j_λ^{rest} we rely on the clustering analysis developed for our sample in Daddi et al. (2003). Using the correlation length estimated at $2 < z < 4$, $r_o = 5.5 h_{100}^{-1}$ Mpc, we calculated the expected 1-sigma fluctuations in the number density of objects in our two highest redshift bins. Because for our high- z samples the poissonian errors are almost identical to the bootstrap errors, we can use the errors in the number density as a good proxy to the errors in j_λ^{rest} . The inclusion of the effects of clustering would increase the bootstrap errors on the luminosity density by a factor of, at most, 1.75 downwards and 2.8 upwards. This implies that the inferred evolution is still robust even in the face of the measured clustering. The COMBO-17 data appear to have a slightly steeper rise towards higher redshift than our data, however there are two effects to remember at this point. First, our lowest redshift point averages over all redshifts $z < 1.6$, in which case we are in reasonably good agreement with what one would predict from the average of the SDSS and COMBO-17 data. Second, our data may simply have an offset in density with respect to the local measurements. Such an offset affects the values of j_λ^{rest} , but as we will show in §4.2, it does not strongly affect the global color estimates. Nonetheless, given the general increase with j_λ^{rest} towards higher redshifts, we fit the changing j_λ^{rest} with a power law of the form $j_\lambda^{\text{rest}}(z) = j_\lambda^{\text{rest}}(0) * (1+z)^\beta$. These curves are overplotted in Figure 2 and the best fit parameters in sets of $(j_\lambda^{\text{rest}}(0), \beta)$ are $(5.96 \times 10^7, 1.41)$, $(6.84 \times 10^7, 0.93)$, and $(8.42 \times 10^7, 0.52)$ in the U , B , and V bands respectively, where $j_\lambda^{\text{rest}}(0)$ has units of $h_{70} L_\odot \text{Mpc}^{-3}$. At the same time, it is important to remember that our power law fit is likely an oversimplification of the true evolution in j_λ^{rest} .

The increase in j_λ^{rest} with decreasing cosmic time can be modeled as a simple brightening of L_* . Performing a test similar to that performed in R01, we determine the increase in $L_{*,V}$ with respect to $L_{*,V}^{\text{local}}$ needed to match the observed increase in j_λ^{rest} from $z = 0.1$ to $2.41 < z < 3.2$, assuming the SDSS Schechter function parameters. To convert between the Schechter function parameters in the SDSS bands and those in the Bessell (1990) filters we transformed the $L_{*,0.1r}^{\text{SDSS}}$ values to the Bessell V filter using the $(V - 0.1r)$ color, where the color was derived from the total luminosity densities in the indicated bands (as given in B03). We then ap-

Table 1 — Rest-Frame Optical Luminosity Density and Integrated Color

z	$\log j_{\text{U}}^{\text{rest}}$ [$h_{70} \text{ L}_{\odot,U} \text{Mpc}^{-3}$]	$\log j_{\text{B}}^{\text{rest}}$ [$h_{70} \text{ L}_{\odot,B} \text{Mpc}^{-3}$]	$\log j_{\text{V}}^{\text{rest}}$ [$h_{70} \text{ L}_{\odot,V} \text{Mpc}^{-3}$]	$(U - B)_{\text{rest}}$	$(B - V)_{\text{rest}}$
0.10 ± 0.10^a	$7.89^{+0.04}_{-0.05}$	$7.87^{+0.04}_{-0.05}$	$7.91^{+0.04}_{-0.05}$	$0.14^{+0.02}_{-0.02}$	$0.75^{+0.02}_{-0.02}$
0.30 ± 0.10^b	$7.84^{+0.05}_{-0.05}$	$7.85^{+0.05}_{-0.05}$	$7.93^{+0.05}_{-0.05}$	$0.21^{+0.02}_{-0.02}$	$0.84^{+0.01}_{-0.01}$
0.50 ± 0.10^b	$8.01^{+0.04}_{-0.05}$	$7.99^{+0.04}_{-0.05}$	$8.01^{+0.04}_{-0.05}$	$0.16^{+0.01}_{-0.01}$	$0.69^{+0.005}_{-0.01}$
0.70 ± 0.10^b	$8.18^{+0.04}_{-0.05}$	$8.13^{+0.04}_{-0.05}$	$8.12^{+0.04}_{-0.05}$	$0.06^{+0.01}_{-0.01}$	$0.64^{+0.01}_{-0.01}$
0.90 ± 0.10^b	$8.22^{+0.04}_{-0.05}$	$8.13^{+0.04}_{-0.05}$	$8.09^{+0.04}_{-0.05}$	$-0.04^{+0.004}_{-0.01}$	$0.55^{+0.005}_{-0.01}$
$1.12^{+0.48}_{-1.12}^c$	$8.11^{+0.08}_{-0.09}$	$8.02^{+0.08}_{-0.08}$	$8.00^{+0.08}_{-0.09}$	$-0.04^{+0.03}_{-0.03}$	$0.61^{+0.02}_{-0.02}$
$2.01^{+0.40}_{-0.41}^c$	$8.21^{+0.08}_{-0.10}$	$8.00^{+0.08}_{-0.10}$	$7.89^{+0.08}_{-0.09}$	$-0.34^{+0.04}_{-0.03}$	$0.38^{+0.04}_{-0.04}$
$2.80^{+0.40}_{-0.39}^c$	$8.58^{+0.07}_{-0.08}$	$8.32^{+0.07}_{-0.08}$	$8.18^{+0.07}_{-0.08}$	$-0.44^{+0.04}_{-0.03}$	$0.29^{+0.04}_{-0.03}$

$j_{\lambda}^{\text{rest}}$ and rest-frame colors calculated for galaxies with $\text{L}_{\text{V}}^{\text{rest}} > 1.4 \times 10^{10} h_{70}^{-2} \text{ L}_{\odot,V}$.

^a SDSS

^b COMBO-17

^c FIRES

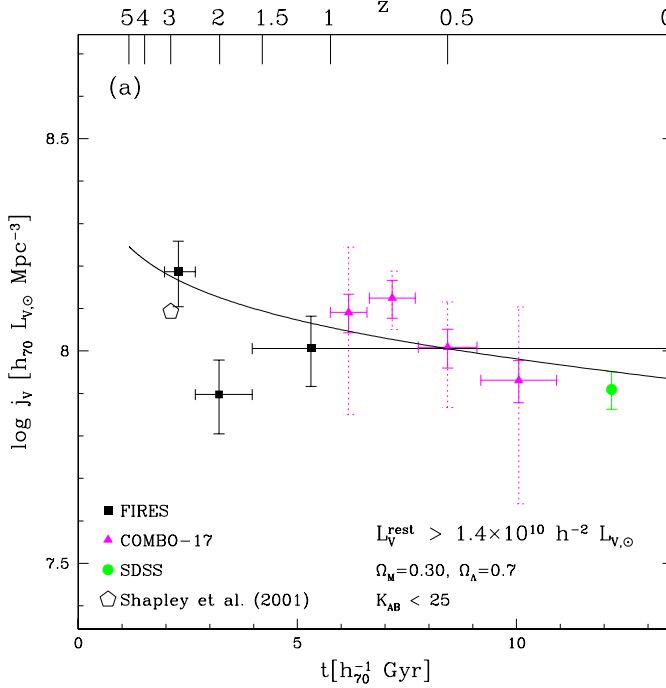
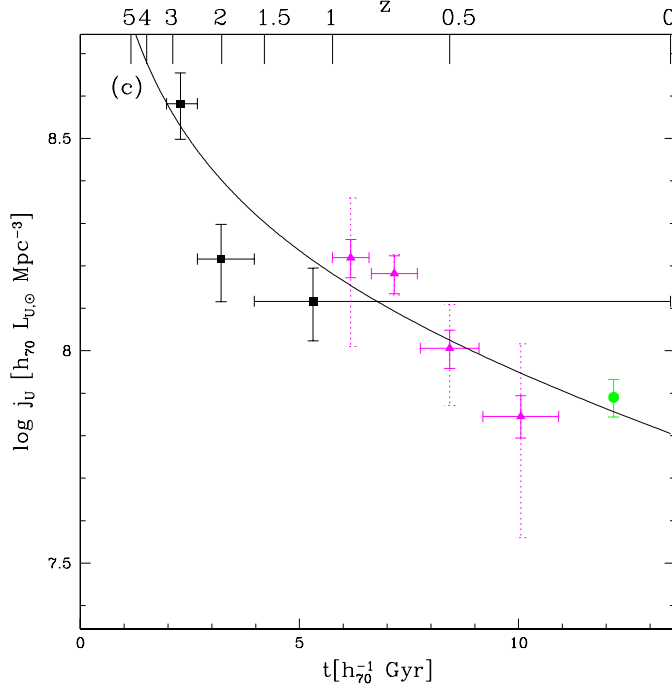
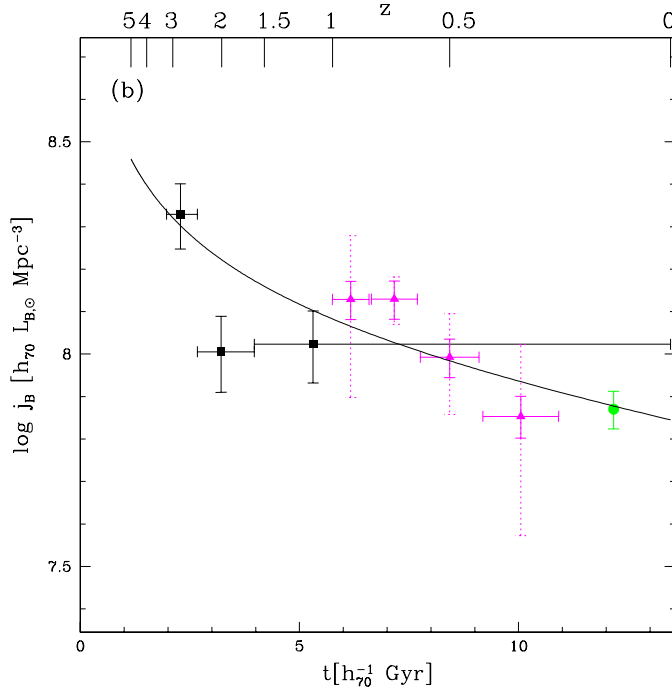


Figure 2 — The rest-frame optical luminosity density vs. cosmic age and redshift from galaxies with $K_{s,AB}^{tot} < 25$ and $L_V^{rest} > L_V^{thresh}$. For comparison we plot j_λ^{rest} determinations from other surveys down to our L_λ^{rest} limits. The squares are those from our data, the triangles are from the Combo-17 survey (Wolf et al. 2003), the circle is that at $z = 0.1$ from the SDSS (B03), and the pentagon is that from Shapley et al. (2001). The dotted errorbars on the COMBO-17 data indicate the rms field-to-field variation derived from the three spatially distinct COMBO-17 fields. The solid line is a power law fit to the FIRES, COMBO-17, and SDSS data of the form $j_\lambda^{rest}(z) = j_\lambda^{rest}(0) * (1+z)^\beta$.

plied the appropriate AB to Vega correction tabulated in Bessell (1990). Because the difference in λ_{eff} is small between the two filters in each of these colors, the shifts between the systems are less than 5%. The luminosity density in the V -band at $2.41 < z < 3.2$ is $j_V^{rest} = 1.53 \pm 0.26 \times 10^8 h_{70} L_\odot \text{ Mpc}^{-3}$. Using the V -band Schechter function parameters for our cosmology, $\phi_*^{SDSS} = 5.11 \times 10^{-3} h_{70}^3 \text{ Mpc}^{-3}$, $\alpha^{SDSS} = -1.05$, and $L_{*,V}^{SDSS} = 2.53 \times 10^{10} h_{70}^{-2} L_\odot$, we can match the increase in j_V^{rest} if $L_{*,V}$ brightens by a factor of 1.7 out to $2.41 < z < 3.2$.

4.2 The Cosmic Color

Using our measures of j_λ^{rest} we estimated the cosmic rest-frame color of all the visible stars which lie in galaxies with $L_V^{rest} > 1.4 \times 10^{10} h_{70}^{-2} L_\odot$. We derived the mean cosmic $(U - B)_{rest}$ and $(B - V)_{rest}$ by using the j_λ^{rest} estimates from



the previous section with the appropriate magnitude zeropoints. The measured colors for the FIRES data, the COMBO-17 data, and the SDSS data are given in Table 1. Emission line corrected $(U - B)_{rest}$ colors may be calculated by applying Equation 1 to the values in the table. For the FIRES and COMBO-17 data, uncertainty estimates are derived from the same bootstrapping simulation used in § 4.1. In this case, however, the COMBO-17 and SDSS errorbars do not include an extra component from errors in the transformation to rest-frame luminosities, since these transformation errors may be correlated in a non-trivial way.

The bluing with increasing redshift which could have been inferred from Figure 2 is seen explicitly in Figure 3. The color change towards higher redshift occurs more smoothly than the evolution in j_{λ}^{rest} , with our FIRES data meshing nicely with the COMBO-17 data. It is immediately apparent that the rms field-to-field errors for the COMBO-17 data are much less than the observed trend in color, in contrast to Figure 2. This explicitly shows that the integrated color is much less sensitive than j_{λ}^{rest} to field-to-field variations, even when such variations may dominate the error in the luminosity density. The COMBO-17 data at $z \lesssim 0.4$ are also redder than the local SDSS data, possibly owing to the very small central apertures used to measure colors in the COMBO-17 survey. The colors in the COMBO-17 data were measured with the package MPIAPHOT using the peak surface brightness in images smoothed all to identical seeing ($1''.5$). Such small apertures were chosen to measure very precise colors, not to obtain global color estimates. Because of color gradients, these small apertures can overestimate the global colors in nearby well resolved galaxies, while providing more accurate global color estimates for the more distant objects. Following the estimates of this bias provided by Bell et al. (2003b), we increased the color errorbars on the blue side to 0.1 for the $z < 0.4$ COMBO-17 data. It is encouraging to see that the color evolution is roughly consistent with a rather simple model and that it is much smoother than the luminosity density evolution, which is more strongly affected by cosmic variance.

We interpreted the color evolution as being primarily driven by a decrease in the stellar age with increasing redshift. Applying the $(B - V)_{rest}$ dependent emission line corrections inferred from local samples (See §3.2.1), we see that the effect of the emission lines on the color is much less than the magnitude of the observed trend. We can also interpret this change in color as a change in mean cosmic M/L^* with redshift. In this picture, which is true for a variety of monotonic SFHs and extinctions, the points at high redshift have lower M/L^* than those at low redshift. At the same time, however, the evolution in j_V^{rest} with redshift is quite weak. Taken together this would imply that the stellar mass density ρ_* is also decreasing with increasing redshift. We will quantify this in §4.3.

To show how our mean cosmic $(U - B)_{rest}$ and $(B - V)_{rest}$ colors compare to those of morphologically normal nearby galaxies, we overplot them in Figure 4 on the locus of nearby galaxies from Larson & Tinsley (1978). The integrated

colors, at all redshifts, lie very close to the local track, which Larson & Tinsley (1978) demonstrated is easily reproduceable with simple monotonically declining SFHs and which is preserved in the presence of modest amounts of reddening, which moves galaxies roughly parallel to the locus. In fact, correcting our data for emission lines moved them even closer to the local track. While we have suggested that M/L^* decreases with decreasing color, if we wish to actually quantify the M/L^* evolution from our data we must first attempt to find a set of models which can match our observed colors and which we will later use to convert between the color and M/L_V^* . We overplot in Figure 4 two model tracks corresponding to an exponential SFH with $\tau = 6$ Gyr and with $E(B - V) = 0, 0.15$, and 0.35 (assuming a Calzetti et al. (2000) reddening law). These tracks were calculated using the 2000 version of the Bruzual A. & Charlot (1993) models and have $Z = Z_\odot$ and a Salpeter (1955) IMF with a mass range of $0.1\text{--}120M_\odot$. Other exponentially declining models and even a constant star forming track all yield similar colors to the $\tau = 6$ Gyr track. The measured cosmic colors at $z < 1.6$ are fairly well approximated either of the reddening models. At $z > 1.6$, however, only the $E(B - V) = 0.35$ track can reproduce the data. This high extinction is in contrast to the results of P01 and Shapley et al. (2001) who found a mean reddening for LBGs of $E(B - V) \sim 0.15$. SY98 and Thompson, Weymann, & Storrie-Lombardi (2001), however, measured extinctions on this order for galaxies in the HDF-N, although the mean extinction from Thompson, Weymann, & Storrie-Lombardi (2001) was lower at $z > 2$. The amount of reddening in our sample is one of the largest uncertainty in deriving the M/L^* values, nonetheless, our choice of a high extinction is the only allowable possibility given the integrated colors of our high redshift data.

Although this figure demonstrates that the measured colors can be matched, at some age, by this simple $E(B - V) = 0.35$ model, we must nevertheless investigate whether the evolution of our model colors are also compatible with the evolution in the measured colors. This is shown by the track in Figure 3. We have tried different combinations of τ , $E(B - V)$, and z_{start} , but have not been able to find a model which fits the data well at all redshifts. The parameterized $SFR(z)$ curve of Cole et al. (2001) also provided a poor fit to the data. Given the large range of possible parameters, our data may not be sufficient to well constrain the SFH.

4.3 Estimating M/L_V^* and The Stellar Mass Density

In this subsection we describe the use of our mean cosmic color estimates to derive the mean cosmic M/L_V^* and the evolution in ρ_* .

The main strength of considering the luminosity density and integrated colors of the galaxy population, as opposed to those of individual galaxies, lies in the simple and robust ways in which these global values can be modeled. When attempting to derive the SFHs and stellar masses of individual high-redshift galaxies, the state-of-the-art models for the broad-band colors only consider stellar populations with at most two separate components (SY98; P01; Shapley et al. 2001).

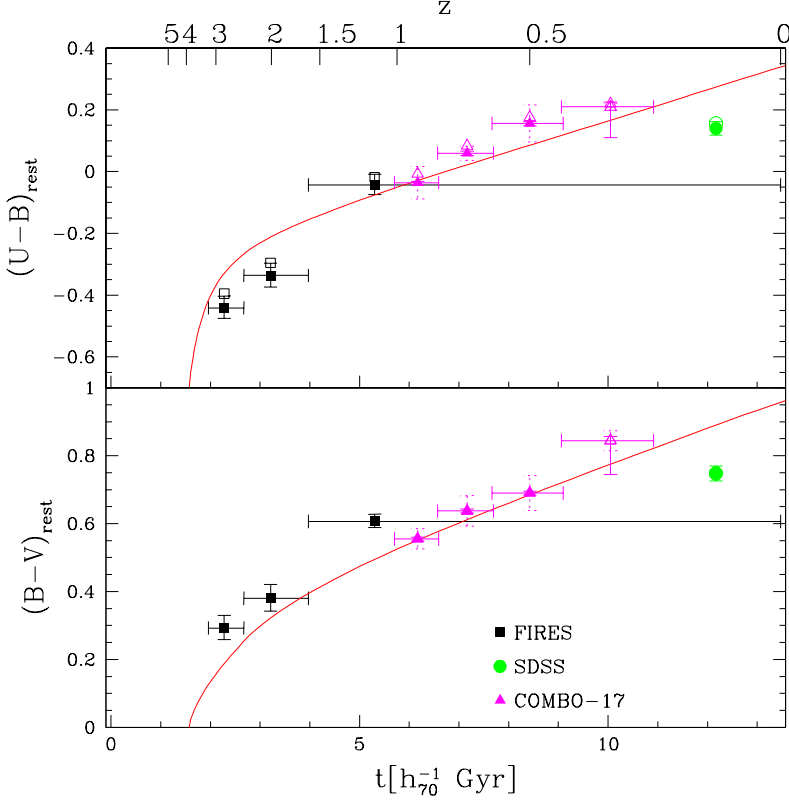


Figure 3 — The evolution of the cosmic color plotted against redshift and cosmic time for our data in addition to data from other $z \lesssim 1$ surveys. The squares are those from our data, the triangles are from the Combo-17 survey (Wolf et al. 2003), and the circle is that at $z = 0.1$ from the SDSS (B03). The open symbols indicate the empirical emission line correction to the integrated colors derived using the spectroscopic and photometric data from the NFGS (Jansen et al. 2000b). The dotted errorbars on the COMBO-17 data indicate the field-to-field variation. Note that the integrated rest-frame color is much more stable than $j_{\lambda}^{\text{rest}}$ against field-to-field variations. The COMBO-17 data point at $z = 0.3$ has been given a color errorbar of 0.1 in the blueward direction and an open symbol to reflect the possible systematic biases resulting from their very small central apertures. We also overplot a model with an exponentially declining SFH with $\tau = 6 \text{ Gyr}$, $E(B - V) = 0.35$, and $z_{\text{start}} = 4.0$ assuming a Calzetti et al. (2000) extinction law.

Using their stellar population synthesis modeling, Shapley et al. (2001) proposes a model in which LBGs likely have smooth SFHs. On the other hand, SY98 concluded that they may only be seeing the most recent episode of star formation and that LBGs may indeed have bursting SFHs. This same idea was supported by P01 and Ferguson, Dickinson, & Papovich (2002) using much deeper NIR data. When using similar simple SFHs to model the cosmic average of the galaxy population,

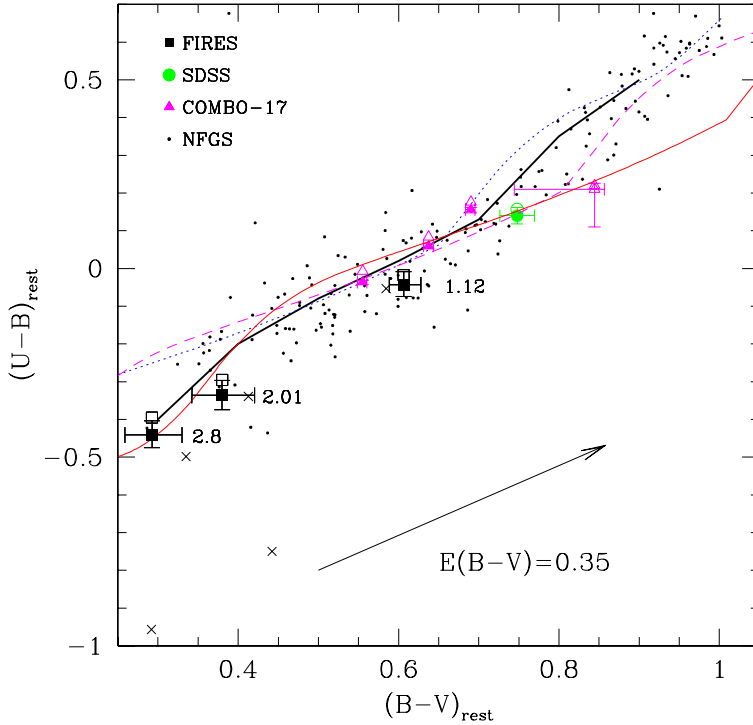


Figure 4 — The $(U - B)_{rest}$ vs. $(B - V)_{rest}$ at $z = 1.12, 2.01$, and 2.8 of all the relatively unobscured stars in galaxies with $L_V^{rest} > 1.4 \times 10^{10} h_{70}^{-2} L_\odot$. The thick solid black line is the local relation derived by Larson & Tinsley (1978) from nearby morphologically normal galaxies. The large symbols are identical to those in Figure 3. For clarity we do not plot the field-to-field errorbars for the COMBO-17 data. The small solid points are the colors of nearby galaxies from the NFGS (Jansen et al. 2000a), which have been corrected for emission lines. The small crosses are the NFGS galaxies which harbor AGN. The thin tracks correspond to an exponentially declining SFH with a timescale of 6 Gyr. The tracks were created using a Salpeter (1955) IMF and the 2000 version of the Bruzual A. & Charlot (1993) models. The dotted track has no extinction, the dashed track has been reddened by $E(B - V) = 0.15$, and the thin solid track has been reddened by $E(B - V) = 0.35$, using the Calzetti et al. (2000) extinction law. The black arrow indicates the reddening vector applied to the solid model track. The emission line corrected data lie very close the track defined by observations of local galaxies and the agreement with the models demonstrates that simple SFHs can be used to reproduce the integrated colors from massive galaxies at all redshifts.

a more self-consistent approach is possible. While individual galaxies may, and probably do, have complex SFHs, the mean SFH of all galaxies is much smoother than that of individual ones.

Encouraged by the general agreement between the measured colors and the simple models, we attempted to use this model to constrain the stellar mass-to-

light ratio M/L_V^* in the rest-frame V -band, by taking advantage of the relation between color and $\log_{10}M/L^*$ found by Bell & de Jong (2001). For monotonic SFHs, the scatter of this relation remains small in the presence of modest variations in the reddening and metallicity because these effects run roughly parallel to the mean relation. Using the $\tau = 6$ Gyr exponentially declining model, we plot in Figure 5 the relation between $(U - V)_{rest}$ and M/L_V^* for the $E(B - V) = 0, 0.15$, and 0.35 models. It is seen that dust extinction moves objects roughly parallel to the model tracks, reddening their colors, but making them dimmer as well and hence increasing M/L_V^* . Nonetheless, extinction uncertainties are a major contributor to our errors in the determination of M/L_V^* . We chose to derive M/L_V^* from the $(U - V)_{rest}$ color instead of from the $(B - V)_{rest}$ color because at blue colors, where our high redshift points lie, $(B - V)_{rest}$ derived M/L_V^* values are much more sensitive to the exact value of the extinction. This behavior likely stems from the fact that the $(U - V)_{rest}$ color spans the Balmer/4000Å break and hence is more age sensitive than $(B - V)_{rest}$. At the same time, while $(U - B)_{rest}$ colors are even less sensitive to extinction than $(U - V)_{rest}$, they are more susceptible to the effects of bursts.

We constructed our relation using a Salpeter (1955) IMF⁵. The adoption of a different IMF would simply change the zeropoint of this curve, leaving the relative M/L^* as a function of color, however, unchanged. As discussed in §4.2, this model does not fit the redshift evolution of the cosmic color very well. Nonetheless, the impact on our M/L_V^* estimates should not be very large, since most smooth SFHs occupy very similar positions in the M/L_V^* vs. $U - V$ plane.

This relation breaks down in the presence of more complex SFHs. We demonstrate this in Figure 6 where we plot the $\tau = 6$ Gyr track and a second track whose SFH is comprised of a 50 Myr burst at $t = 0$, a gap of 2 Gyr, and a constant SFR rate for 1 Gyr thereafter, where 50% of the mass is formed in the burst. It is obvious from this figure that using a smooth model will cause errors in the M/L_V^* estimate if the galaxy has a SFR which has an early peak in the SFH. At blue colors, such a early burst of SFR will cause an underestimate of M/L_V^* , a result similar to that of P01 and D03. At red colors, however, M/L_V^* would be overestimated with the exact systematic offset as a function of color depending strongly on the detailed SFH, i.e. the fraction of mass formed in the burst, the length of the gap, and the final age of the stellar population.

The models show that the method may over- or underestimate the true stellar mass-to-light ratio if the galaxies have complex SFHs. It is important to quantify the errors on the global M/L_V^* based on the mean $(U - V)_{rest}$ color and how these errors compare to those when determining the global M/L_V^* value from individual M/L_V^* estimates. To make this comparison we constructed a model whose SFH consist of a set of 10 Myr duration bursts separated by 90 Myr gaps. We drew galaxies at random from this model by randomly varying the phase and

⁵We do not attempt to model an evolving IMF although evidence for a top-heavy IMF at high redshifts has been presented by Ferguson, Dickinson, & Papovich (2002)

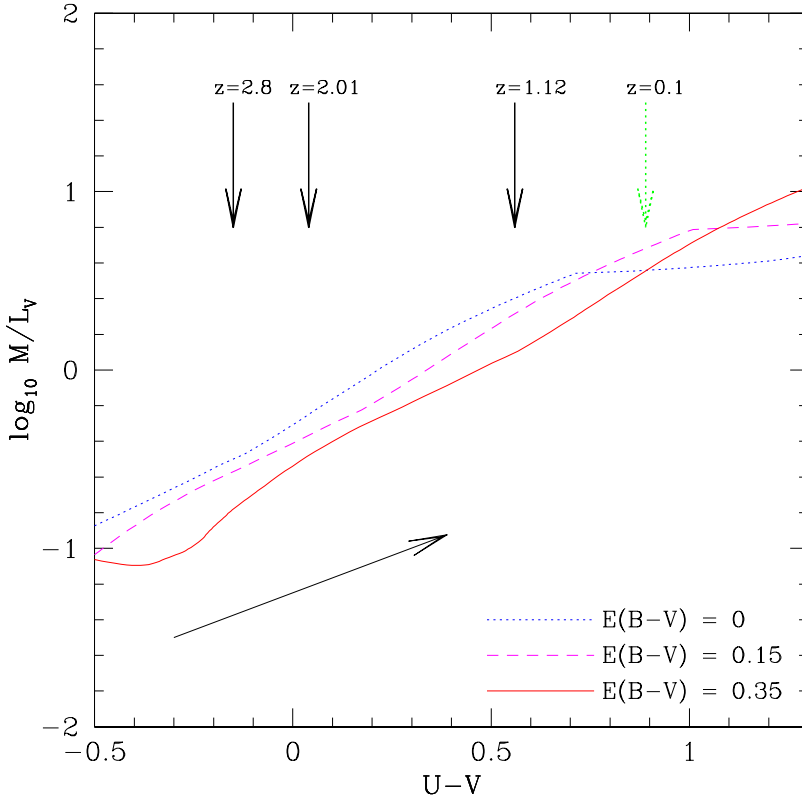


Figure 5 — The relation between $(U - V)$ and M/L_V^* for a model track with an exponential timescale of 6 Gyr. The dotted line is for a model with $E(B - V) = 0$, the dashed line for a model with $E(B - V) = 0.15$, and the solid line is for a model reddened by $E(B - V) = 0.35$ (using a Calzetti extinction law), which we adopt for our M/L_V^* conversions. The vertical solid arrows indicate the colors of the three FIRES data points, the vertical dotted arrow indicates the color of the SDSS data, and the diagonal solid arrow indicates the vector used to redden the $E(B - V) = 0$ model to $E(B - V) = 0.35$. The labels above the vertical arrows correspond to the redshifts of the FIRES and SDSS data.

age of the burst sequence, where the maximum age was 4 Gyr. Next we estimated the total mass-to-light ratios of this sample by two different methods; first we determined the M/L_V^* for the galaxies individually assuming the simple relation between color and mass-to-light ratio, and we took the luminosity weighted mean of the individual estimates to obtain the total M/L_V^* . This point is indicated by a large square in Figure 7 and overestimates the total M/L_V^* by $\sim 70\%$. Next we first add the light of all the galaxies in both U and V , then use the simple relation between color and M/L_V^* to convert the integrated $(U - V)_{rest}$ into a mass-to-light ratio. This method overestimates M/L_V^* by much less, $\sim 35\%$. This

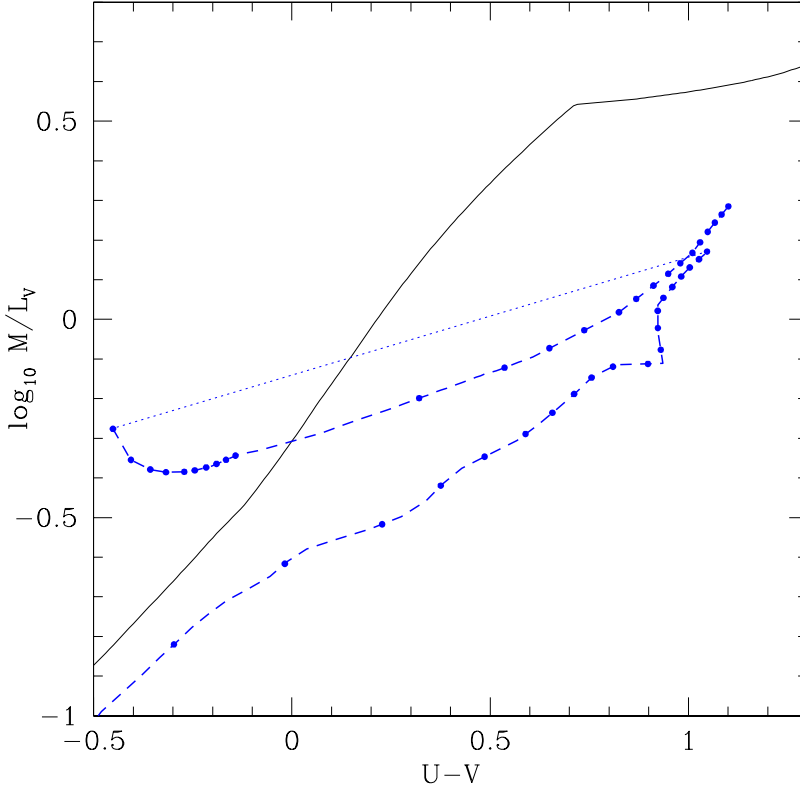


Figure 6 — The effect of an early burst of star formation on the relation between $(U - V)$ and M/L_V^* . The relation between $(U - V)$ and M/L_V^* for a model track with an exponential timescale of 6 Gyr is shown by the solid line. We also show a track for a SFH which includes a 50 Myr burst at $t = 0$ followed by a gap of 2 Gyr and then a constant SFR rate for 1 Gyr thereafter, where the fraction of mass formed in the burst is 0.5. The track continues for a total time of 4.5 Gyr. The dots are placed at 100 Myr intervals and the dotted section of the line indicates the very rapid transition in color caused by the onset of the second period of star formation. Both tracks have the same extinction.

comparison shows clearly that it is best to estimate the mass using the integrated light. This is not unexpected; the star formation history of the universe as a whole is more regular than the star formation history of individual galaxies. If enough galaxies are averaged, the mean star formation history is naturally fairly smooth.

Using the relationship between color and M/L_V^* we convert our $(U - V)_{rest}$ and j_V^{rest} measurements to stellar mass density estimates ρ_* . The resulting ρ_* values are plotted vs. cosmic time in Figure 8. We have included points for the SDSS survey created in an analogous way to those from this work, i.e. using the M/L_V^* derived from the rest-frame color and multiplying it by j_V^{SDSS} for all galaxies with

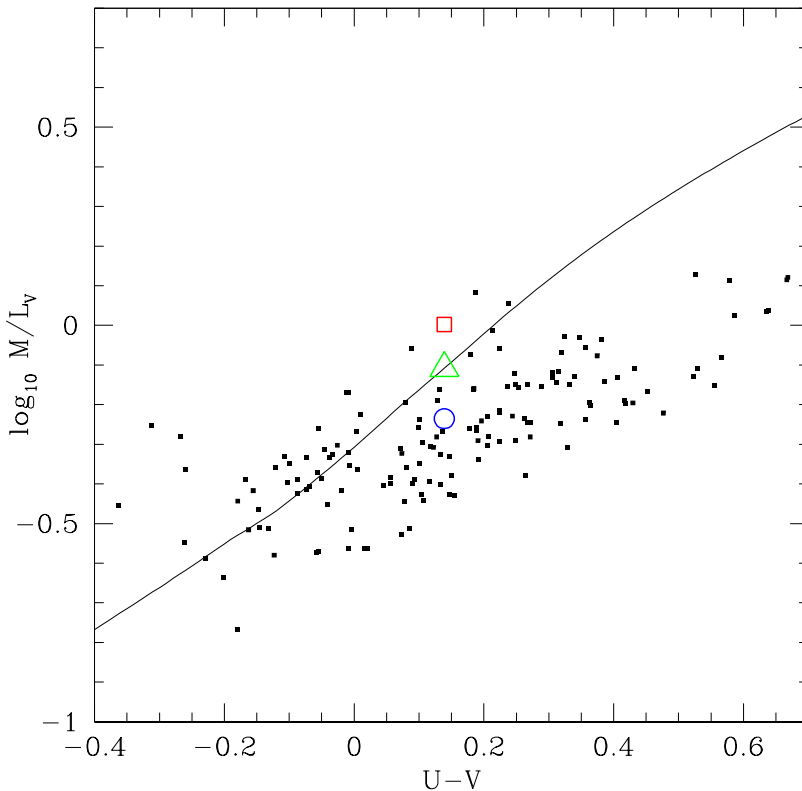


Figure 7 — A comparison of different measures of the global M/L_V^* for a mock catalog of galaxies with bursting SFHs. The solid line represents the relation between $(U - V)$ and M/L_V^* for a model track with an exponential timescale of 6 Gyr. The black dots show the true M/L_V 's of the model starbursting galaxies, as described in the text; the open circle shows the true luminosity weighted $M_{\text{tot}}/L_{\text{tot}}$ of the mock galaxies. The square shows the luminosity weighted M/L_V^* derived by applying the simple model to the individual galaxies - in this case, the mean $M_{\text{tot}}/L_{\text{tot}}$ is overestimated by 70%. The triangle is the $M_{\text{tot}}/L_{\text{tot}}$ derived from the luminosity weighted mean color (or $(U - V)_{\text{tot}}$) of the model galaxies. It overestimates $M_{\text{tot}}/L_{\text{tot}}$ by only 35%.

$L_V^{\text{rest}} > L_V^{\text{thresh}}$. The ρ_* estimates are listed in Table 2. We have derived the statistical errorbars on the ρ_* estimates by creating a Monte-Carlo simulation where we allowed j_V^{rest} and $(U - V)_{\text{rest}}$ (and hence M/L_V^*) to vary within their errorbars. We then took the resulting distribution of ρ_* values and determined the 68% confidence limits. As an estimate of our systematic uncertainties corresponding to the method we also determined M/L_V^* from the $(U - B)_{\text{rest}}$ and $(B - V)_{\text{rest}}$ data using an identical relation as for the $(U - V)_{\text{rest}}$ to M/L_V^* conversion. The $(U - B)_{\text{rest}}$ derived M/L_V^* values were different from the $(U - V)_{\text{rest}}$ values

Table 2 — M/L_V^* and Stellar Mass Density Estimates

z	$\log M/L_V^*$ [$\frac{M_\odot}{L_\odot}$]	$\log \rho_*$ [$h_{70} M_\odot \text{Mpc}^{-3}$]
0.1 ± 0.1^a	$0.54^{+0.03}_{-0.03}$	$8.49^{+0.04}_{-0.05}$
$1.12^{+0.48}_{-1.12}^b$	$0.13^{+0.07}_{-0.06}$	$8.14^{+0.11}_{-0.10}$
$2.01^{+0.40}_{-0.41}^b$	$-0.42^{+0.09}_{-0.10}$	$7.48^{+0.12}_{-0.16}$
$2.80^{+0.40}_{-0.39}^b$	$-0.70^{+0.11}_{-0.12}$	$7.49^{+0.12}_{-0.14}$

^a SDSS

^b FIRES

by a factor of 1.02, 0.80, 0.95, and 1.12 for the $z=0.1$, 1.12, 2.01, and 2.8 redshift bins respectively. Likewise the $(B - V)_{rest}$ determined M/L_V^* values changed by a factor of 0.99, 1.11, 1.15, and 0.80 with respect to the $(U - V)_{rest}$ values. While the $(U - V)_{rest}$ values are very similar to those derived from the other colors, the $(U - V)_{rest}$ color is less susceptible to dust uncertainties than the $(B - V)_{rest}$ data and less susceptible to the effects of bursts than the $(U - B)_{rest}$ data.

The derived mass density rises monotonically by a factor of ~ 10 all the way to $z \sim 0.1$, with our low redshift point meshing nicely with the local SDSS point.

5 Discussion

5.1 Comparison with other Work

Figure 8 shows a consistent picture of the build-up of stellar mass, both for the luminous galaxies and the total galaxy population. It is remarkable that the results from different authors appear to agree well given that the methods to derive the densities were different and that the fields are very small.

We compared our results to the total mass estimates of other authors in Figure 8. In doing this we must remember, because of our $L_\lambda^{\text{thresh}}$ cut, that we are missing significant amounts of light, and hence, mass. Assuming the SDSS luminosity function parameters, we lose 46% of the light at $z = 0$. At $z = 2.8$, however, we inferred a brightening of $L_{*,V}$ by a factor of 1.7, implying that we go further down the luminosity function at high redshift, sampling a larger fraction of the total starlight. If we apply this brightening to the SDSS $L_{*,V}$ we miss 30% of the light below our luminosity threshold at $z = 2.8$. Hence, the fraction of the total starlight contained in our sample is rather stable as a function of redshift. To graphically compare our data to other authors we have scaled the two different axes in Figure 8 so that our derivation of the SDSS ρ_* is at the same height of the total ρ_* estimate of Cole et al. (2001). At $z < 1$ we compared our mass estimates to those of Brinchmann & Ellis (2000). Following D03, we have corrected their

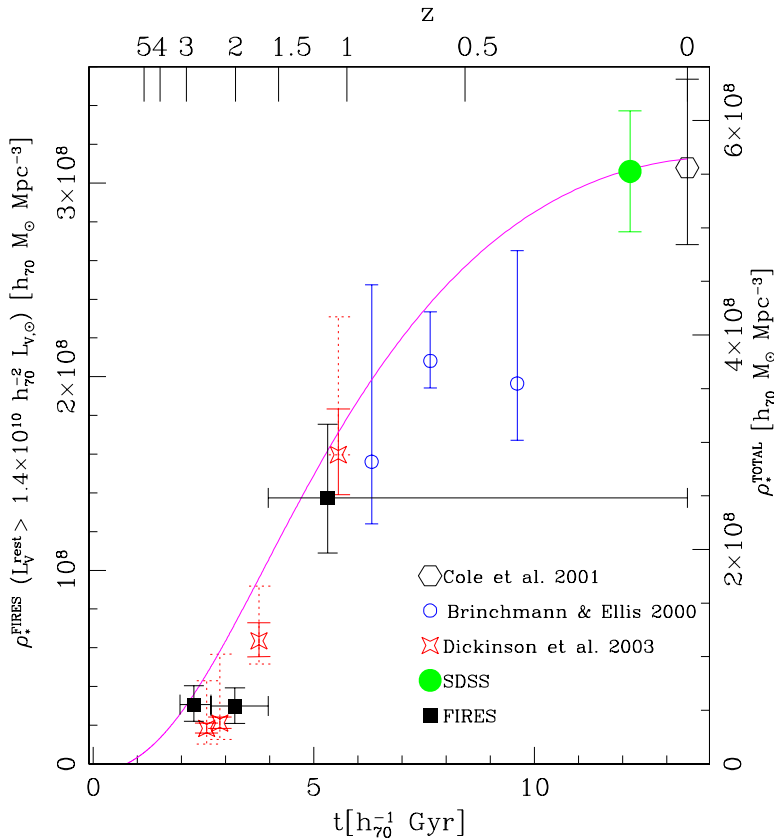


Figure 8 — The build-up of the stellar mass density as a function of redshift. The solid points are for galaxies with $L_V^{rest} > 1.4 \times 10^{10} h_{70}^{-2} L_\odot$ and were derived by applying the $E(B-V) = 0.35$ relation in Figure 5 to the $(U - V)_{rest}$ colors and j_λ^{rest} measurements from the FIREs (solid squares) and SDSS data (solid circle). The y-axis scale on the left side corresponds to the ρ_* values for these points. The hollow points show the total stellar mass density measurements from the one-component models in the HDF-N (D03; hollow stars; calculated assuming solar metallicity), the CFRS (Brinchmann & Ellis 2000; hollow circles), and the 2dFGRS + 2MASS (Cole et al. 2001; hollow hexagon). The dotted errorbars on the D03 points reflect the systematic mass uncertainties resulting from metallicity and SFH changes. The y-axis scale on the right hand side corresponds to the ρ_* estimates for these points. The relative scaling of the two axes was adjusted so that our SDSS ρ_* estimate was at the same height as the total ρ_* estimate of Cole et al. The solid curve is an integral of the SFR(z) from Cole et al. (2001) which has been fit to extinction corrected data at $z \lesssim 4$.

published points to total masses by correcting them upwards by 20% to account for their mass incompleteness. The fraction of the total stars formed at $z < 1$ agrees well between our data and that of SDSS and Brinchmann & Ellis. At $z > 0.5$, we compared our results to those of D03. D03 calculated the total mass density, using the integrated luminosity density in the rest-frame B-band coupled with M/L

measurements of individual galaxies. The fractions of the total stars formed in our sample (60%, 13%, and 9% at $z=1.12$, 2.01, and 2.8) are almost twice as high as those of D03. The results, however, are consistent within the errors.

We explored whether field-to-field variations may play a role in the discrepancy between the two datasets. D03 studied the HDF-N, which has far fewer "red" galaxies than HDF-S (e.g., Labbé et al. 2003, Franx et al, 2003). If we omit the $J-K$ selected galaxies found by Franx et al. (2003) in the HDF-S, the formal M/L_V^* decreases to 45% and 43% of the total values and the mass density decreases to 57% and 56% of the total values in the $z = 2.01$ and $z = 2.8$ bins respectively, bringing our data into better agreement with that from D03. This reinforces the earlier suggestion by Franx et al. (2003) that the $J-K$ selected galaxies contribute significantly to the stellar mass budget.

The errors in both determinations are dominated by systematic uncertainties, although our method should be less sensitive to bursts than that of D03 as it uses the light integrated over the galaxy population.

We note that Fontana et al. (2003) have also measured the stellar mass density in the HDF-S using a catalog derived from data in common with our own. They find a similar, although slightly smaller evolution in the stellar mass density, consistent with our result to within the uncertainties.

5.2 Comparison with $SFR(z)$

We can compare the derived stellar mass to the mass expected from determinations of the SFR as a function of redshift. We use the curve by Cole et al. (2001), who fitted the observed SFR as determined from various sources at $z \lesssim 4$. To obtain the curve in Figure 8 we integrated the $SFR(z)$ curve taking into account the time dependent stellar mass loss derived from the 2000 version of the Bruzual A. & Charlot (1993) population synthesis models.

We calculated a reduced χ^2 of 4.3 when comparing all the data to the model. If, however, we omit the Brinchmann & Ellis (2000) data, the reduced χ^2 decreases to 1.8, although the results at $z > 2$ lie systematically below the curve. This suggests that some systematic errors may play a role, or that the curve is not quite correct. The following errors can influence our mass density determinations:

- Dusty, evolved populations: it is assumed that the dust is mixed in a simple way with the stars, leading to a Calzetti extinction curve. If the dust is distributed differently, e.g., by having a very extincted underlying evolved population, or by having a larger R value, the current assumptions lead to a systematic underestimate of the mass. If an underlying, extincted evolved population exists, it would naturally explain the fact that the ages of the Lyman-break galaxies are much younger than expected (e.g., P01, Ferguson et al. 2002). There may also be galaxies which contribute significantly to the mass density but are so heavily extincted that they are undetected, even with our very deep K_s -band data. If such objects

are also actively forming stars, they may be detectable with deep submillimeter observations or with rest-frame NIR observations from SIRTf.

-Cosmic variance: the two fields which have been studied are very small. Although we use a consistent estimate of clustering from Daddi et al. (2003), red galaxies make up a large fraction of the mass density in our highest redshift bins. Since red galaxies have a very high measured clustering from $z \sim 1$ (e.g., Daddi et al. 2000, McCarthy et al. 2001) up to possibly $z \sim 3$ (Daddi et al. 2003), large uncertainties remain.

-Evolving Initial Mass Function: the light which we see is mostly coming from the most massive stars present, whereas the stellar mass is dominated by low mass stars. Changes in the IMF would immediately lead to different mass estimates but if the IMF everywhere is identical (as we assume), then the relative masses should be robust. If the IMF evolves with redshift, however, systematic errors in the mass estimate will occur.

-A steep galaxy mass function at high redshift: if much of the UV light which is used to measure the SFR at high redshifts comes from small galaxies which would fall below our rest-frame luminosity threshold then we may be missing significant amounts of stellar mass. Even the mass estimates of D03, which were obtained by integrating the luminosity function, are very sensitive to the faint end extrapolation in their highest redshift bin.

5.3 The Build-up of the Stellar Mass

The primary goal of measuring the stellar mass density is to determine how rapidly the universe assembled its stars. At $z \sim 2-3$, our results indicate that the universe only contained $\sim 10\%$ of the current stellar mass, regardless of whether we refer only to galaxies at $L_V^{rest} > 1.4 \times 10^{10} h_{70}^{-2} L_\odot$ or whether we use the total mass estimates of other authors. The galaxy population in the HDF-S was rich and diverse at $z > 2$, but even so it was far from finished in its build-up of stellar mass. By $z \sim 1$, however, the total mass density had increased to roughly half its local value, indicating that the epoch of $1 < z < 2$ was an important period in the stellar mass build-up of the universe.

A successful model of galaxy formation must not only explain our global results, but also reconcile them with the observed properties of individual galaxies at all redshifts. For example, a population of galaxies at $z \sim 1-1.5$ has been discovered (the so called extremely red objects or EROs), roughly half of which can be fit with formation redshifts higher than 2.4 (Cimatti et al. 2002) and nearly passive stellar evolution thereafter. Our results, which show that the universe contained only $\sim 10\%$ as many stars at $z \sim 2-3$ as today would seem to indicate that any population of galaxies which formed most of its mass at $z \gtrsim 2$ can *at most* contribute $\sim 10\%$ of the present day stellar mass density. At $z \sim 1-1.5$, where the EROs reside, the universe had assembled roughly half of its current stars. Therefore, this would imply that the old EROs contribute about $\sim 20\%$ of the

mass budget at their epoch. Likewise, it should be true that a large fraction of the stellar mass at low redshift should reside in objects with mass weighted stellar ages corresponding to a formation redshift of $1 < z < 2$. In support of this, Hogg et al. (2002) recently have shown that $\sim 40\%$ of the local luminosity density at $0.7\mu m$, and perhaps $\sim 50\%$ of the stellar mass comes from centrally concentrated, high surface brightness galaxies which have red colors. In agreement with the Hogg et al. (2002) results, Bell et al. (2003a) and Kauffmann et al. (2003) also found that $\sim 50 - 75\%$ of the local stellar mass density resides in early type galaxies. Hogg et al. (2002) suggest that their red galaxies would have been formed at $z \gtrsim 1$, fully consistent with our results for the rapid mass growth of the universe during this period.

6 Summary & Conclusions

In this paper we presented the globally averaged rest-frame optical properties of a K_s -band selected sample of galaxies with $z < 3.2$ in the HDF-S. Using our very deep $0.3 - 2.2\mu m$, seven band photometry taken as part of the FIRE Survey we estimated accurate photometric redshifts and rest-frame luminosities for all galaxies with $K_{s,AB}^{\text{tot}} < 25$ and used these luminosity estimates to measure the rest-frame optical luminosity density $j_{\lambda}^{\text{rest}}$, the globally averaged rest-frame optical color, and the stellar mass density for all galaxies at $z < 3.2$ with $L_V^{\text{rest}} > 1.4 \times 10^{10} h_{70}^{-2} L_{\odot}$. By selecting galaxies in the rest-frame V -band, we selected them in a way much less biased by star formation and dust than the traditional selection in the rest-frame UV and much closer to a selection by stellar mass.

We have shown that $j_{\lambda}^{\text{rest}}$ in all three bands rises out to $z \sim 3$ by factors of 4.9 ± 1.0 , 2.9 ± 0.6 , and 1.9 ± 0.4 in the U , B , and V -bands respectively. Modeling this increase in $j_{\lambda}^{\text{rest}}$ as an increase in L_* of the local luminosity function, we derive that L_* must have brightened by a factor of 1.7 in the rest-frame V -band.

Using our $j_{\lambda}^{\text{rest}}$ estimates we calculate the $(U - B)_{\text{rest}}$ and $(B - V)_{\text{rest}}$ colors of all the visible stars in galaxies with $L_V^{\text{rest}} > 1.4 \times 10^{10} h_{70}^{-2} L_{\odot}$. Using the COMBO-17 data we have shown that the mean color is much less sensitive to density fluctuations and field-to-field variations than either $j_{\lambda}^{\text{rest}}$ or ρ_* . Because of their stability, integrated color measurements are ideal for constraining galaxy evolution models. The luminosity weighted mean colors lie close to the locus of morphologically normal local galaxy colors defined by Larson & Tinsley (1978). The mean colors monotonically bluen with increasing redshift by 0.55 and 0.46 magnitudes in $(U - B)_{\text{rest}}$ and $(B - V)_{\text{rest}}$ respectively out to $z \sim 3$. We interpret this color change primarily as a change in the mean stellar age. The joint colors can be roughly matched by simple SFH models if modest amounts of reddening ($E(B - V) < 0.35$) are applied. In detail, the redshift dependence of $(U - B)_{\text{rest}}$ and $(B - V)_{\text{rest}}$ cannot be matched exactly by the simple models, assuming a constant reddening and constant metallicity. However, we show that the models can still be used, even in the face of these small disagreements, to robustly predict the stellar

mass-to-light ratios M/L_V^* of the integrated cosmic stellar population implied by our mean rest-frame colors. Variations in the metallicity does not strongly affect this relation and it holds for a variety of smooth SFHs. Even the IMF only affects the normalization of this relation, not its slope, assuming that the IMF everywhere is the same. The reddening, which moves objects roughly along this relation is, however, a large source of uncertainty. Using these M/L_V^* estimates coupled with our j_λ^{rest} measurements, we derive the stellar mass density ρ_* . These globally averaged estimates of the mass density are more reliable than those obtained from the mean of individual galaxies determined using smooth SFHs, primarily because the cosmic mean SFH is plausibly much better approximated as being smooth, whereas the SFHs of individual galaxies are almost definitely not.

The stellar mass density, ρ_* , increases monotonically with increasing cosmic time to come into good agreement with the other measured values at $z \lesssim 1$ with a factor of ~ 10 increase from $z \sim 3$ to the present day. Within the random uncertainties, our results agree well with those of Dickinson, Papovich, Ferguson, & Budavári (2003) in the HDF-N although our ρ_* estimates are systematically higher than in the HDF-N. Taken together, the HDF-N and HDF-S paint a picture in which only $\sim 5 - 15\%$ of the present day stellar mass was formed by $z \sim 2$. By $z \sim 1$, however, the stellar mass density had increased to $\sim 50\%$ of its present value, implying that a large fraction of the stellar mass in the universe today was assembled at $1 < z < 2$. Our ρ_* estimates slightly underpredict the stellar mass density derived from the integral of the SFR(z) curve at $z > 2$. A resolution of the small apparent discrepancy between different fields, and between the predictions from optical observations will in part require deeper NIR data, to probe further down the mass function, and wider fields with multiple pointings to control the effects of cosmic variance. In addition, large amounts of follow-up optical/NIR spectroscopy are required to help control systematic effects in the z_{phot} estimates. The 25 square arcminute MS1054-03 data taken as part of FIRES and the ACS/ISAAC GOODS observations of the CDF-S region will be very helpful for such studies. Observations with SIRTf will also improve the situation by accessing the rest-frame NIR, where obscuration by dust becomes much less important. Finally, systematics in the M/L^* estimates may exist because of a lack of constraint on the faint end slope of the stellar IMF.

We still have to reconcile global measurements of the galaxy population with what we know about the ages and SFHs of individual galaxies. Our globally determined quantities are quite stable and may serve as robust constraints on theoretical models, which must correctly model the global build-up of stellar mass in addition to matching the detailed properties of the galaxy population.

Acknowledgments

GR would like to thank Jarle Brinchmann and Frank van den Bosch for useful discussions in the process of writing this paper, Christian Wolf for providing ad-

ditional COMBO-17 data products, and Eric Bell and Marcin Sawicki for giving comments on an earlier version of the paper. GR would also like to acknowledge the generous travel support of the Lorentz center and the Leids Kerkhoven-Bosscha Fonds, and the financial support of Sonderforschungsbereich 375.

Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

References

- Arnouts, S. et al. 2002, MNRAS, 329, 355
 Baldry, I. K. et al. 2002, ApJ, 569, 582
 Barger, A. J., Cowie, L. L., & Richards, E. A. 2000, AJ, 119, 2092
 Bell, E. F. & de Jong, R. S. 2001, ApJ, 550, 212
 Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJ, 585, L117
 Bell, E. F., Wolf, C., Meisenheimer, K., Rix, H.-W., Borch, A., Dye, S., Kleinheinrich, K., McIntosh, D. H. 2003, submitted to ApJ, *astro-ph/0303394*
 Benítez, N., Broadhurst, T., Bouwens, R., Silk, J., & Rosati, P. 1999, ApJ, 515, L65
 Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
 Bessell, M. S. & Brett, J. M. 1988, PASP, 100, 1134
 Bessell, M. S. 1990, PASP, 102, 1181
 Blanton, M. R. et al. 2001, AJ, 121, 2358
 Blanton, M. R. et al. 2003a, ApJ, in press (*astro-ph/0209479*)
 Blanton, M. R. et al. 2003b, AJ, 125, 2348
 Blanton, M. R. et al. 2003c, ApJ, 592, 819
 Bolzonella, M., Pelló, R., & Maccagni, D. 2002, A&A, 395, 443
 Brinchmann, J. & Ellis, R. S. 2000, ApJ, 536, L77
 Bruzual A., G. & Charlot, S. 1993, ApJ, 405, 538
 Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., & Storchi-Bergmann, T. 2000, ApJ, 533, 682
 Casertano, S. et al. 2000, AJ, 120, 2747
 Charlot, S. & Longhetti, M. 2001, MNRAS, 323, 887
 Chen, H. et al. 2003, ApJ, 586, 745
 Cimatti, A. et al. 2002, A&A, 381, L68
 Cohen, J. G., Hogg, D. W., Blandford, R., Cowie, L. L., Hu, E., Songaila, A., Shopbell, P., & Richberg, K. 2000, ApJ, 538, 29

- Cole, S. et al. 2001, MNRAS, 326, 255
- Coleman, G. D., Wu, C. -C., & Weedman, D. W. 1980, ApJS, 43, 393
- Colless, M. et al. 2001, MNRAS, 328, 1039
- Cowie, L. L., Songaila, A., & Barger, A. J. 1999, AJ, 118, 603
- Cox, A. N. 2000, Allen's astrophysical quantities, (4th ed. New York: AIP Press; Springer)
- Cristiani, S. et al. 2000, A&A, 359, 489
- Daddi, E., Cimatti, A., Pozzetti, L., Hoekstra, H., Röttgering, H. J. A., Renzini, A., Zamorani, G., & Mannucci, F. 2000, A&A, 361, 535
- Daddi, E. et al. 2003, ApJ, 588, 50
- Dickinson, M., Papovich, C., Ferguson, H. C., & Budavári, T. 2003, ApJ, 587, 25
- Eisenhardt, P. & Dickinson, M. 1992, ApJ, 399, L47
- Erb, D. K., Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., Hunt, M. P., Moorwood, A. F. M., & Cuby, J. 2003, ApJ, 591, 101
- Ferguson, H. C., Dickinson, M., & Papovich, C. 2002, ApJ, 569, L65
- Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, ApJ, 513, 34
- Folkes, S. et al. 1999, MNRAS, 308, 459
- Förster Schreiber, N. et al. in preparation
- Fontana, A., D'Odorico, S., Poli, F., Giallongo, E., Arnouts, S., Cristiani, S., Moorwood, A., & Saracco, P. 2000, AJ, 120, 2206
- Fontana, A. et al. 2003, ApJ, 594, L9
- Franx, M. et al. 2003, ApJ, 587, L79
- Frenk, C. S., White, S. D. M., Efstathiou, G., & Davis, M. 1985, Nature, 317, 595
- Fried, J. W. et al. 2001, A&A, 367, 788
- Fukugita, M., Shimasaku, K., & Ichikawa, T. 1995, PASP, 107, 945
- Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748
- Gialalisco, M., Steidel, C. C., Adelberger, K. L., Dickinson, M. E., Pettini, M., & Kellogg, M. 1998, ApJ, 503, 543
- Gialalisco, M. & Dickinson, M. 2001, ApJ, 550, 177
- Glazebrook, K. et al. 2003, in preparation
- Glazebrook, K. et al. 2003, ApJ, 587, 55
- Hauschildt, P. H., Allard, F., & Baron, E. 1999, ApJ, 512, 377
- Hogg, D. W. et al. 2002, AJ, 124, 646
- Huchra, J. P. 1977, ApJ, 217, 928
- Jansen, R. A., Franx, M., Fabricant, D., & Caldwell, N. 2000, ApJS, 126, 271
- Jansen, R. A., Fabricant, D., Franx, M., & Caldwell, N. 2000, ApJS, 126, 331
- Kashikawa, N. et al. 2003, AJ, 125, 53
- Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201
- Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 303, 188
- Kauffmann, G. et al. 2003, MNRAS, 341, 33
- Kinney, A. L., Calzetti, D., Bohlin, R. C., McQuade, K., Storchi-Bergmann, T., & Schmitt, H. R. 1996, ApJ, 467, 38
- Labbé, I. et al. 2003, AJ, 125, 1107
- Lanzetta, K. M., Yahata, N., Pascarelle, S., Chen, H., & Fernández-Soto, A. 2002, ApJ, 570, 492
- Larson, R. B. & Tinsley, B. M. 1978, ApJ, 219, 46
- Lilly, S. J., Le Fevre, O., Crampton, D., Hammer, F., & Tresse, L. 1995, ApJ, 455, 50
- Lilly, S. J., Le Fevre, O., Hammer, F., & Crampton, D. 1996, ApJ, 460, L1
- Madau, P. 1995, ApJ, 441, 18

- Madau, P., Pozzetti, L., & Dickinson, M. 1998, ApJ, 498, 106
- Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996, MNRAS, 283, 1388
- McCarthy, P. J. et al. 2001, ApJ, 560, L131
- Moorwood, A. F. 1997, Proc. SPIE, 2871, 1146
- Norberg, P. et al. 2002, MNRAS, 336, 907
- Papovich, C., Dickinson, M., Ferguson, H. C., 2001, ApJ, 559, 620
- Pei, Y. C., Fall, S. M., & Hauser, M. G. 1999, ApJ, 522, 604
- Pettini, M., Steidel, C. C., Adelberger, K. L., Dickinson, M., & Giavalisco, M. 2000, ApJ, 528, 96
- Pettini, M., Shapley, A. E., Steidel, C. C., Cuby, J., Dickinson, M., Moorwood, A. F. M., Adelberger, K. L., & Giavalisco, M. 2001, ApJ, 554, 981
- Pettini, M., Rix, S. A., Steidel, C. C., Adelberger, K. L., Hunt, M. P., & Shapley, A. E. 2002, ApJ, 569, 742
- Poli, F., Menci, N., Giallongo, E., Fontana, A., Cristiani, S., & D'Odorico, S. 2001, ApJ, 551, L45
- Poli, F. et al. 2003, ApJ, 593, L1
- Rigopoulou, D. et al. 2000, ApJ, 537, L85
- Rudnick, G. et al. 2003, in preparation
- Rudnick, G., Franx, M., Rix, H.-W., Moorwood, A., Kuijken, K., van Starkenburg, L., van der Werf, P., Röttgering, H., Labbé, I. 2001, AJ, 122, 2205
- Salpeter, E. E. 1955, ApJ, 121, 161
- Sawicki, M. J., Lin, H., & Yee, H. K. C. 1997, AJ, 113, 1
- Scalo, J. M. 1986, Fundamentals of Cosmic Physics, 11, 1
- Schade, D. et al. 1999, ApJ, 525, 31
- Seaton, M. J. 1979, MNRAS, 187, 73P
- Shapley, A. E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., & Pettini, M. 2001, ApJ, 562, 95
- Skrutskie, M. F. et al. 1997, ASSL Vol. 210: The Impact of Large Scale Near-IR Sky Surveys, 25
- Somerville, R. S., Primack, J. R., & Faber, S. M. 2001, MNRAS, 320, 504
- Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJ, 462, L17
- Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1
- Stoughton, C. et al. 2002, AJ, 123, 485
- Strauss, M. A. et al. 2002, AJ, 124, 1810
- Thompson, R. I., Weymann, R. J., & Storrie-Lombardi, L. J. 2001, ApJ, 546, 694
- Thompson, R. I. 2003, ApJin press, *astro-ph/0306594*
- van de Ven, G., van Dokkum, P. G., & Franx, M. 2003, submitted to MNRAS, *astro-ph/0211566*
- van Dokkum, P. G., Franx, M., Kelson, D. D., & Illingworth, G. D. 2001, ApJ, 553, L39
- van Dokkum, P. G. et al. 2003, ApJ, 587, L83
- Vanzella, E. et al. 2002, A&A, 396, 847
- Williams, R. E. et al. 1996, AJ, 112, 1335
- Wilson, G., Cowie, L. L., Barger, A. J., & Burke, D. J. 2002, AJ, 124, 1258
- Wolf, C., Meisenheimer, K., Rix, H.-W., Borch, A., Dye, S., & Kleinheinrich, M. 2003, A&A, 401, 73
- York, D. G. et al. 2000, AJ, 120, 1579

Zepf, S. E. 1997, Nature, 390, 377

Appendix A. Derivation of z_{phot} Uncertainty

Given a set of formal flux errors, one way to broaden the redshift confidence interval without degrading the accuracy (as noticed in R01) is to lower the absolute χ^2 of every $\chi^2(z)$ curve without changing its shape (or the location of the minimum). By scaling up all the flux errors by a constant factor, we can retain the relative weights of the points in the χ^2 without changing the best fit redshift and SED, but we do enlarge the redshift interval over which the templates can satisfactorily fit the flux points. Since we believe the disagreement between z_{spec} and z_{phot} is due to our finite and incomplete template set, this factor should reflect the degree of template mismatch in our sample, i.e., the degree by which our models fail to fit the flux points. To estimate this factor we first compute the fractional difference between the model and the data $\Delta_{i,j}$ for the j^{th} galaxy in the i^{th} filter,

$$\Delta_{i,j} = \frac{(f_{i,j}^{\text{mod}} - f_{i,j}^{\text{dat}})}{f_{i,j}^{\text{dat}}} \quad (\text{A.1})$$

where f^{mod} are the predicted fluxes of the best-fit template combination and f^{dat} are our actual data. For each galaxy we calculated

$$\Delta_j = \sqrt{\frac{1}{N_{\text{filt}} - 1} \sum_{i=2}^{N_{\text{filt}}} \Delta_{i,j}^2} \quad (\text{A.2})$$

where we have ignored the contributions of the U -band. While the U -band is important in finding breaks in the SEDs, the exact shapes of the templates are poorly constrained blueward of the rest-frame U -band and the U -data often deviates significantly from the best-fit model fluxes.

To determine the mean deviation of all of the flux points from the model Δ_{dev} we then averaged over all galaxies in our complete FIRES sample with $K_{s,AB}^{\text{tot}} < 22$ (for which the systematic z_{phot} errors should dominate over those resulting from photometric errors) to obtain

$$\Delta_{dev} = \frac{1}{N_{\text{gal}}} \sum_{j=1}^{N_{\text{gal}}} |\Delta_j|. \quad (\text{A.3})$$

We find $\Delta_{\text{dev}} = 0.08$, which includes both random and systematic deviations from the model. We modified the Monte-Carlo simulation of R01 by calculating, for each object j ,

$$\left\langle \frac{S}{N} \right\rangle_j = \sqrt{\frac{\sum_{i=2}^{N_{\text{filt}}} \left(\frac{f_i}{\delta f_i} \right)^2}{N_{\text{filt}} - 1}} \quad (\text{A.4})$$

again excluding the U -band. We then scaled the flux errors, for each object, using the following criteria:

$$\delta f'_i = \begin{cases} \delta f_i & : \left\langle \frac{S}{N} \right\rangle_j \leq \frac{1}{\Delta_{\text{dev}}} \\ \delta f_i \Delta_{\text{dev}} \left\langle \frac{S}{N} \right\rangle_j & : \left\langle \frac{S}{N} \right\rangle_j > \frac{1}{\Delta_{\text{dev}}} \end{cases} . \quad (\text{A.5})$$

The photometric redshift error probability distribution is computed using the $\delta f'_i$'s. Note that this procedure will not modify the z_{phot} errors of the objects with low S/N where the z_{phot} errors are dominated by the formal photometric errors. The resulting probability distribution is highly non-Gaussian and using it we calculate the upper and lower 68% confidence limits on the redshift $z_{\text{phot}}^{\text{hi}}$ and $z_{\text{phot}}^{\text{low}}$ respectively. As a single number which encodes the total range of acceptable z_{phot} 's, we define $\delta z_{\text{phot}} \equiv 0.5 * (z_{\text{phot}}^{\text{hi}} - z_{\text{phot}}^{\text{low}})$.

Figure 6 from L03 shows the comparison of z_{phot} to z_{spec} . For these bright galaxies, it is remarkable that our new photometric redshift errorbars come so close to predicting the difference between z_{phot} and the true value. Some galaxies have large δz_{phot} values even when the local χ^2 minimum is well defined because there is another χ^2 minimum of comparable depth that is contained in the 68% redshift confidence limits. There are galaxies with $\delta z_{\text{phot}} < 0.05$. Some of these are bright low redshift galaxies with large rest-frame optical breaks, which presumably place a strong constraint on the allowed redshift. Many of these galaxies, however, are faint and the δz_{phot} is unrealistically low. Even though these faint galaxies have $\left\langle \frac{S}{N} \right\rangle_j \leq \frac{1}{\Delta_{\text{dev}}}$, they still can have high S/N in the B_{450} or V_{606} bandpasses and hence have steep χ^2 curves and small inferred redshift uncertainties. In addition, many of these galaxies have $z_{\text{phot}} > 2$ and very blue continuum longward of $\text{Ly}\alpha$. The imposed sharp discontinuity in the template SEDs at the onset of HI absorption causes a very narrow minimum in the $\chi^2(z)$ curve, and hence a small δz_{phot} , but likely differs from the true shape of the discontinuity because we use the mean opacity values of Madau (1995), neglecting its variance among different lines of sight.

It is difficult to develop a scheme for measuring realistic photometric redshift uncertainties over all regimes. The δz_{phot} estimate derives the z_{phot} uncertainties individually for each object, but can underpredict the uncertainties in some cases. Compared to the technique of R01 however, a method based completely on the Monte-Carlo technique is preferable because it has a straightforwardly computed redshift probability function. This trait is desirable for estimating the errors in the rest-frame luminosities and colors and for this reason we will use δz_{phot} as our uncertainty estimate in this paper.

Appendix B. Rest-Frame Photometric System

To define the rest-frame U , B , and V fluxes we use the filter transmission curves and zeropoints tabulated in Bessell (1990), specifically his UX , B , and V filters. The Bessell zeropoints are given as magnitude offsets with respect to a source which has constant f_ν and $AB = 0$. The AB magnitude is defined as

$$AB_\nu = -2.5 * \log_{10} \langle f_\nu \rangle - 48.58 \quad (\text{B.1})$$

where $\langle f_\nu \rangle$ is the flux $f_\nu(\nu)$ observed through a filter $T(\nu)$ and in units of $\text{ergs s}^{-1} \text{cm}^{-2} \text{Hz}^{-1}$. Given the zeropoint offset ZP_ν for a given filter, the Vega magnitude m_ν is then

$$m_\nu = AB_\nu - ZP_\nu = -2.5 * \log_{10} \langle f_\nu \rangle - 48.58 - ZP_\nu. \quad (\text{B.2})$$

All of our observed fluxes and rest-frame template fluxes are expressed in f_λ . To obtain rest-frame magnitudes in the Bessell (1990) system, we must calculate the conversion from f_λ to f_ν for the redshifted rest-frame filter set. The flux density of an SED with $f_\lambda(\lambda)$ integrated through a given filter with transmission curve $T(\lambda)$ is

$$\langle f_\lambda \rangle = \frac{\int f_\lambda(\lambda') T'(\lambda') d\lambda'}{\int T'(\lambda') d\lambda'} \quad (\text{B.3})$$

or

$$\langle f_\nu \rangle = \frac{\int f_\nu(\nu') T'(\nu') d\nu'}{\int T'(\nu') d\nu'}. \quad (\text{B.4})$$

Since

$$\int f_\lambda(\lambda') T'(\lambda') d\lambda' = \int f_\nu(\nu') T'(\nu') d\nu' \quad (\text{B.5})$$

we can convert to $\langle f_\nu \rangle$ through

$$\langle f_\nu \rangle = \langle f_\lambda \rangle * \frac{\int T'(\lambda') d\lambda'}{\int T'(\nu') d\nu'} \quad (\text{B.6})$$

and use $\langle f_\nu \rangle$ to calculate the apparent rest-frame Vega magnitude through the redshifted filter via Eq. B.2.

Appendix C. Estimating Rest-Frame Luminosities

We derive for any given redshift, the relation between the apparent AB magnitude m_{λ_z} of a galaxy through a redshifted rest-frame filter, its observed fluxes $\langle f_{\lambda_i, obs} \rangle$ in the different filters i , and the colors of the spectral templates. At redshift z , the rest-frame filter with effective wavelength λ_{rest} has been shifted to an observed wavelength

$$\lambda_z = \lambda_{rest} \times (1 + z) \quad (\text{C.1})$$

and we define the adjacent observed bandpasses with effective wavelengths λ_l and λ_h which satisfy

$$\lambda_l < \lambda_z \leq \lambda_h. \quad (\text{C.2})$$

We now define

$$C_{obs} \equiv m_{obs,\lambda_l} - m_{obs,\lambda_h} \quad (\text{C.3})$$

where m_{obs,λ_l} and m_{obs,λ_h} are the AB magnitudes which correspond to the fluxes $\langle f_{\lambda_l,obs} \rangle$ and $\langle f_{\lambda_h,obs} \rangle$ respectively. We then shift each template in wavelength to the redshift z and compute,

$$C_{templ} \equiv m_{templ,\lambda_l} - m_{templ,\lambda_h}, \quad (\text{C.4})$$

where m_{templ,λ_l} and m_{templ,λ_h} are the AB magnitudes through the λ_l and λ_h observed bandpasses (including the atmospheric and instrument throughputs). We sort the templates by their C_{templ} values, $C_{templ,a}$, $C_{templ,b}$, etc., and find the two templates such that

$$C_{templ,a} \leq C_{obs} < C_{templ,b}. \quad (\text{C.5})$$

We then define for the a^{th} template

$$C_{\lambda_l,z,a} \equiv m_{templ,\lambda_l} - m_{templ,\lambda_z} \quad (\text{C.6})$$

where m_{templ,λ_z} is the apparent AB magnitude of the redshifted a^{th} template through the redshifted λ_{rest} filter. We point out that because our computations always involve colors, they are not dependent on the actual template normalization (which cancels out in the difference). Taking our observed color C_{obs} and the templates with adjacent “observed” colors $C_{templ,a}$ and $C_{templ,b}$, we can interpolate between $C_{\lambda_l,z,a}$ and $C_{\lambda_l,z,b}$

$$m_{obs,\lambda_l} - m_{\lambda_z} = C_{\lambda_l,z,a} + (C_{obs} - C_{templ,a}) \times \left(\frac{C_{\lambda_l,z,b} - C_{\lambda_l,z,a}}{C_{templ,b} - C_{templ,a}} \right) \quad (\text{C.7})$$

and solve for m_{λ_z} .

When C_{obs} lies outside the range of the C_{templ} ’s, we simply take the two nearest templates in observed C_{templ} space and extrapolate Eq. C.7 to compute m_{λ_z} .

Equation C.7 has the feature that $m_{\lambda_z} \approx m_{obs,\lambda_l}$ when $\lambda_z = \lambda_l$ (and hence when $C_{\lambda_l,z,a}$ and $C_{\lambda_l,z,b} \approx 0$). While this method still assumes that the templates are reasonably good approximations to the true shape of the SEDs it has the advantage that it does not rely on exact agreement. Galaxies whose observed colors fall outside the range of the templates can also be easily flagged. A final advantage of this method is that the uncertainty in m_{λ_z} can be readily calculated from the errors in the observed fluxes.

From m_{λ_z} , we compute the rest-frame luminosity by applying the K-correction and converting to luminosity units

$$\frac{L^{\text{rest}}}{L_{\odot}} = 10^{-0.4(m_{\lambda_z} - M_{\odot,\lambda_{rest}} - ZP_{\lambda_{rest}})} \times \left(\frac{D_L}{10pc} \right)^2 \times (1+z)^{-1} \times h^{-2} \quad (\text{C.8})$$

where $M_{\odot, \lambda_{rest}}$ is the absolute magnitude of the sun in the λ_{rest} filter ($M_{\odot, U} = +5.66$, $M_{\odot, B} = +5.47$, and $M_{\odot, V} = +4.82$ in Vega magnitudes; Cox 2000), $ZP_{\lambda_{rest}}$ is the zeropoint in that filter (as in Eq. B.2 but expressed at λ and not at ν), and D_L is the distance modulus in parsecs. Following R01, we correct this luminosity by the ratio of the K_s^{tot} flux to the modified isophotal aperture flux (see L03). This adjustment factor, which accounts both for the larger size of the total aperture and the aperture correction, changes with apparent magnitude and it ranges from 1.23, at $20 < K_{s, AB}^{tot} \leq 24$, to 1.69, at $24 < K_{s, AB}^{tot} \leq 25$, and it has an RMS dispersions of 0.17 and 0.49 in the two magnitude bins respectively.

The uncertainty in the derived L_{λ}^{rest} has contributions both from the observational flux errors and from the redshift uncertainty, which causes λ_z to move with respect to the observed filters. The first effect is estimated by propagating the observed flux errors through Eq. C.7. As an example, object 531 at $z_{phot} = 2.20$ has $K_{s, AB}^{tot} = 24.91$ and signal-to-noise in the K_s -band of 8.99 and 5.43 in our modified isophotal and total apertures respectively. The resultant error in L_V^{rest} purely from flux errors is then 26%. At $K_{s, AB}^{tot} \approx 24$, the typical signal-to-noise in the K_s -band increases to ≈ 13 and ≈ 6.3 in our modified isophotal and total apertures respectively and the error L_{λ}^{rest} decreases accordingly.

To account for the redshift dependent error in the calculated luminosity, we use the Monte-Carlo simulation described first in R01 and updated in §Appendix A.. For each Monte-Carlo iteration we calculate the rest-frame luminosities and determine the 68% confidence limits of the resulting distribution. The 68% confidence limits in L_{λ}^{rest} can be highly asymmetric, just as for z_{phot} . For objects with $K_{s, AB}^{tot} \lesssim 25$ we find that the contributions to the total L_{λ}^{rest} error budget are dominated by the redshift errors rather than by the flux errors.

Field Galaxies at $1 < z < 3$ The Color Magnitude Distribution of the evolution and modeling of the blue sequence

ABSTRACT

We use deep near-infrared VLT/ISAAC imaging to study the rest-frame color-magnitude distribution of infrared selected galaxies in the redshift range $1 < z < 3$. We find a well-defined blue peak of star-forming galaxies at all redshifts. The blue galaxies populate a color-magnitude relation (CMR), such that more luminous galaxies in the rest-frame V -band tend to have redder ultraviolet-to-optical colors. The slope of the CMR does not evolve with time, and is similar to the slope of blue, late-type galaxies in the local universe. Analysis of spectra of nearby late-type galaxies suggests that the steepness of the slope can be fully explained by the observed correlation of dust content with optical luminosity. The zeropoint of the blue CMR at a given magnitude reddens smoothly from $z = 3$ to $z = 0$, likely reflecting an increase of the mean stellar age and an increase in the mean dust opacity of blue-sequence galaxies. A distinct feature is that the color scatter around the $z \sim 3$ CMR is asymmetric, with a blue “ridge” and a skew towards red colors. We have explored which types of star formation histories can reproduce the scatter and the skewed shape of the color distribution. These included models with constant star formation rates and sudden cutoffs, exponentially declining star formation rates, burst models, and models with episodic star formation. The episodic models reproduced the color distribution best, with quiescent periods lasting 30-50% of the length of an active period, and duration of the duty cycle between 150 Myr to 1 Gyr. The episodic star formation in these models rejuvenate the galaxies during each episode, making it significantly bluer than a galaxy with constant star formation of the same age. This could be a solution of the enigmatic observation that $z = 3$ galaxies are much bluer than expected if they were as old as the universe. Finally, the color distribution has a strong tail of very red galaxies. The relative number of red galaxies increases sharply from $z \sim 3$ to $z \sim 1$. The rest-frame V -band luminosity density in luminous blue-sequence galaxies is constant, or decreases, whereas that in red galaxies rises with time. We are viewing the progressive formation of red, passively evolving galaxies.

Ivo Labbé, Marijn Franx, Gregory Rudnick, Natascha M. Förster Schreiber,
Emanuele Daddi, Pieter G. van Dokkum, Konrad Kuijken, Alan Moorwood,
Hans-Walter Rix, Huub Röttgering, Ignacio Trujillo, Arjen van der Wel, Paul van
der Werf, & Lottje van Starkenburg

1 Introduction

THE formation and evolution of galaxies is a complex process, involving the growth of dark-matter structure from the gradual hierarchical merging of smaller fragments (e.g., White & Frenk 1991; Kauffmann & White 1993), the accretion of gas, the formation of stars, and feedback from supernovae and black holes. Presently, the theories describing the growth of large-scale dark-matter structure are thought to be well-constrained (Freedman et al. 2001; Efstathiou et al. 2002; Spergel et al. 2003), but the formation history of the stars inside the dark-matter halos is still poorly understood. Neither hydrodynamical simulations (e.g., Katz & Gunn 1991; Springel et al. 2001; Steinmetz & Navarro 2002) nor state-of-the-art semianalytic models (e.g., Kauffmann, White, & Guiderdoni 1993; Somerville & Primack 1999; Cole, Lacey, Baugh, & Frenk 2000) provide unique predictions for the formation of the stars in galaxies, given the large parameter space available to these models. Direct observations are critical to constrain them.

Specifically the observations of massive galaxies provide strong tests for pictures of galaxy formation, as their build-up can be directly observed from high redshift to the present epoch. The largest samples of high-redshift galaxies to date have been selected by their rest-frame UV light, through the Lyman Break technique (LBGs; Steidel et al. 1996a,b, 2003). Unfortunately, the rest-frame UV light is highly susceptible to dust and not a good measure of the number of intermediate and low mass stars, which may dominate the stellar mass. In fact, rest-frame optical observations have shown LBGs to be relatively low mass ($M \sim 10^{10} M_{\odot}$), unobscured, star-forming galaxies (e.g., Papovich, Dickinson, & Ferguson 2001; Shapley et al. 2001).

The rest-frame optical light is already much less sensitive to the effects of dust obscuration and on-going star formation than the UV, and it is expected to be a better tracer of stellar mass. Recent advances in near-infrared (NIR) capabilities on large telescopes are now making it possible to select statistically meaningful samples of galaxies by their rest-frame optical light out to $z \sim 3$. In this context we started the Faint Infrared Extragalactic Survey (FIRES; Franx et al. 2000), a deep optical-to-infrared multicolor survey of NIR-selected galaxies. The rest-frame optical observations are also useful as much of our knowledge in the local universe is based on studies at these wavelengths.

A particular diagnostic of galaxy formation is the shape of the galaxy luminosity and mass functions as a function of the spectral type. In the local universe these distributions have been determined in great detail (e.g., Strateva et al. 2001; Norberg et al. 2002; Blanton et al. 2003; Kauffmann et al. 2003). While a precise characterisation of the local distributions can strongly constrain the models, powerful tests can also be made using their evolution as a function of redshift. For example, Bell et al. (2004a, 2004b) show that the V -band luminosity density of photometrically selected early-type galaxies is nearly constant in the range $0.2 < z < 1.1$, suggesting an increase in stellar mass of passively evolving galaxies

over this range. This is not expected from the simplest collapse models, where all galaxies form at high redshift and evolve passively to the present-day.

Different constraints on galaxy formation come from correlations between integrated galaxy properties. For example, morphologically early-type galaxies in the local universe and distant clusters populate a well-defined color-magnitude relation (CMR), which is thought to reflect a sequence of increasing metallicity with stellar mass (e.g., Bower, Lucey, & Ellis 1992; Schweizer & Seitzer 1992; Kodama & Arimoto 1997). The small scatter of this relation implies that early type galaxies formed most of their stars at high redshift (e.g., van Dokkum et al. 1998). Blue, star forming, late-type galaxies also display a color-magnitude relation (Chester & Roberts 1964; Visvanathan 1981; Tully, Mould, & Aaronson 1982), but the scatter is larger (Griensmith 1980), and its origin is probably more complex; it has been interpreted as a sequence in the mean stellar age (Peletier & de Grijs 1998), dust attenuation (Tully et al. 1998), and/or metallicity (Zaritsky, Kennicutt, & Huchra 1994, Bell & De Jong 2000).

Interestingly, Papovich, Dickinson, & Ferguson (2001) found a color-magnitude relation for blue, star-forming galaxies at $z \sim 3$, from a sample of NIR-selected galaxies in the Hubble Deep Field North (HDFN). The trend, which is also seen in the similar field of the Hubble Deep Field South (Labbé et al. 2003), is such that galaxies more luminous in the rest-frame V -band, tend to have redder ultraviolet-to-optical colors. A distinct aspect of the high-redshift CMR, is that the galaxies are asymmetrically distributed around the relation, with a well-defined blue envelope (Papovich et al. 2004).

Studies of colors and magnitudes of star-forming galaxies at $z > 2$ also raised questions. A particular puzzle was presented by modeling of their stellar populations, which implied average luminosity weighted ages of a few 100 Myr, much younger than the age of the universe at these redshifts (e.g., Papovich, Dickinson, & Ferguson 2001; Shapley et al. 2001). From this, and from the relative absence of candidates for red, non star-forming galaxies in the HDFN it was suggested that star formation in LBGs occurs with short duty cycles and a timescale between star formation events of $\lesssim 1$ Gyr.

In this paper, we investigate the evolution of the rest-frame colors of galaxies as a function of redshift in the range $1 \lesssim z \lesssim 3$. The deep optical-to-NIR imaging and the homogeneous photometry of the FIRES project is excellently suited for such studies, and we use it to analyze rest-frame ultraviolet-to-optical colors and magnitudes of a sample of 1475 K_s -band selected galaxies.

We focus on the galaxies that populate the blue peak of galaxies at low and high redshift. We wish to understand the nature of the blue color-magnitude relation, the evolution of the colors towards high redshift, and the origin of the conspicuous skewed color distribution around the CMR at $z \sim 3$. For the latter we explore models with different types of star formation histories that might produce such distributions. Finally, we chart the evolution of the relative number of red

galaxies over the redshift range $1 \lesssim z \lesssim 3$.

This paper is organized as follows. We present the data in §2, describe the color magnitude distribution of FIRES galaxies in §3, analyze the blue color-magnitude relation in §4, and model the scatter of galaxies around the blue CMR in §5. Finally, §6 presents the evolution of the red galaxy fraction. Where necessary, we adopt an $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ cosmology. We use magnitudes calibrated to models for Vega throughout.

2 The Data

2.1 The Observations and Sample Selection

The observations were obtained as part of the public Faint Infrared Extragalactic Survey (FIRES; Franx et al. 2000) the deepest groundbased NIR survey to date. We cover two fields with existing deep optical WFPC2 imaging from the *Hubble Space Telescope* (HST): the WFPC2-field of HDFs, and the field around the $z = 0.83$ cluster MS1054-03. The observations, data reduction, and assembly of the catalog source catalogs are presented in detail by Labbé et al. (2003) for the HDFs and Förster Schreiber et al. (2004a) for the MS1054 field.

Briefly, we observed in the NIR J_s , H , and K_s bands with the Infrared Spectrometer and Array Camera (ISAAC; Moorwood 1997) at the *Very Large Telescope* (VLT). In the HDFs, a total of 101.5 hours was invested in a single $2.5' \times 2.5'$ pointing, resulting in the deepest groundbased NIR imaging, and the deepest K -band to date, even from space. We complemented the existing deep optical HST WFPC2 imaging in the U_{300} , B_{450} , V_{606} , I_{814} bands (Casertano et al. 2000). A further 77 hours of NIR imaging was spent on a mosaic of four ISAAC pointings centered on the $z = 0.83$ foreground cluster MS1054-03, reaching somewhat shallower depths. We complemented the data with WFPC2 mosaics in the V_{606} and I_{814} bands (van Dokkum et al. 2000), and collected additional imaging with the VLT FORS1 instrument in the U , B , and V bands (Förster Schreiber et al. 2004a). In both surveyed fields the effective seeing in the final NIR images was $\approx 0''.45 - 0''.55$ FWHM.

We detect objects in the K_s -band using version 2.2.2 of the SExtractor software ((Bertin & Arnouts 1996). For consistent photometry accross all bands, all images were aligned, and accurately PSF-matched to the filter in which the image quality was worst. Stellar curve of growth analysis indicates that the fraction of enclosed flux agrees to better than 3% for the apertures relevant to our color measurements. The color measurements were done in a customized isophotal aperture defined from the K_s -image. The estimate of total flux in the K_s -band was computed using SExtractor's *AUTO* aperture for isolated sources, and in an adaptive circular aperture for blended sources. In both cases, a minimal aperture correction for the light lost by a point source was applied. Photometric uncertainties were derived empirically from the flux distribution in apertures placed on empty parts of the

maps. For details concerning all aspects of the photometric measurements, see Labbé et al. 2003. The total $5\text{-}\sigma$ limiting depth for point sources are $K_s^{tot} = 23.8$ for the HDFS, and 23.1 for the MS1054-field.

2.2 Photometric Redshifts and Rest-Frame Colors

Photometric redshifts were estimated by fitting a linear combination of redshifted empirical galaxy spectra, and a 10 Myr old simple stellar population model (1999 version of Bruzual A. & Charlot 1993) to the observed flux points. The algorithm is described in detail by Rudnick et al. (2001, 2003). We adopt a minimux flux error of 5% for all bands to account for zeropoint uncertainties and for mismatches between the observations and the photo-z template set.

Monte-Carlo simulations were used to estimate the errors $\delta z_{ph,MC}$ on the photometric redshifts. These errors reflect the photometric uncertainties, template mismatch, and the possibility of secondary solutions. We determined the accuracy of the technique from comparisons to the available spectroscopy in each field. We find $\delta z = \langle |z_{spec} - z_{phot}| / (1 + z_{spec}) \rangle = 0.07$, and $\delta z = 0.05$ for sources at $z \geq 2$. The errors calculated from simulations $\delta z_{ph,MC}$ are consistent with this. We identified and removed stars using the method described in Rudnick et al. (2003).

We combine the observed SEDs and photometric redshifts to derive rest-frame luminosities L_λ^{rest} . We used a method of estimating L_λ^{rest} that interpolates directly between the observed fluxes, using the templates as a guide. The rest-frame photometric system, and details on estimating L_λ^{rest} are described extensively by (Rudnick et al. 2003). Throughout we will use the rest-frame UX , B , and V filters Beers et al. (1990) and the HST/FOC $F140W$, $F170W$, and $F220W$ filters, which we will call 1400, 1700, and 2200 throughout. We adopted the photometric system of Bessell (1990) for the optical filters, which was calibrated to the Dreiling and Bell (1980) model spectrum for Vega. The HST/FOC UV zeropoints were calibrated to the Kurucz (1992) model for Vega.

The rest-frame luminosities are sensitive to the uncertainties in the photometric redshifts. Therefore we only analyze the sample of galaxies with $\delta z_{ph,MC} / (1 + z_{ph}) < 0.2$, keeping 1354 out of 1475 galaxies. The median $\delta z_{ph,MC} / (1 + z_{ph})$ for the remaining sample is 0.05. We checked that the color distribution of the rejected galaxies was consistent with that of the galaxies we kept.

The reduced images, photometric catalog, redshifts, and rest-frame luminosities are all available on-line through the FIRES homepage¹.

¹<http://www.strw.leidenuniv.nl/~fires>

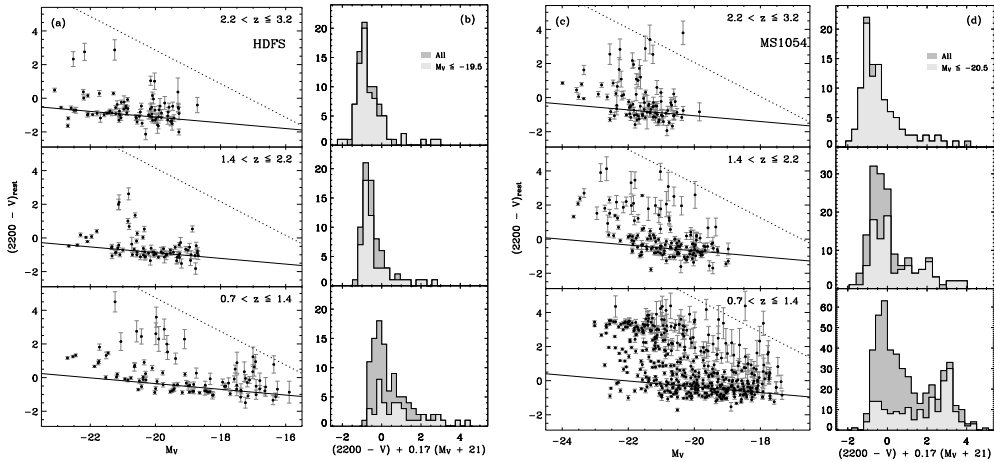


Figure 1 — The rest-frame $2200 - V$ colors against absolute magnitude in the V -band for galaxies in the field of the HDFS (a) and in the field of MS1054 (c). The two samples are split into three redshift bins. The errorbars represent the 1σ uncertainties on the rest-frame colors. The dotted line roughly marks a conservative color limit, corresponding to the 2σ flux uncertainty in the observed filter that is closest to 2200-band at the maximum redshift of the bin. Galaxies redward of this line have uncertain rest-frame colors, but can be observed. The solid line shows a fit of a linear relation with a fixed slope of -0.17 to the galaxies in the blue peak of the color-magnitude distribution. Panels b (HDFS) and d (MS1054) show histograms of the colors after the slope is subtracted. The peak of this distribution is normalized to the intercept of the CMR at $M_V = -21$. We show the color distributions of all detected galaxies in each bin (dark gray histograms) and those to a limiting magnitude of $M_V \leq -19.5$ and $M_V \leq -20.5$ in the field of the HDFS and MS1054 respectively (light gray histograms).

3 The rest-frame Color-Magnitude Distribution of Galaxies from $z \sim 1$ to $z \sim 3$

A well-studied diagnostic in the local universe is the optical color-magnitude distribution (e.g., Strateva et al. 2001; Hogg et al. 2002; Baldry et al. 2004). Quantifying this distribution out to higher redshifts, will provide strong constraints on models of galaxy formation, which must reproduce these observations.

To quantify the evolution of galaxy colors and magnitude with redshift, we compare the distribution at the rest-frame $2200 - V$ color versus absolute V -band magnitude. We present evolution in terms of the $2200 - V$ color in panels a and c of Figure 1. These colors are sensitive to variations in stellar population age, as they span the Balmer and $4000 \text{ \AA}$ breaks, and attenuation by dust, by virtue of the long wavelength baseline extending into the UV. We prefer the rest-frame 2200-band to the more commonly used U -band, because at $z = 2 - 3$ the 2200-band is probed by our deepest optical observations. In addition, the uncertainties in our photometric redshifts tend to increase the scatter more in the U and B colors than in the 2200 colors as the former bands directly straddle the Balmer

and 4000 Å breaks.

We divide the galaxies in three redshift bins centered at $z=1.0$, 1.8 , and $z=2.7$. The redshift range is defined so that the rest-frame 2200 and V -band always lies within the range of our observed filters. The number of redshift bins is a compromise between keeping a statistically meaningful sample at each redshift, while reducing the effects of evolution over the time span in each bin. The limits of the bins are defined so that galaxies in the smaller field of the HDFs are equally divided in number. The redshift bins of the larger MS1054 field are matched to those of the HDFs.

Clearly, in both fields and in all redshift bins there is a well-defined ridge of blue galaxies. Furthermore, the ridge is tilted: the more luminous galaxies along the ridge tend to have redder $2200 - V$ colors. This ridge is what we define as the color magnitude relation of blue galaxies.

There is no reason to think that this color-magnitude relation is due to some selection effect; we can easily detect galaxies that are bright in M_V and blue in $2200 - V$, and our photometry is sufficiently deep to ensure that the trend is not caused by a lack of faint red galaxies. Only at low redshifts ($z \sim 1$) and faint magnitudes may we miss some of the bluest galaxies, as the sample was selected the observed K_s -filter, which is significantly redder than the redshifted V -band at $z \sim 1$.

It is striking that the blue color magnitude relation has a very well-defined boundary to the blue, and a much more extended tail to the red. This red tail extends up to 4 magnitudes, and defines a clear red color magnitude relation in the lowest redshift bin in the MS1054 field. This is mainly caused by the color-magnitude relation of the cluster galaxies at $z = 0.83$, and is also observed in the field up to $z = 1$ (e.g., Bell et al. 2004a)

We now proceed to analyze the properties of the blue color magnitude in sections 4 and 5, and will turn to the red galaxies in section 6.

4 The Color-Magnitude Relation of Blue Field Galaxies

Two key features of the blue CMR in Fig. 1 are that the peak of the distribution, and hence the zeropoint of the blue sequence, reddens from $z \sim 3$ to $z \sim 1$, while on first sight the slope does not seem to vary appreciably. We proceed by quantifying the slope and zeropoint of the relation and their evolution with redshift.

4.1 The Slope and its Evolution

We assume that the distribution of colors and magnitudes can be characterized by a linear relation. We obtain the best fitting slope as follows: for each value of the slope we obtain a distribution of color residuals. We select the slope that maximizes the peak of the histogram of the residuals. This so-called “mode re-

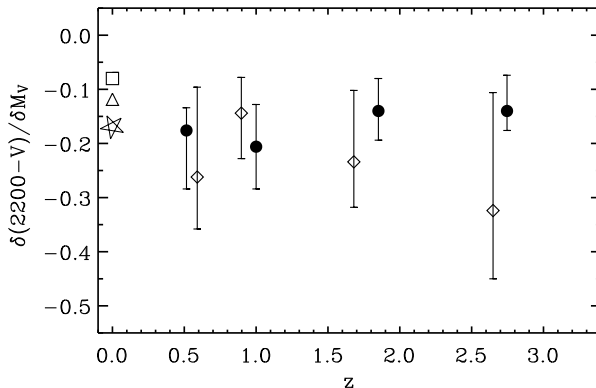


Figure 2 — The evolution of the slope of the blue color-magnitude relation from linear fits to the galaxy distributions in the HDFS (*filled circles*) and MS1054 (*diamonds*). The results are plotted at the median redshift of the galaxies in the bin. The uncertainties correspond to the 68% confidence interval obtained with bootstrap resampling. The data are consistent with a constant slope with redshift. Also drawn are values for blue late type galaxies from the Nearby Field Galaxy Survey (Jansen, Franx, & Fabricant 2000a), where we assume the CMR is caused by a systematic trend with dust reddening (*star*), or stellar age (*triangle*). Finally, we show the value for the metallicity-luminosity relation of local early type galaxies (*square*; Bower, Lucey, & Ellis 1992).

gression” technique is very insensitive to outliers. The histogram is not calculated in discrete bins, but using a kernel density estimator with a small gaussian kernel of $\sigma = 0.2$, comparable to the photometric uncertainty in the colors of individual galaxies. The results depend slightly on the smoothing parameter used but these effects are small compared to the intrinsic uncertainty caused by the limited number of galaxies in the sample.

These uncertainties were calculated by bootstrap resampling the color-magnitude distributions 200 times, and repeating the fitting procedures. We selected the central 68% of the best-fit slopes as our confidence interval.

Figure 2 shows the values of the slope $\delta(2200 - V)/\delta M_V$ versus redshift in fields of the HDFS and MS1054. We have included an additional low-redshift bin centered at $z \sim 0.5$, also determined from our data. To compare to observations at $z = 0$, we derived the $2200 - V$ slope from the $U - V$ slope under three different assumptions:

In the first case we assume the slope to result from a metallicity-luminosity sequence identical to that of nearby early-type galaxies, where more luminous galaxies are more metal-rich. We used observations of Coma cluster galaxies (Bower, Lucey, & Ellis 1992) and converted the observed slope in $U - V$ colors to $2200 - V$ colors with the help of Bruzual & Charlot (2003) models. These models were constructed to match observed colors of the cluster galaxies on the red CM-relation,

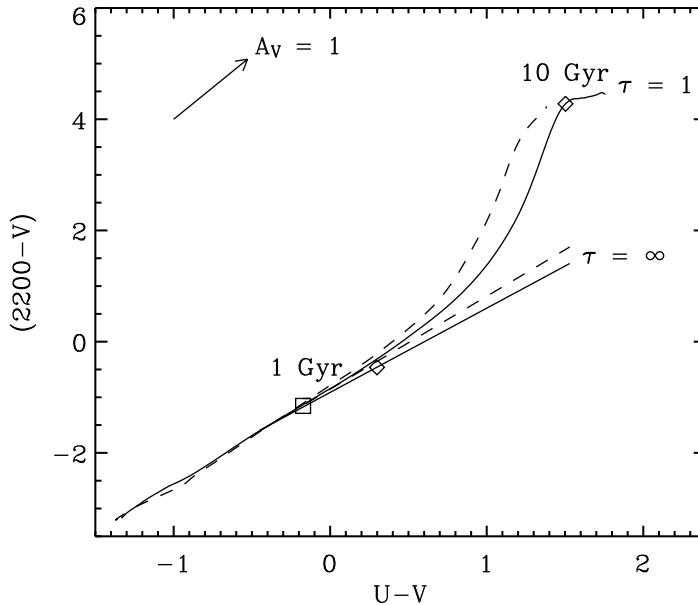


Figure 3 — The relation between $2200 - V$ and $U - V$ colors for Bruzual & Charlot(2003) stellar population models with declining star formation rates and a range of timescales τ . Solar metallicity (*solid lines*) and 1/3 solar models (*dashed lines*) are shown. The square indicates stellar ages of ~ 1 Gyr, the diamonds indicate ages of ~ 10 Gyr. Over a range of stellar population ages and metallicities the relationship is tight for blue galaxies, allowing a fairly accurate transformation of their colors. Also drawn is a Calzetti et al. (2000) attenuation vector of $A_V = 1$. In the presence of dust, the transformation of $U - V$ to $2200 - V$ colors is done with stellar tracks that already include reddening.

assuming the expected star formation history, i.e., high formation redshift and passive evolution.

The other two are derived from a linear fit to the $U - V$ colors of nearby blue-sequence galaxies from the Nearby Field Galaxy Survey (NGFS; Jansen et al. 2000a). Here we synthesized the $2200 - V$ slope under two assumptions. First, we assumed the CMR to reflect a systematic age trend, where higher-luminosity galaxies have higher stellar ages, and we used Bruzual & Charlot (2003) models to transform $U - V$ colors to $2200 - V$ colors (see Figure 3) For galaxies with blue colors, the models give $\Delta(2200 - V) \approx 1.5\Delta(U - V)$ for a range of metallicities and stellar ages. Second, we assume the slope is the result of increasing dust opacity with luminosity. Adopting the Calzetti et al. (2000) dust law yields $\Delta(2200 - V) \approx 2.3\Delta(U - V)$. An SMC extinction law (Gordon et al. 2003) would yield similar values.

Fig. 2 shows that the slope of the CM relation does not depend on redshift.

Only if we use the $z = 0$ slope derived for very red early-type galaxies in Coma is there any hint of evolution. This model is rather extreme, however, and will not be considered further. The measured evolution of the slope is $0.01z \pm 0.03$, consistent with zero. The error-weighted mean value of all FIRES measurements is

$$\delta(2200 - V)/\delta M_V = -0.17 \pm 0.021$$

If the slope were due to an age gradient as a function of magnitude, one might expect the slope to steepen with redshift, but we see no significant effect in the data presented here. We analyze the cause of the relation later in §4.3.

We note that Barmby et al. (2004) found a linear CMR for faint galaxies, and an upturn of the CMR at bright magnitudes. Figure 1 shows some evidence for an upturn at the bright end, but obviously our sample is too small to describe this properly, and we hence focus entirely on the linear part of the relation.

4.2 The Zeropoint and its Evolution

We determine the zeropoint by assuming that the slope of the CMR does not evolve and remains at the mean value of $\delta(2200 - V)/\delta M_V = -0.17$. We subtract the slope so that all galaxy colors are normalized to the color at $M_V = -21$. The histogram of normalized colors was determined as described before, and the zeropoint was determined from the location of the peak.

The evolution of the CMR zeropoint at $M_V = -21$ with redshift is shown in figure 4. The errorbars reflect the central 68% confidence interval of the zeropoint, obtained with the bootstrapping technique. We also present the CMR zeropoint of a local galaxy sample from the Nearby Field Galaxy Survey (Jansen, Fabricant, Franx, & Caldwell 2000). This value was derived from a fit to the $U - V$ colors of blue-sequence galaxies. We used tracks in Figure 3 to transform from $U - V$ colors to $2200 - V$, applying a small correction for a dust reddening of $E(B - V) = 0.15$. The color excess was obtained from the spectra of the galaxies (see §4.3), and adopting a Calzetti et al. (2000) dust law.

Clearly, the color of the blue CMR at fixed absolute magnitude reddens monotonically from $z \sim 3$ to $z \sim 0.5$. The galaxies become redder in $2200 - V$ by ≈ 1.2 mag from $z \sim 2.7$ to $z \sim 0.5$ (from the FIRES data alone), or by ≈ 1.45 from $z \sim 2.7$ to $z = 0$. Surprisingly a straight line describes the points very well, and we find

$$2200 - V = 0.24 \pm 0.06 - (0.46 \pm 0.03)z$$

The colors in the independent fields of the HDFS and MS1054 agree very well, and if extrapolated to $z = 0$, are consistent with the $z = 0$ point from the NFGS.

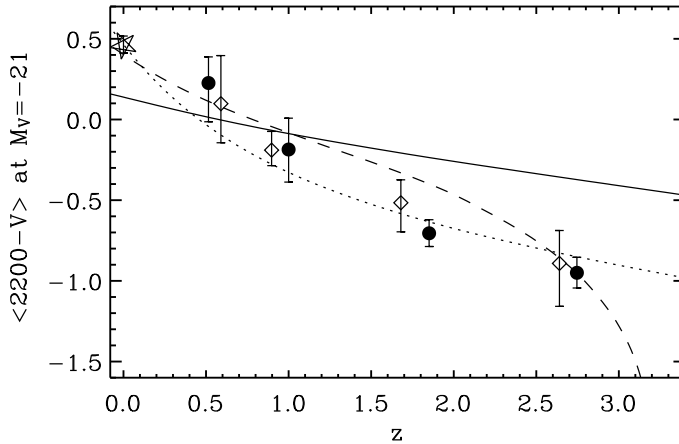


Figure 4 — The intercepts of fits to the blue CMR at fixed $M_V = -21$, marking the color evolution of the blue sequence as a function of redshift. We show the results in the field of the HDFs field (*filled circles*) and in the field of MS1054 (*diamonds*). The errorbars correspond to the 68% confidence interval derived from bootstrap resampling. The star indicates the $z = 0$ relation from the NFGS (Jansen, Franx, & Fabricant 2000a). The lines represent tracks of Bruzual & Charlot (2003) stellar population models. We show a model with formation redshift $z_f = 3.2$, a star formation timescale $\tau = 10$ Gyr, and fixed reddening of $E(B - V) = 0.15$ (*dashed line*); one with $z_f = 10$, constant star formation, and fixed $E(B - V) = 0.15$ (*solid line*); a model with $z_f = 10$, $\tau = 10$ Gyr, and $E(B - V)$ evolving linearly in time from 0 at $z = 10$ to 0.15 at $z = 0$ (*dotted line*).

This is encouraging considering that absolute calibration between such different surveys is difficult, and that the NFGS data has been transformed to $2200 - V$ colors from other passbands.

Next we compare the observed color evolution to predictions from simple stellar populations models. We assume that the galaxies remain on the ridge of the CMR throughout their life. This assumption may very well be wrong, but more extensive modeling is beyond the scope of this paper. Obviously for such simple models, the galaxies have all the same color, and these follow directly from the colors of the stellar population model, depending on star formation history, and the dust absorption only.

We calculate rest-frame colors using Bruzual & Charlot (2003) stellar population synthesis models with solar metallicity and exponentially declining star formation rates. We added dust reddening to the model colors using a Calzetti et al. (2000) dust law. To generate colors at constant absolute magnitude we apply a small color correction to account for V -band luminosity evolution of the stellar population. It reflects the fact that galaxies that were brighter in the past populated a different, redder part of the CMR. Similarly, when we allow varying levels

of dust attenuation, we apply a correction to account for dimming of the V -band light. We use the measured slope to apply the correction, the total amplitude of the effect is less than $\Delta(2200 - V) \lesssim 0.1$ for most models.

The simple model tracks are shown in Figure 4. A model with constant star formation and formation redshift $z_f = 10$ fits remarkably bad: the evolution is much slower than the observed evolution. When we explore models with exponentially declining star formation rates, we find that a decline time scale $\tau = 10 \text{ Gyr}$ and formation redshift $z_f = 3.2$ fits best. We note that we restricted the fit to $z_f \geq 3.2$ as the color-magnitude relation is already in place at redshifts lower than that.

Naturally, the models can be made to fit perfectly by allowing variable reddening by dust. As an example we show a $\tau = 10 \text{ Gyr}$ model with formation redshift $z_f = 10$ with reddening evolving linearly with time from $E(B - V) = 0$ at $z_f = 10$ to $E(B - V) = 0.15$ at $z = 0$.

We conclude that the relatively strong color evolution in the interval $0 < z < 3$ is likely caused by both aging of the stellar population and increasing levels of dust attenuation with time.

4.3 The Origin of the Blue Sequence in the Local Universe

It is impossible to determine the cause of the color-magnitude relation of blue galaxies from broad band photometry alone. Fits of models to the photometry produce age and dust estimates which are very uncertain (e.g., Shapley et al. 2001, Papovich et al. 2001, Förster-Schreiber et al. 2004). Spectroscopy is needed for more direct estimates of the reddening and ages. Unfortunately, the restframe optical spectroscopy of distant galaxies is generally not deep enough for the detection of $H\beta$, and the balmer decrement cannot be determined to high enough accuracy. Hence we can only analyze the relation for nearby galaxies which have spectroscopy with high signal-to-noise ratio.

We use the local sample of galaxies in the Nearby Field Galaxy Survey (Jansen, Fabricant, Franx, & Caldwell 2000; Jansen, Franx, Fabricant, & Caldwell 2000), a spectrophotometric survey of a subsample of 196 galaxies in the CfA redshift survey (Huchra et al 1993). It was carefully selected from ~ 2400 galaxies to closely match the distributions of morphology and magnitude of the nearby galaxy population, and covers a large range of magnitudes $-15 < M_B < -23$. The strength of the NFGS is the simultaneous availability of integrated broadband photometry, and integrated spectrophotometry of all galaxies, including line fluxes and equivalent widths of $[OII]$ H_α , and H_β . Integrated spectra and photometry are essential to enable a fair comparison to the integrated photometry of our high-redshift galaxy sample.

The trends of $U - V$ colors, dust absorption, metallicity, and $H\alpha$ equivalent width for normal nearby galaxies are illustrated in Figure 5. The distribution

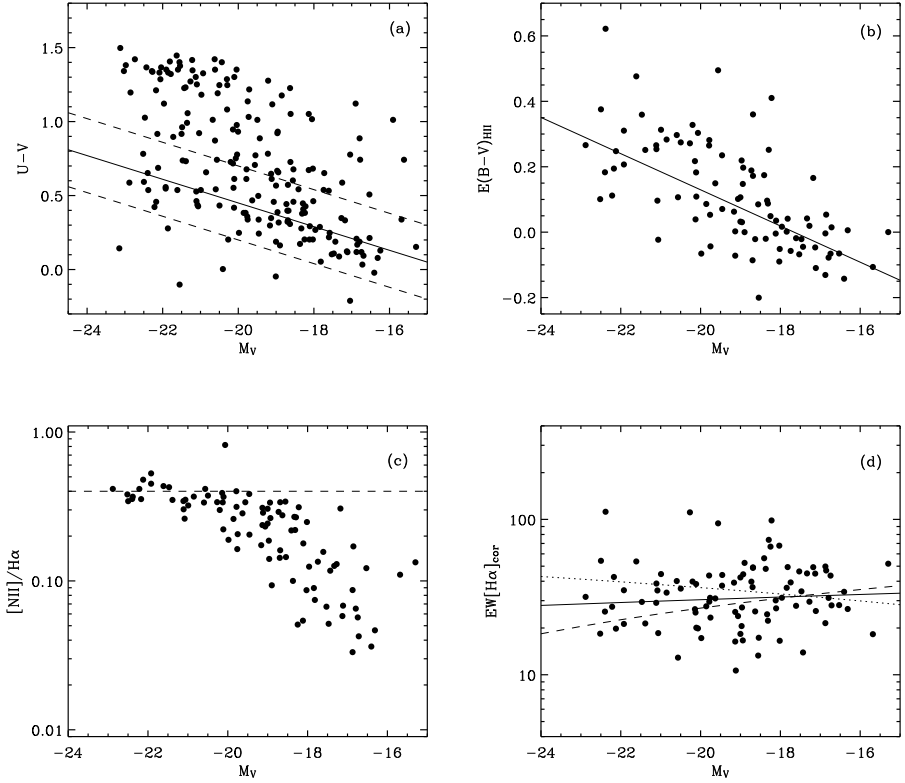


Figure 5 — $U - V$ colors and nebular emission lines properties versus absolute V -band magnitude of nearby normal galaxies from the Nearby Field Galaxy Survey (Jansen et al. 2000a). (a) The $U - V$ color versus absolute V magnitude. A linear fit to the blue sequence is shown (*solid line*). In the other panels we plot only the galaxies that are within 0.25 mag of the blue color-magnitude relation (*dashed lines*). (b) The $E(B - V)_{HII}$ color excess versus absolute V -band magnitude, derived from the observed ratio of integrated fluxes of $H\alpha$ and $H\beta$. The fluxes are corrected for Balmer absorption and Galactic reddening. The line shows a linear fit to the data. (d) The metallicity sensitive $[N II]\lambda 6584/H\alpha$ ratio. The dashed line indicates Solar metallicity. (c) The $H\alpha$ equivalent width $EW[H\alpha]$, corrected for Balmer absorption and attenuation by dust. The correction for dust absorption was obtained from the measured $E(B - V)_{HII}$ and assuming an absorption of the stellar continuum $E(B - V)_{cont} = rE(B - V)_{HII}$, where $r = 0.7$. The solid line is a linear fit to the data. Also shown are linear fits to the data in the case that $r = 1$ (*dashed line*) and $r = 0.2$ (*dotted line*).

$U - V$ colors versus absolute V magnitudes is shown in (a). A blue sequence and a red sequence of galaxies are visible. A linear fit to the blue sequence using the same technique as described in §4.1 gives a slope of $\delta(U - V)/\delta M_V = -0.08$. We isolate the blue galaxies and only consider the spectral properties of galaxies with colors $\Delta(U - V) < 0.25$ from the linear fit.

In (b) we show the reddening $E(B - V)_{HII}$ towards HII regions versus absolute V -band magnitude. The reddening was computed by Jansen et al. (2000) from the observed Balmer decrement $H\alpha/H\beta$, assuming the intrinsic ratio of 2.85 (case B recombination). As noted by Jansen et al., the sample shows a clear correlation; more luminous galaxies tend to have higher dust opacities.

The slope of the reddening-luminosity relation is $\delta E(B - V)_{HII} / \delta M_V = 0.06$. This reddening estimate applies to the H II regions in the galaxies, and it is generally thought that the mean reddening towards the stars is lower by a factor r : $E(B - V)_{cont} = rE(B - V)_{HII}$ with $0.5 \leq r \leq 1$ (Kennicutt 1998; Calzetti, Kinney, & Storchi-Bergmann 1996; Erb et al. 2003). For a Calzetti et al. (2000) dust law, the implied range of $U - V$ slope is 0.06 - 0.11, close to the observed value. Other forms of the extinction law, appropriate for the Milky Way (MW; Allen 1976) or the Small Magellanic Cloud (SMC; Gordon et al. 2003), yield similar values.

In (b) we show the relation between metallicity and absolute magnitude, as traced by nebular emission from H II regions. The correlation of the $[N II]\lambda 6584/H\alpha$ line ratio with integrated M_V shows that brighter galaxies are more metal-rich. To estimate the implications of this relation on the broadband colors, we explore two possibilities. Firstly, we use BC03 models to calculate the expected $U - V$ color variations for stellar populations of a wide variety of metallicities. We explored a grid of models with a range of exponentially declining star formation rates ($\tau = 0 - \infty$) and ages ($t = 0 - 13$ Gyr). We find that, regardless of age, models with blue colors $U - V < 0.5$ show variations with metallicities of $\Delta(U - V) \lesssim 0.1$, very small compared to the observed trend.

Finally, we show in (d) the relation between the $H\alpha$ equivalent width and absolute magnitude, after correction for Balmer absorption and dust absorption. The $EW[H\alpha]$ measures the instantaneous star formation rate per unit optical luminosity, and can be interpreted as the ratio of present to past averaged star formation rate or the time since the onset of star formation.

The correlation of $EW[H\alpha]$ with M_V depends on the value of r used for the correction for dust extinction. It is straightforward to see that the extinction correction for $\log EW[H\alpha]$ is proportional to $(1 - r)$: if $r = 1$, the H II regions and continuum are equally extinguished, whereas for lower values of r , the $H\alpha$ is more extinguished than the continuum.

We illustrate the effect in (d). The filled circles and the solid line show the data and a linear fit after correction with $r = 0.7$. There is no relation with absolute magnitude, implying that all the galaxies have the same age. The dashed line shows a linear fit to the data in the case of equal absorption $r = 1$. Now more luminous galaxies have slightly higher mean stellar ages. The dotted line shows $r = 0.2$, leading to a decreasing ages with increasing luminosity.

We can use the derived slopes to calculate the dependence of age and reddening as a function of magnitude. If metallicity variations are ignored, an exact solution

can be derived: The reddening of the H II regions follows immediately from the data, r and the slope of log age versus magnitude follow directly from the observed slopes of $U - V$ versus magnitude and $\text{EW}[\text{H}\alpha]$ versus magnitude. The solution is

$$r = 0.78$$

and

$$\delta \log(\text{age}) / \delta M_V = 0.027$$

This age gradient is very small, and produces a very shallow color magnitude slope of 0.01, opposite to the observed slope. The main cause of the color magnitude relation is the variation of dust reddening with magnitude. These results does not change significantly if metallicity contributes -0.01 to the color magnitude relation.

We conclude that the data on local galaxies imply that the CMR for blue galaxies is mostly caused by variations in reddening, with only small contributions from age and metallicity variations.

4.4 Comparison to $z \sim 3$ Galaxies

We turn to our broadband observations at $z \sim 3$ and compare the situation in the local universe to that at high redshift. We select the 186 galaxies in the FIRES sample at $2.2 < z < 3.2$, and inspect the properties of the galaxies that lie within $\Delta(2200 - V) < 1$ mag of the blue CMR.

4.4.1 The Models

Following previous studies of high redshift galaxies (Papovich, Dickinson, & Ferguson 2001; Shapley et al. 2001; Förster Schreiber et al. 2004), we fit stellar population models to the broadband SEDs of individual galaxies, and interpret the distribution of best-fit ages and extinctions. We use the publicly available HYPERZ fitting code (Bolzonella, Miralles, & Pello 2000), updated with the synthetic template spectra from the latest version of the Bruzual & Charlot (2003) stellar population synthesis code. We use the Basel 3.1 library (Westera et al. 2002) of theoretical stellar spectra, selected the Padova1994 stellar evolutionary tracks, and adopted a Salpeter IMF with upper and lower mass cut-offs of 0.1 and 100 $M_\odot$.

For simplicity, we only discuss the results for models with constant star formation rates and solar metallicity. Metallicities of blue, star-forming galaxies $z \sim 2-3$ are not well constrained, and may be solar (Shapley et al. 2004) or somewhat lower (Pettini et al. 2001). We prefer solar metallicity models as they have been directly calibrated against empirical stellar spectra. We refer to earlier studies for detailed discussions how the model assumptions affect the distribution of best-fit parameters Papovich, Dickinson, & Ferguson (2001); Shapley et al. (2001). Finally, we

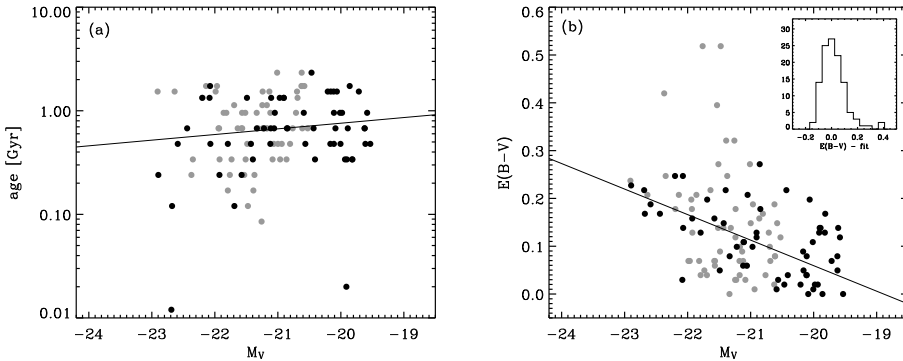


Figure 6 — The result of Bruzual & Charlot stellar population fits to the broadband SEDs FIRES fields at $2.2 < z < 3.2$. We show the distribution of best-fit stellar ages and extinctions as a function of absolute V -band magnitude for solar metallicity, constant star forming models, reddened with according to a Calzetti et al.(2000) dust law. Only the results for the galaxies within 1 mag in 2200- V of the blue CMR are drawn. (a) Best-fit ages and (b) best fit color excesses $E(B-V)$ versus absolute magnitude. The solid lines shows linear fits to the data. and the inset shows the distribution of residuals around the best fit relation for the color excess.

include reddening by dust by adopting the Calzetti et al. (2000) starburst attenuation law. Hence, the models are characterized by 3 free parameters: age (time since onset of star formation), star formation rate (SFR), and the level of extinction.

4.4.2 Results

We kept 152 out of a total 186 galaxies for which the models could find an acceptable fit, based on the value of χ^2 per degree of freedom. The galaxies that did not fit well, could have had SEDs that are not well-described by our (incomplete) template set, or suffer from contamination by emission lines. The distribution of colors and magnitudes of the rejected galaxies is similar to that of the galaxies with good fits, and their exclusion is not likely to affect the results.

In Fig 6(a) we show the between best-fit age against M_V . A weak trend is indicated, in the sense that the faint galaxies are slightly older than the bright ones. A least squares fit indicates a slope of $\log(\text{age})$ versus M_V 0.05 ± 0.06 , significant at the 0.8σ level.

In contrast, the reddening $E(B - V)$ shows a much stronger correlation with M_V in (c). We fitted a linear relation to the data finding a $\delta E(B - V)/\delta M_V = 0.05 \pm 0.01$. Remarkably, this is almost the same as the slope implied for the continuum reddening of nearby galaxies $\delta E(B - V)/\delta M_V = 0.047$ (see §4.3).

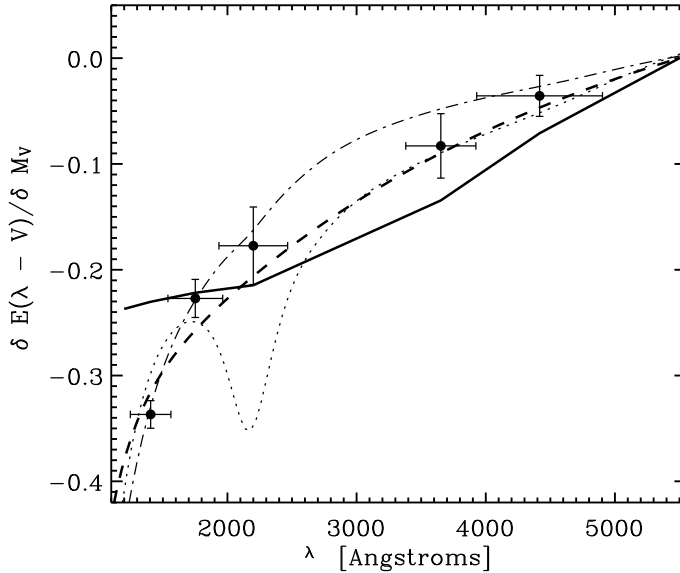


Figure 7 — The steepness of the CMR slope versus wavelength in the field of the HDFs. The data show the slope in the rest-frame $\lambda - V$ color versus M_V as a function of the filter λ . Overplotted are expectations for three extinction laws: the Calzetti et al. (2000) dust law (*dashed line*), the MW extinction law (Allen 1976; *dotted*), and the SMC extinction law (Gordon et al. (2003); *dash-dot line*). The normalization is different for each law: $\delta E(B - V)/\delta M_V = 0.04$ for the Calzetti, 0.05 for the MW, and 0.02 for the SMC. The thick solid line represents the color-dependence of the CMR slope in the case that stellar population age correlates with M_V . We show the track for a log-metallicity, constant star-forming model. The normalization is $\log(\text{age}) \propto -0.30 M_V$.

4.4.3 The Blue CMR in Various Rest-Frame Colors

As another way to show why the data imply dust as the main cause of the CMR slope, we derived the slope of the CMR for many different restframe filters, and show the results in Figure 7. We described our rest-frame luminosities and colors earlier in §2.2. We overplot the expected relation as a function of wavelength for several extinction curves. These curves were scaled with an arbitrary constant to provide the best fit. As we can see, the Calzetti curve provides the best fit, whereas the SMC and MW curve fit progressively worse. We also show the expected dependence if the CMR is caused by age variations. Again, the amplitude of this curve is fitted to the data points. It does not fit at all to the bluest point, which is the CMR slope in $1400 - V$ color versus M_V .

We conclude that, models fits to SEDs at $z = 3$ are in agreement with the picture in local universe. Reddening by dust explains the slope of the CMR consistently.

5 Constraints of the Color-Magnitude Relation on the Star Formation Histories of Blue Galaxies at $z \sim 3$

In this section we explore models to produce the narrow and asymmetric color distribution of galaxies along the CMR. We focus on the redshift range $2.2 < z < 3.2$, where the fraction of red galaxies is the lowest, and where perhaps the color distribution may be described by simpler models than at lower redshift. The distribution (Figure 1) is characterized by a blue peak and a skewness to much redder colors. The color distribution has a sharp cutoff at the blue side of the peak, and the scatter of the colors around the peak is quite low: most of the high redshift galaxies occupy a narrow locus in color space.

Here, our basic assumption is that the color scatter around the CMR is caused by age variations. We explore 4 different scenarios. In each scenario, we generate the complete star formation history of model galaxies, typically characterized by 2 or 3 free parameters. We then compile a large library of Monte-Carlo realizations, and generate the expected galaxy color distributions, taking into account observational errors and biases. We compare the model distributions to the observed color distribution to identify the best fitting parameterizations of the star formation histories. Kauffmann et al. (2003) used a similar method to constrain the star formation histories of local galaxies from the Sloan Digital Sky Survey.

A desirable property of this method, as will become apparent below, is that by using the information contained in the galaxy color distribution, we can resolve some of the degeneracies produced by fitting SEDs to the broadband colors of individual galaxies (e.g., Papovich, Dickinson, & Ferguson 2001), in particular the degeneracies in prior star formation history.

With this in mind, we will now discuss the model ingredients, the generation of the library, the fitting methodology, and the results for several different parameterizations of the star formation histories.

5.1 A library of Star Formation Histories

As the basic ingredient, we used the solar metallicity BC03 models as discussed in §4.4, and we fixed the BC03 model parameters except the star formation history: we only explore the effect of aging on the galaxy color distributions. We will compare the predictions of the models directly to our rest-frame luminosities and colors (see §2.2).

We generated a large library of Monte-Carlo realizations of stellar populations with different star formation histories. We classify the formation histories into 4 separate scenarios, characterized by distinct parameterizations of the star formation rates. They range from simple constant star formation, to complex models including bursts. Details of the parameterizations are given together with the results in §5.3. We do not introduce additional complexity, such as a distribution of metallicities, or distributions in the star formation rate parameters, as we do not

have sufficient observational constraints to justify more freedom in the models.

For each scenario, and for each SFH parameter combination, we generated between 100-200 realizations of individual star formation histories. For every instance we saved a selection of parameters, including colors, magnitudes, star formation rates, and masses. The final libraries of the scenarios contain on the order of $\sim 10^5$ unique star formation histories each.

5.2 Fitting Method

Briefly, we draw samples from the libraries, generate distributions of colors and magnitudes, and compare them to rest-frame luminosity and colors of FIRES galaxies in the redshift range $2.2 < z < 3.2$. We fit the models separately to the color distributions of the HDFs and MS1054 fields, as these have different depths and areas.

5.2.1 Creating Mock Observations

We need to account for three essential aspects of the data: the observations are magnitude limited, contain photometric errors, and have additional color scatter by variations in dust content.

It is evident that magnitude limits can substantially alter the resulting color distribution if galaxies in the models evolve strongly in luminosity. For example, an otherwise undetected galaxy undergoing a massive burst will temporarily increase in luminosity, and can enter a magnitude-limited sample, changing the color distribution. This effect is enhanced if there are many more galaxies below the magnitude limit than above. Hence, the steepness of the faint end slope of the luminosity function also plays a role.

We adopt a faint-end slope of $\alpha = -1.6$ according to the rest-frame far-UV luminosity function of Steidel et al. (1999). Shapley et al. (2001) found a steeper slope, but it resulted from a positive correlation of observed R and $R - K_s$ photometry, which we do not see in our data. We applied the luminosity function in the following way.

From the models we generate sets of galaxies with the same redshift distribution as the observed distribution. We then draw a luminosity from a luminosity function with $\alpha = -1.6$ and scale the model galaxy to that luminosity. Instead of using the instantaneous luminosity of the mock galaxy at the time of observation to compute the scaling, we use its median luminosity over the redshift range $2.2 < z < 3.2$.

Furthermore, we add photometric errors to the model colors as a function of model luminosity. The standard deviations of the errors were determined from a linear fit to the errors on the rest-frame luminosities and colors as a function of rest-frame M_V . Hence, these include the photometric redshift uncertainties. The fit gives a mean error in the rest-frame $2200 - V$ color of 0.15 in the HDFs and

0.19 for the MS1054.

Finally, we include reddening by dust in the models, adopting the Calzetti et al. (2000) starburst attenuation law. We add a distribution of color excesses

$$E(B - V) = 0.1 \pm 0.05(1\sigma)$$

to the model colors and magnitudes, where $E(B - V)$ is required to be greater than 0. The mean value is appropriate for $M_V = -21$ galaxies from the best-fit models in §4.3, and we used a scatter of 0.05 equal to that found in local galaxies (see §4.2) and similar to the distribution of $E(B - V)$ in §4.3. Using this distribution of extinctions, the scatter added to the model colors is $\sigma(2200 - V)_{dust} \sim 0.2$. Adopting an SMC extinction law would result in $\sigma(2200 - V)_{dust} \sim 0.15$. We remark that the extinction variations of the FIRES galaxies at $z = 2 - 3$ are well-constrained by the low scatter in the rest-frame far-UV colors.

Summing up, the average $(2200 - V)$ scatter introduced into the models by including both photometric uncertainties and dust variations is 0.25 mag for the HDFs, and 0.28 mag for the MS1054 field. Note that the width of the observed scatter at $2.2 < z < 3.2$, which we characterize by the central 32% of the color distribution, is 0.58 and 0.83 (for the HDFs and MS1054, respectively; see also Table 2).

Finally, to complete the mock color distributions, we imposed a magnitude limit of $M_V = -19.5$ for the HDFs field, and $M_V = -20.5$ for the field of MS1054. We are complete for all SED types down to these magnitude limits.

5.2.2 *The Fitting*

First, we subtracted the color-magnitude relation $\delta(2200 - V)/\delta M_V$ from the observed colors as in §4.2, and we normalize the $2200 - V$ color distribution to the color of the CMR intercept at $M_V = -21$. Then, for each scenario and for each field, we found the best-fit parameters by performing the two-sided Kolmogorov-Smirnov (KS) test on the unbinned color distributions of the models and the data. Next, we multiplied the KS-test probabilities of the individual fields, and we selected the parameter combination that yielded the highest probability.

A property of the KS-test is that it tends to be more sensitive around the median values of the distribution than at the extreme ends (Press et al. 1992). Given this behaviour, it is easy to see that our assumptions for the stellar population model, such as IMF and metallicity, may impose strong, yet undesirable, constraints on the fits. For example, a steeper IMF slope less rich in high-mass stars (e.g., Scalo instead of Salpeter) results in inherently redder spectra, whereas lower metallicity of the stellar populations leads to intrinsically bluer spectra. The KS-test would be sensitive to these systematic color changes in color, violation our basic assumption that the variations in color are caused by variations in age alone.

Furthermore, both the FIRES fields contain some very red galaxies (Franx et al. 2003; van Dokkum et al. 2003). These galaxies are thought to have much larger dust extinctions, higher ages, and perhaps different star formation histories than the $z \sim 3$ blue sequence galaxies we model here (see, e.g., van Dokkum et al. 2004; Förster Schreiber et al. 2004a). In principle, we do not expect our simple models to account for them, and while the KS-test is relatively insensitive to a small numbers of outliers, we might consider excluding them from the fit.

For these reasons we perform 3 different KS-tests, which we will refer to as KS1, KS2, and KS3. Firstly, we do a straightforward two-sided KS-test on the model and observed distributions. For the second, we normalize both the model and the observed color distribution to the median, and hence we only test for the shape of the color distribution, and not the absolute value of the color. Lastly, we compare the median normalized distributions after excluding galaxies that are $\Delta(2200 - V) > 1.75$ redder than the mode of the distributions, thereby removing the reddest outliers. The mode is calculated as in §4.1.

In the subsequent presentation of the results we focus on the second test (KS2), which is sensitive to the shape only, but we discuss the outcome of the other tests where appropriate.

We note in advance, that due to inherent differences between the fields the combined KS probabilities will never reach high values. For example, a two-sided KS-test comparing the MS1054 data to the observations of the HDFs (to a magnitude limit of $M_V = -20.5$) gives a probability of 70%. We do not expect any model that fits both data sets to exceed this probability.

5.3 Results

5.3.1 Constant Star Formation

In scenario 1, we assume galaxies start forming stars at random redshifts z_f . We take z_f to be distributed uniformly in time from $z = 2.2$ up to a certain maximum redshift z_{max} , where z_{max} is between $3.2 < z_{max} < 10$ in steps of 0.2. The star formation rate of each individual galaxy is constant for a certain time t_{sf} after which star formation ceases. We construct predictions for values of t_{sf} sampled logarithmically in 10 steps from 0.05 to 3 Gyr. Hence, this model is characterized by two parameters: z_{max} and t_{sf} .

In Figure 8(a) we show the color evolution as a function of time for a characteristic star formation history in this model, together with a schematic representation of the star formation rate. We also overplot the color intercept of the observed CMR in the HDFs and MS1054 fields at the median redshift of observation. In (b) we show the corresponding color-magnitude evolution. Figure 8 (c) shows the entire color-magnitude distribution of the best-fit models in the fields of the HDFs and MS1054. The observations are overplotted after subtracting the color-magnitude relation $-0.17(M_V + 21)$ (see also §4.1). In (d) we show the absolute

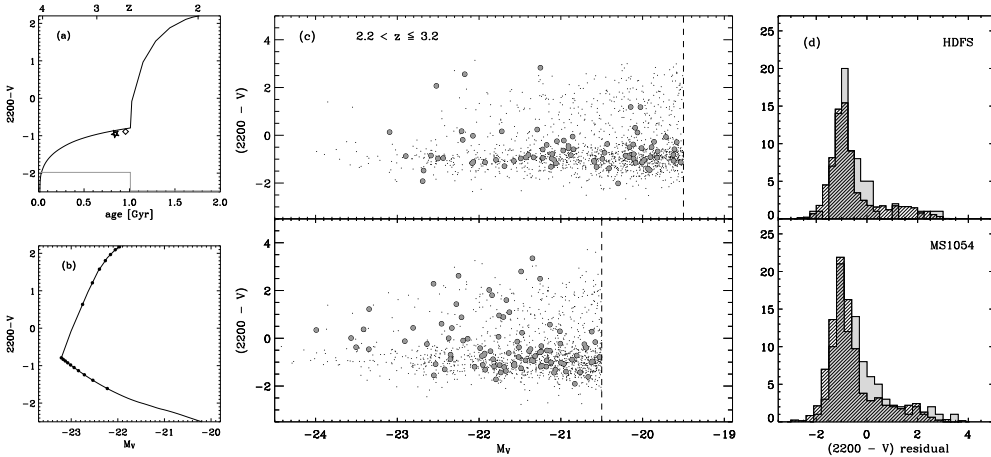


Figure 8 — The results of model 1. Galaxies in this model start off at random redshifts $z_0 < z_{max}$, and form stars at a constant rate for a duration t_{sf} before stopping. The maximum formation redshift z_{max} and t_{sf} are the free parameters. (a) The stellar population track of $2200 - V$ color against age for a characteristic galaxy ($z_0 = 4$, $t_{sf} = 1$ Gyr). The star formation rate is illustrated schematically by the gray line. Also drawn is the observed $2200 - V$ color of the blue CMR at fixed $M_V = -21$ in the HDF5 (*star*) and MS1054 (*diamond*). A mean dust reddening of $E(B - V) = 0.1 \approx E(2200 - V) = 0.4$ (Calzetti et al 2000) is added to the model color. (b) The track of $2200 - V$ color against absolute V -band magnitude for the same galaxy in steps of 100 Myr (*filled circles*) (c) The full color-magnitude diagram of the best-fit model (*black points*) in the redshift range $2.2 < z \leq 3.2$. The filled gray circles are the data from the HDF5 (*top*) and MS1054 (*bottom*). The color distribution of the model is broadened to account for photometric errors and scatter in the dust properties, where we used $\sigma(2200 - V)_{dust} = 0.2$. Only galaxies brighter than the absolute magnitude cut-off (*dashed line*) are included in the fit. (d) The histograms of residual $2200 - V$ colors show the data (*gray histograms*) and the best-fit model (*hatched histograms*). The best-fit parameters are $z_{max} = 4.6$ and $t_{sf} = 1$ Gyr.

color histograms of the best-fit model together with the data. We note that the “shape-sensitive” KS2 test was used to find the best fit.

The best-fit model in the KS2 test has a maximum formation redshift $z_{max} = 4.6$ and a constant star formation timescale $t_{sf} = 1$ Gyr. Some characteristics are reproduced, such as the narrow blue peak, but the asymmetric profile in the blue peak is not, although there is a low-level tail to red colors containing passively evolving galaxies. The fit is rather poor; the KS2-test assigns a 0.11 probability to the fit, the KS1 gives a maximum 0.01, while the KS3 gives 0.04.

In order to understand the behaviour of this model, we explored the z_{max} and t_{sf} parameter space. We found this model cannot produce the distinct skewness towards red colors in the blue peak, even if it cuts off star formation in a substantial fraction of the galaxies around the epoch of observation. This would be only mechanism to produce a red skew in this model, but the reddening of the passively

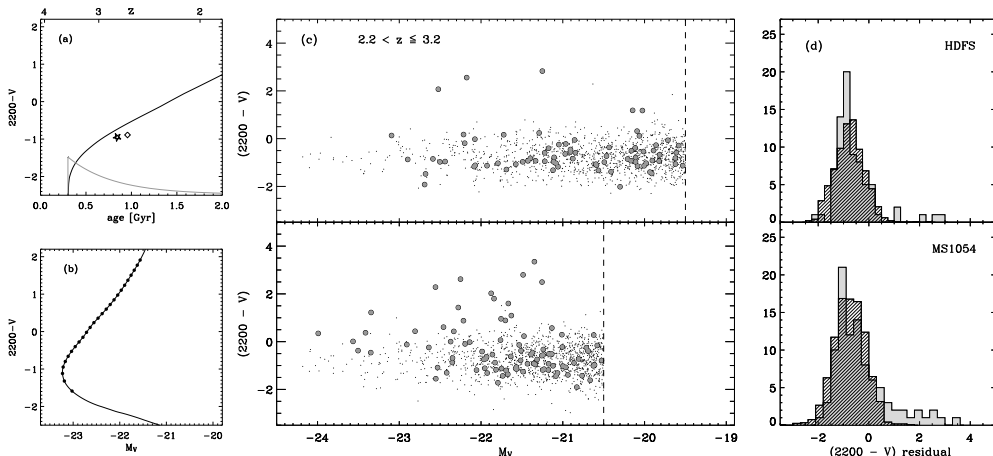


Figure 9 — Same as Figure 8 for model 2. Galaxies start forming stars at random redshifts $z_0 < z_{max}$, and form stars at an exponentially declining rate with e-folding time τ . The maximum formation redshift z_{max} and τ are the free parameters in this model. A characteristic galaxy shown in (a,b) has $z_0 = 3.5$ and $\tau = 0.5$ Gyr. The best-fit parameters of the in model (c,d) are $z_{max} = 4.2$ and $\tau = 0.5$ Gyr.

evolving galaxies is so rapid (see Fig. 8a) that instead of a red wing, it produces a prominent second red peak of passively evolving systems.

5.3.2 Exponentially Declining Star Formation

Scenario 2 is almost identical to the first, but now each individual galaxy has a single exponentially declining star formation rate with timescale τ . This model is characterized by two parameters: τ and z_{max} , where τ is sampled logarithmically in 10 steps from 0.05 to 3 Gyr.

Figure 9 shows a characteristic star formation history, and the fitting results of scenario 2. The KS2 test yields a best-fit model with a maximum formation redshift $z_{max} = 4.2$ and a constant star formation timescale $\tau = 0.5$ Gyr. It fits very poorly however, as the distribution is too broad and symmetric around the median. The KS2 test rules out this model at the 99% confidence level. The KS1 and KS3 test give similar answers.

Exploring the z_{max} and τ parameter space, we can understand why the fit is always poor. The only way for this model to produce a red skew in the color distribution, is by having a relatively small value for the star formation timescale τ compared to the mean age of the galaxies at $2.2 < z < 3.2$. In that case, a substantial number of galaxies is entering a “post-starburst” phase, where the

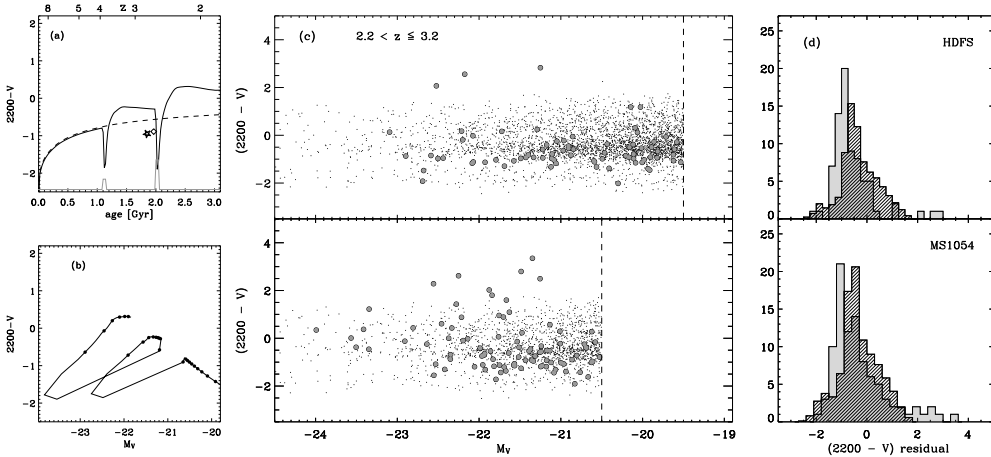


Figure 10 — Same as Figure 8 for model 3. Repeated burst models, with a multicomponent star formation history: an underlying constant star formation rate, and superimposed bursts of a certain strength $r = M_{burst}/M_{tot}$ and frequency n . The formation redshift is fixed to $z = 10$. A characteristic galaxy shown in (a,b) has $n = 1$ and $r = 1 \text{ Gyr}^{-1}$. The best-fit parameters of the model in (c,d) are $n = 0.3 \text{ Gyr}^{-1}$ and $r = 4$. The dashed line in (a) represents a constant star forming model (Bruzual & Charlot 2003).

mean stellar age is $t > \tau$ and the instantaneous SFR becomes much smaller than the past average. These galaxies gradually move away from the blue CMR, creating a skewness to red colors.

However, $\tau = 0.5 \text{ Gyr}$ models redden more quickly than CSF models at all ages, as can be seen comparing Fig 9a with Fig 8a. This results in somewhat redder mean colors, and a broader spread on the blue side of the peak, created by the newly formed galaxies that are continuously added to the sample.

5.3.3 Repeated Bursts

In scenario 3, all galaxies start forming stars at a fixed $z = 10$. The stars form in two modes: a mode of underlying constant star formation, and superimposed on this, random star bursts. The bursts are distributed uniformly in time with frequency n : the average number of bursts per Gyr. The amplitude of the burst is parameterized as the mass fraction $r = M_{burst}/M_{tot}$ where M_{burst} is the stellar mass formed in the burst and M_{tot} is the total mass formed by the constant star formation and any previous bursts combined. During a burst, stars form at a constant, elevated rate for a fixed time $t_{burst} = 100 \text{ Myr}$. Thus, there are two free parameters in this model: the burst frequency n , and the burst strength r . We sample n as $n^{1/2}$ from 0.1 to 6 Gyr^{-1} , and r as $r^{1/3}$ from 0.01 to 4.

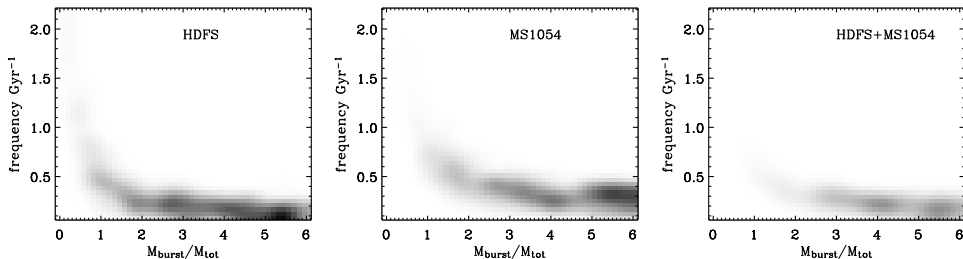


Figure 11 — The map of KS-probabilities versus model parameter repeated burst models of scenario 3. The free parameter were the burst frequency, and the strength $r = M_{burst}/M_{tot}$ of the bursts. The grayscale linearly encodes the probability of the fit. It shows the degeneracies of the model parameters, and regions of parameter space that are clearly excluded. The panels correspond to the model fits to the HDFS data, MS1054 data, and the data sets combined.

Figure 10 shows a characteristic star formation history and the best-fit results of scenario 3. The KS2 test yields a best-fit model with an average burst frequency of 0.3 Gyr^{-1} and a mass fraction formed in each burst of $r = M_{burst}/M_{tot} = 4$ ($= 400\%$): extremely massive, but relatively infrequent bursts. This solution reproduces the correct shape of the color distribution, i.e., the blue cut-off and red skew, but not the absolute colors. The median $2200 - V$ color is 0.4 mag too red, reflecting the $z = 10$ formation redshift. The KS2 test probability for this model is 0.35 , with the KS3 test giving a similar value. The KS1 test rejects the model at the 99% confidence level, as a result of the wrong median color.

We explore in Figure 11 the KS2-fit probability over the relevant part of n, r parameter space. In both fields, only models with infrequent massive bursts are allowed, although the exact strength is relatively unconstrained. There is notable difference between the HDFS and MS1054 fields. Specifically, the broader red wing in MS1054 observations compared to the HDFS favors a higher burst frequency.

As illustrated in Figure 10(a), the skewness towards red colors is produced by post-starburst galaxies. These galaxies have just formed large numbers of A-stars in the previous burst that outshine for some time the O- and B-stars formed in the underlying mode of constant star formation, hence biasing the integrated colors to the red. Interestingly, the color of a galaxy after a starburst always remain redder compared to constant star formation (or no bursts). Concluding, recurrent bursts do not “rejuvenate” the ensemble of galaxies. Rather, after a very brief blue period during the burst, the galaxy is redder forever after.

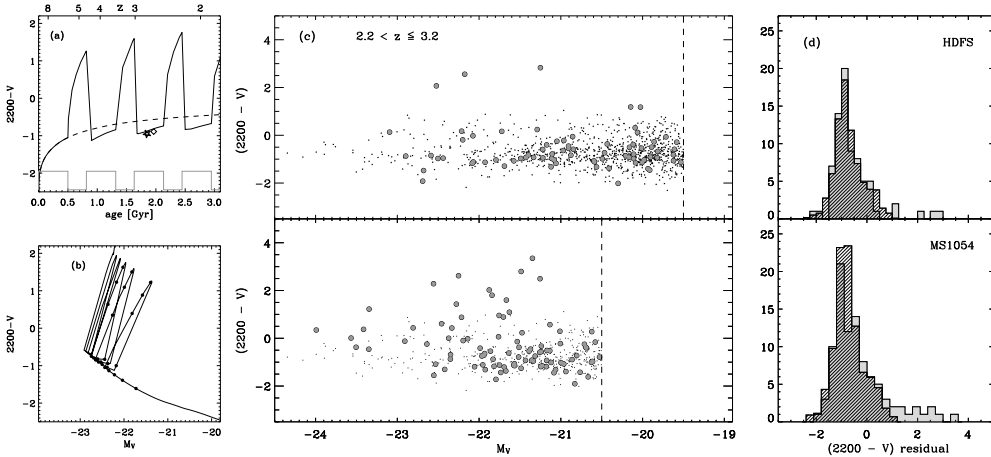


Figure 12 — Same as Figure 8 for model 4. Episodic star formation model with active periods of star formation and times of quiescence. The free parameter are the star formation duration t_{sf} , the fractional duration of quiescence r_q , and the remaining fraction of star formation r_{sfr} in the non-active period. The formation redshift is fixed to $z = 10$. A characteristic galaxy shown in (a,b) has $t_{sf} = 500$ Myr, $r_{tq} = 0.6 \text{ Gyr}^{-1}$, and $r_{sfr} = 0.02$. The best-fit parameters of the model in (c,d) are star formation duration $t_{sfr} = 200$ Myr, $r_{tq} = 0.4$, and $r_{sfr} = 0.02$. Note that the stellar population after resuming star formation is bluer than before it stopped.

5.3.4 Episodic Star Formation: The Duty Cycle

In the final scenario, galaxies again start at a fixed $z = 10$, and subsequently form stars at a constant rate for an “active” period of length t_{sf} . Then star formation suddenly drops to some fraction r_{sfr} of the nominal value, while the galaxy passes through a quiescent period that takes a fraction r_{tq} of the “active” episode $r_{tq} = t_q/t_{sf}$. This constitutes one duty cycle, and these star formation histories are characterized by repeating cycles of fixed length. We randomize only the phases of the cycles. There are three free parameters in this model: t_{sf} , r_{tq} , and r_{sfr} . We sampled t_{sf} logarithmically in 10 steps from 0.05 to 3.0 Gyr, r_{tq} logarithmically in 16 steps from 0.04 to 40, and r_{sfr} logarithmically in 10 steps from 0.01 to 0.8.

Figure 12 shows a characteristic star formation history and the fitting results of scenario 4. Using the KS2 test, the best-fit model galaxies have active periods of $t_{sf} = 200$ Myrs, and quiescent periods of 50 Myr ($r_{tq} = 0.4$), during which star formation drops to $r_{sfr} = 2\%$ of original rate. Figure 12(d) shows that the key features of the observed color distribution are correctly reproduced: the blue mean colors and the asymmetric shape of the observed distribution. The fit probability of the KS2 test is 0.58, and the KS1 test yields comparable probabilities, indicating that the model also produces the correct blue absolute colors. The KS3 test, which

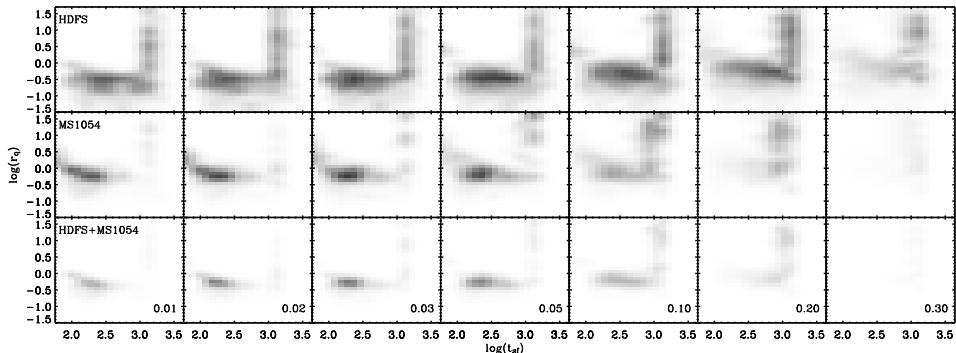


Figure 13 — The map of KS-probabilities versus model parameters in the duty-cycle models of scenario 4. The free parameters were the star formation duration t_{sf} , the fractional duration of quiescence r_q , and the remaining fraction of star formation r_{sfr} in the non-active period. The seven panels in each row correspond to 2-dimensional slices of the 3-dimensional parameter cube. Each slice is taken at a fixed r_{sfr} , which is indicated in the lower-right corner. The three rows correspond to the model fits to the HDFS data, MS1054 data, and the data sets combined. The grayscale encodes linearly the probability of the fit.

excludes the reddest galaxies, yields 0.89, reflecting that the mismatch with the model, but also between the two fields mutually, is mainly the enhanced number of very red galaxies in the field of MS1054 (see also Förster Schreiber et al. 2004). We discussed in §5.2 that our models are not necessarily expected to describe the small number of very red galaxies, and we conclude that this model provides the best description of the shape of the blue peak.

In Figure 13 we show the distribution of the (KS2) fit probabilities over the entire 3-dimensional parameter space. We present 2-dimensional slices of parameter space at steps in fraction of residual star formation rate r_{sfr} . The parameter space allowed by the two datasets are somewhat different. The scatter in the HDFS is intrinsically smaller, leading to smaller values of r_q and allowing longer star formation durations t_{sf} . Also noticeable is the “plume” of reasonably high probabilities in t_{sf}, r_{tq} . This region of parameter space resembles the model discussed in §5.3.1, where star formation is constant and stops exactly at the age of observation. The broader red wing of MS1054 clearly favors longer quiescent periods and short duty-cycles. It can also be seen that the length of the cycles is not well-constrained, and depends on the level of residual star formation rate r_{sfr} . If the intra-burst star formation rates are high, then the quiescent periods can be longer.

A striking aspect of the color evolution of a galaxy in this model is that when star formation resumes after the quiescent period, then the color of the population is at least as blue as, or even bluer than before quiescence (see Fig 12a). The high probabilities of the KS1 test already suggested that, in contrast to the bursting

model, galaxies with cycling star formation histories manage to maintain their extremely blue, colors despite the high formation redshift $z = 10$. The repeating “rejuvenating” of the colors leads to a substantial slower color evolution with time than a constant star forming model, as evidenced from the blue “ridge” of the color evolution track in Fig 12(a).

5.4 Discussion

The results presented here show that if we use the information contained in the observed color distribution of the galaxies, we can effectively constrain the simple models of the star formation history.

The first two scenarios (constant SFR and cut-off, and declining SFRs) were selected for their simplicity, and because these type of star formation histories are often assumed in SED modeling of high redshift galaxies (e.g., Shapley et al. 2001; Papovich, Dickinson, & Ferguson 2001; van Dokkum et al. 2004; Förster Schreiber et al. 2004a).

It is worrying that these scenarios generally fail to reproduce the observed color distribution. In addition, the parameter values for the best fit are not realistic. The maximum formation redshift is generally low $z = 4 - 4.5$, and the timescale of star formation (t_{sf} or τ) is comparable to the mean age of the galaxies at the $2.2 < z < 3.2$ redshift of observation. Hence, this is a special moment in the formation history, with many galaxies switching from active star formation to passive evolution or much lower SFRs. Such models predict profound evolution in the color distribution from $z = 4$ through $z = 2$, in contrast to the modest changes in the observed color distribution (Papovich, Dickinson, & Ferguson 2001; Papovich et al. 2004, this work).

The more complex models we considered are characterized by a fixed, high formation redshift ($z = 10$) and constant star formation, which is modulated by random events, such as starbursts.

The repeated burst model (scenario 3) with underlying constant star formation, does explain the shape of the observed color distribution, but we disfavor it for two reasons. Firstly, the adding of bursts does not rejuvenate the galaxy colors but leads to exceedingly red mean colors, mismatching the observations. We note that part of it can be resolved by tuning model assumptions, e.g., modifying metallicity, IMF, formation redshift, etc.

The second, more suspicious aspect is that the model needs to produce bursts with a high mass fraction, so that the longer-lived, but intrinsically fainter stars produced in the burst, outshine the luminous O- and B-stars created in the underlying mode of constant star formation. High mass fractions imply extreme instant star formation rates. Given a burst duration of 100 Myr, an $r = M_{burst}/M_{tot} \approx 4$ burst at $z = 3$ leads to a 60-fold increase of the instant star formation rate, transforming an ordinary L^* Lyman break galaxy into a monster with SFRs of 2000

$M_{\odot}\text{yr}^{-1}$. Such galaxies are not observed, unless they are also temporarily highly obscured, and show up at sub-mm wavelengths as “SCUBA” sources. Slightly longer burst durations do not alter this picture.

It is evident that the underlying constant star formation artificially introduced the need for massive bursts. Should it temporarily cease, then the bursts would not need to involve such high mass fractions to produce a red wing in the color distribution. In fact, if the underlying star formation temporarily ceased, one does not need bursts at all: hence the episodic model.

The episodic models (scenario 4) reproduced the color distribution best of the scenarios explored here. The best-fit parameters correspond to quiescent periods lasting 30-50% of the time of an active period, and typical duration of the total duty cycle of 150 Myr to 1 Gyr. The relative fraction of the quiescent period is better constrained than the length of the duty cycle, which correlates with the level of residual star formation in the passive periods.

The episodic star formation in these models rejuvenate the galaxies during each episode, making it significantly bluer than a galaxy with constant star formation of the same age. This also demonstrates the dependence of the derived ages on prior star formation history. If the broadband SEDs of our mock galaxies at $z \sim 3$ were fit with constant star forming stellar populations, then the best-fit ages would be ~ 500 Myr, or a “formation redshift” of the galaxies of $z_f \sim 4$, instead of the true $z_f = 10$.

The episodic model could be a solution to the enigma presented by studies of the broadband SEDs of $z \sim 3$ Lyman Break Galaxies (Papovich, Dickinson, & Ferguson 2001; Shapley et al. 2001). When the SEDs of LBGs were fit with constant or declining star forming models, the resulting best-fit ages (100-300 Myr) were much smaller than the cosmic time span of the $2 < z < 3.5$ observation epoch or the age of the universe (1 Gyr, and 2 Gyr, respectively). Papovich, Dickinson, & Ferguson (2001) already cautioned that constant or declining SFH are probably wrong for most galaxies, and suggested that episodic star formation could provide a mechanism to rejuvenate the appearance of the galaxies. Here we have formulated such a model, subjected it to a quantitative test, and constrained its parameters.

Even more constraints can be placed by taking into account the observed evolution of the color distribution as a function of redshift. The models described here may be tested directly by the data at lower redshift. However, one reason why we focused on the redshift range $2.2 < z < 3.2$ is that here the fraction of red galaxies is the lowest, and our simple models are perhaps appropriate. At lower redshift the color distribution is more complex with a prominent blue and red peak (see Figure 1), and such simple models might not apply.

6 The Onset of the Red galaxies

We could already see in Figure 1 that the color distribution in the FIRES fields evolves strongly from $z \sim 3$ to $z \sim 1$. This trend is particularly clear in the lightgray color histograms in (b) and (d). Most notably, at $z \sim 3$ most galaxies are on the blue sequence, and there is no evidence for a well-populated red peak. The onset of a red peak is tentatively observed in the histograms at $z \lesssim 2$, in particular in the field of MS1054. The red peak in the $z \sim 1$ bin of the MS1054 field contains a major contribution of the cluster of galaxies at $z = 0.83$. A prominent red sequence is also observed in photometrically selected samples in the field up to $z = 1$ (e.g., Bell et al. 2004b, Kodama et al. 2004).

We calculate the red galaxy fraction as a function of redshift using a simple color criterion to separate red from blue galaxies. Classically, color-limited fractions are defined relative to the red color-magnitude relation (Butcher & Oemler 1984), but such a definition is unusable here, as the red sequence is virtually absent at $z \sim 3$ in our sample. We therefore define red galaxies as all galaxies more than 1.5 magnitudes redder than the blue color-magnitude relation: $(2200 - V) + 0.17M_V > 1.5$. The threshold was set to be 2 – 3 times the scatter in $2200 - V$ color around the blue sequence. There is no evidence from the data that the scatter is a function of redshift (see Table 2).

Figure 14(a) shows the evolution of the fraction of red galaxies by number (N_{red}/N_{tot}). We also show the absolute luminosity density contributed by the red and blue galaxies separately in (b), and the fraction of total luminosity contributed by red galaxies (c). We calculated the fractions to a fixed rest-frame magnitude limit to which we are complete for all SED types at $z \sim 3$. The limits are $M_V = -19.5$ in the field of the HDFS, and $M_V = -20.5$ in the field of MS1054. The luminosity densities were computed by adding the luminosities of the galaxies above the magnitude limit, and dividing it by the cosmic volume of the redshift bin. The uncertainties in all estimates were obtained by bootstrapping the color-magnitude distributions.

In both fields, we find a sharp increase from $z \sim 3$ to $z \sim 1$ in the relative number, the absolute rest-frame V -band luminosity density, and the relative V -band luminosity density of red galaxies. At the same time V -band luminosity density in luminous blue-sequence galaxies remains constant, or decreases.

The differences between the fractions in the fields are substantial. In the lowest redshift bin ($z \sim 1$), part of it can be explained by the contribution of the cluster. At higher redshift however, small number statistics and the variations in space density of the red galaxies due to large scale structure are the probable cause (see, e.g., Daddi et al. 2003). The FIRES fields are still small; the total surveyed area is less than 30 arcmin^2 . We note that effect of the $2200 - V$ color scatter introduced by the limited accuracy of our photometric redshift technique, 0.32 in the median for the red galaxies, is minor. The dominant error is the limited number of galaxies in the sample, which is accounted for in the bootstrap calculation of

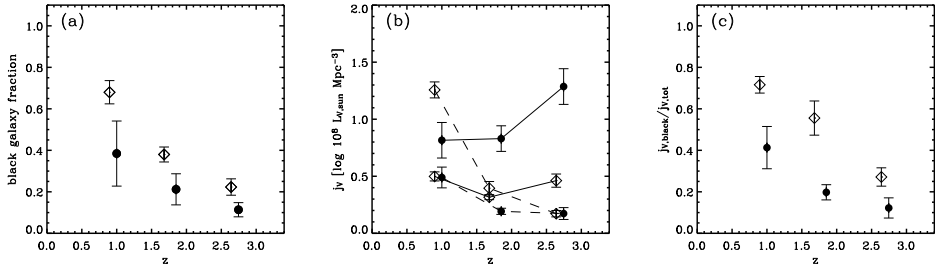


Figure 14 — Evolution of the fraction of red galaxies as a function of redshift. in the HDFs (filled circles) and MS1054 (diamonds). In (a) we show the number fraction of galaxies that are $(2200 - V) > 1.5$ mag redder than the blue sequence. (b) The absolute luminosity density in luminous blue (solid line) and red galaxies (dashed line). (c) The relative fraction of the total luminosity density in red galaxies. There is a sharp increase in the luminosity density in red galaxies between $z=3$ and $z=1$. We note that the field of MS1054 contains a massive cluster at $z = 0.83$.

the uncertainties.

The evolution of the red galaxies at high redshift appears to be different from the evolution between $z = 1$ and $z = 0$. Classic studies of the colors of galaxies at $z < 1$ found no evidence luminosity evolution in red galaxies, although the errorbars were fairly large (e.g., Lilly et al. 1995; Lin et al. 1999; Pozzetti et al. 2003). Recently, Bell et al.(2004b) found a constant luminosity density in photometrically selected red galaxies in the range $0.2 < z < 1.1$, and interpreted this as an increase in stellar mass in the early-type galaxy population, in apparent agreement with the hierarchical models of galaxy formation of Cole, Lacey, Baugh, & Frenk (2000).

We must defer such interpretations for our sample, until we better understand the nature of the red galaxies at $z > 1$. The foremost questions are whether all red galaxies are truly early-types with passively evolving stellar populations, or do other factors, such as reddening by dust play a role. Furthermore, when does the narrow red sequence establish it self?

In the local universe, $\sim 70\%$ of the galaxies in the narrow red sequence are morphologically early-type (Strateva et al. 2001; Hogg et al. 2002). The morphologies of $z \sim 0.7$ red peak galaxies in the GEMS survey (Rix et al. 2004), indicate that 85% of their rest-frame V-band luminosity density comes from visually classified E/S0/Sa (Bell et al. 2004a). In contrast, 70% of the Extremely red objects (EROs) at $1 < z < 2$ are believed to be dust reddened star-forming objects (e.g., Yan & Thompson 2003; Moustakas et al. 2004), while many of the $z \sim 3$ red galaxies in the FIRES sample also show signs of highly reddened star formation (van Dokkum et al. 2004; Förster Schreiber et al. 2004). Recent mid-IR imaging with IRAC on the Spitzer Space Telescope in the HDFs, indicates that only 25% of them are

dominated by passively evolving stellar populations (Labbé et al. 2004). Clearly, future study through spectra and MIR-imaging is necessary to give us a better understanding of the nature of the red colors

The next step in this kind of analysis would be to establish at what redshift the narrow red sequence establishes itself. Unfortunately, our photometric redshifts are too uncertain to allow an determination rest-frame colors with an accuracy better than 0.04 mag, typical of the scatter in the red color-magnitude relation (Bower, Lucey, & Ellis 1992). Hence spectroscopy is needed to establish the onset of the red color-magnitude relation.

7 Summary and Conclusions

We used deep near-infrared VLT/ISAAC imaging to study the rest-frame color-magnitude distribution of infrared selected galaxies in the redshift range $1 < z < 3$. We found a well-defined blue peak of star-forming galaxies at all redshifts. The blue galaxies populate a color-magnitude relation (CMR), such that more luminous galaxies in the rest-frame V -band tend to have redder ultraviolet-to-optical colors.

The slope of the CMR does not evolve with time, and is similar to the slope of blue, late-type galaxies in the local universe. Analysis of the spectra of nearby late-type galaxies from the NFGS suggests that the slope can be fully explained by the observed correlation of dust content with optical luminosity. The zeropoint of the blue peak at a given magnitude reddens smoothly from $z = 3$ to $z = 0$, likely reflecting an increase of the mean stellar age and an increase in the mean dust opacity of blue-sequence galaxies.

A key feature of the blue CMR relation is that the color distribution around it is asymmetric, with a blue “ridge” and a skew towards red colors. Assuming the scatter is caused by variations in mean stellar age, we have constructed models to explore the constraints that these observations place on the star formation history of blue field galaxies at $z = 2 - 3$. In the best-fitting models, galaxies form stars in short “duty-cycles”, characterized by alternating episodes of active star formation and quiescence. In these particular models, the best constrained parameter is the relative duration of the quiescent period, which is 30-50% of the length of an active period. The best-fit total length of the duty-cycle is uncertain, as it correlates with the amount of residual star formation during quiescence; the data allow a range of 150 Myr to 1 Gyr.

The models can explain the color distribution well, and, surprisingly, can also produce the very blue colors of $z = 2 - 3$ galaxies. The colors are bluer than the colors of galaxies with the same age with constant star formation, because the galaxies are rejuvenated by each burst after each quiescent period. These models can therefor solve the paradox that $z \sim 3$ galaxies appear much younger than the age of the universe when fit with constant star formation models (e.g., Papovich, Dickinson, & Ferguson 2001). The solution is simply that the luminosity weighted

ages of the stars derived from model fitting should not be interpreted as the time since the galaxy first started forming stars.

Finally, we find a sharp increase from $z \sim 3$ to $z \sim 1$ in the relative number, the absolute rest-frame V -band luminosity density, and the relative luminosity density of luminous red galaxies, while the V -band luminosity density in luminous blue-sequence galaxies is constant, or decreases. Studies at redshifts $z \leq 1$ imply very little evolution, suggesting that the bulk of the evolution takes place round $z \sim 2$. Obviously, our fields are very small, and the variations between the fields are large enough to suggest that the uncertainties in the red galaxy fraction are dominated by large scale structure (see e.g., Daddi et al. 2003).

On the other hand, the scatter and evolution of the blue galaxy sequence is similar between the fields, indicating that field-to-field variations do not play a large role there. Here the challenge is to confirm directly the cause of the relation between color and magnitude, and the cause of the skewness to red colors of the scatter around the relation. Very high signal-to-noise spectroscopy in the Near-IR will be needed for this purpose, to measure the balmer emission lines, in order to estimate ages and reddening. Furthermore, these studies need to be performed at higher redshifts. The advent of multi-object NIR spectrographs on 8-10m class telescopes will make such studies feasible in the near future.

Acknowledgments

We thank the staff at ESO for their dedicated work in taking these data and making them available. This research was supported by grants from the Netherlands Foundation for Research (NWO), the Leids Kerkhoven-Bosscha Fonds, and the Lorentz Center.

References

- Allen, D. A. 1976, MNRAS, 174, 29P
- Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, ApJ, 600, 681
- Bell, E. F., et al. 2004, ApJ, 600, L11
- Bell, E. F., et al. 2004, ApJ, 608, 752
- Bell, E. F. & de Jong, R. S. 2000, MNRAS, 312, 497
- Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
- Bessel, M. S. 1990, A&AS, 83, 357
- Blanton, M. R., et al. 2003, ApJ, 592, 819
- Bower, R. G., Lucey, J. R., & Ellis, R. S. 1992, MNRAS, 254, 601
- Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, A&A, 363, 476
- Bruzual A., G. & Charlot, S. 1993, ApJ, 405, 538
- Bruzual et Charlot et al. 2003 (BC03)
- Butcher, H. & Oemler, A. 1984, ApJ, 285, 426
- Casertano, S. et al., AJ, 120, pp. 2747–2824, 2000

- Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1996, *ApJ*, 458, 132
- Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., & Storchi-Bergmann, T. 2000, *ApJ*, 533, 682
- Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, *MNRAS*, 319, 168
- Chester, C. & Roberts, M. S. 1964, *AJ*, 69, 635
- Daddi, E., et al. 2003, *ApJ*, 588, 50
- Dreiling, L. A. & Bell, R. A. 1980, *ApJ*, 241, 736
- Efstathiou, G., et al. 2002, *MNRAS*, 330, L29
- Erb, D. K., Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., Hunt, M. P., Moorwood, A. F. M., & Cuby, J. 2003, *ApJ*, 591, 101
- Franx, M. et al., *The Messenger* **99**, pp. 20–22, 2000
- Franx, M. et al. 2003, *ApJ*, 587, L79
- Forster Schreiber et al. 2004, *ApJ*, accepted
- Forster Schreiber et al. 2004, in preparation
- Freedman, W. L., et al. 2001, *ApJ*, 553, 47
- Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, *ApJ*, 594, 279
- Griest, D. 1980, *AJ*, 85, 1295
- Hogg, D. W., et al. 2003, *ApJ*, 585, L5
- Jansen, R. A., Fabricant, D., Franx, M., & Caldwell, N. 2000, *ApJS*, 126, 331
- Jansen, R. A., Franx, M., Fabricant, D., & Caldwell, N. 2000, *ApJS*, 126, 271
- Jansen, R. A., Franx, M., & Fabricant, D. 2001, *ApJ*, 551, 825
- Katz, N. & Gunn, J. E. 1991, *ApJ*, 377, 365
- Kauffmann, G. & White, S. D. M. 1993, *MNRAS*, 261, 921
- Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, *MNRAS*, 264, 201
- Kauffmann, G., Charlot, S., & White, S. D. M. 1996, *MNRAS*, 283, L117
- Kauffmann, G., et al. 2003, *MNRAS*, 341, 33
- Kennicutt, R. C. 1998, *ApJ*, 498, 541
- Kodama, T. & Arimoto, N. 1997, *A&A*, 320, 41
- Labbé, I., et al. 2003, *AJ*, 125, 1107
- Labbé et al. 2004, in preparation
- Le Borgne, J.-F., et al. 2003, *A&A*, 402, 433
- Moorwood, A. F. 1997, *Proc. SPIE*, 2871, 1146
- Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, *ApJ*, 559, 620
- Papovich, C., et al. 2004, *ApJ*, 600, L111
- Peletier, R. F. & de Grijs, R. 1998, *MNRAS*, 300, L3
- Pettini, M., Shapley, A. E., Steidel, C. C., Cuby, J., Dickinson, M., Moorwood, A. F. M., Adelberger, K. L., & Giavalisco, M. 2001, *ApJ*, 554, 981
- Press et al. 1992, *Numerical Recipes in C*
- Rix, H., et al. 2004, *ApJS*, 152, 163
- Rudnick, G. et al. 2001, *AJ*, 122, 2205
- Rudnick, G., et al. 2003, *ApJ*, 599, 847
- Schweizer, F. & Seitzer, P. 1992, *AJ*, 104, 1039
- Shapley, A. E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., & Pettini, M. 2001, *ApJ*, 562, 95
- Shapley et al., 2004, *ApJ*, accepted
- Spergel, D. N., et al. 2003, *ApJS*, 148, 175
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, *MNRAS*, 328, 726
- Somerville, R. S. & Primack, J. R. 1999, *MNRAS*, 310, 1087

- Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, *AJ*, 112, 352
- Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, *ApJ*, 462, L17
- Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, *ApJ*, 519, 1
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., Pettini, M., Dickinson, M., & Giavalisco, M. 2003, *ApJ*, 592, 72
- Steinmetz, M. & Navarro, J. F. 2002, *New Astronomy*, 7, 155
- Tully, R. B., Mould, J. R., & Aaronson, M. 1982, *ApJ*, 257, 527
- Tully, R. B., Pierce, M. J., Huang, J., Saunders, W., Verheijen, M. A. W., & Witchalls, P. L. 1998, *AJ*, 115, 2264
- van Dokkum, P. G., Franx, M., Kelson, D. D., Illingworth, G. D., Fisher, D., & Fabricant, D. 1998, *ApJ*, 500, 714
- van Dokkum, P. G., Franx, M., Fabricant, D., Illingworth, G. D., & Kelson, D. D. 2000, *ApJ*, 541, 95
- van Dokkum, P. G. et al. 2003, *ApJ*, 587, L83
- van Dokkum, P. G., et al. 2004, *ApJ*, 611, 703
- Visvanathan, N. 1981, *A&A*, 100, L20
- Westera, P., Lejeune, T., Buser, R., Cuisinier, F., & Bruzual, G. 2002, *A&A*, 381, 524
- White, S. D. M. & Frenk, C. S. 1991, *ApJ*, 379, 52
- Zaritsky, D., Kennicutt, R. C., & Huchra, J. P. 1994, *ApJ*, 420, 87

Table 1 — The Blue Color Magnitude Relation

	HDFS		MS1054	
z^1	a^2	b^3	a^2	b^3
$0.5 - 0.7^4$	-0.176	0.226	-0.262	0.098
$0.7 - 1.4$	-0.206	-0.186	-0.144	-0.190
$1.4 - 2.2$	-0.140	-0.705	-0.234	-0.516
$2.2 - 3.2$	-0.139	-0.950	-0.324	-0.892

¹ The redshift range of the subsample

² The best-fit slope of the blue CMR.

³ The intercept of the blue CMR, after fitting *with a slope fixed at -0.17* .

⁴ The lower redshift bound was extended to $z=0.35$ for the HDFS, as the F300W filter is bluer than the classical U-band.

Table 2 — The Scatter around the Blue CM

	HDFS		MS1054	
z^1	σ_{obs}^2	σ_{true}^3	σ_{obs}^2	σ_{true}^3
$0.5 - 0.7^4$	0.53	0.46	0.79	0.73
$0.7 - 1.4$	0.66	0.61	0.72	0.65
$1.4 - 2.2$	0.37	0.27	0.73	0.66
$2.2 - 3.2$	0.57	0.51	0.83	0.77

¹ The redshift range of the subsample.

² The measured central 32% of the color distribution after rejecting galaxies with $\Delta(2200 - V) + 0.17M_V > 1.5$.

³ The intrinsic color distribution after subtracting the observational errors.

⁴ The lower redshift bound was extended to $z=0.35$ for the HDFS, as the F300W filter is bluer than the classical U-band.

IRAC Mid-Infrared Imaging of Red Galaxies at $z > 2$ new constraints on age, dust, and mass

ABSTRACT

We present deep 3.6 – 8 micron imaging with IRAC on the Spitzer Space Telescope of a population of galaxies with red rest-frame optical colors at $z > 2$. The 13 distant red galaxies (DRGs) were selected in the field of the Hubble Deep Field South on the simple color criterion $J_s - K_s > 2.3$ and we compare their properties to those of 23 Lyman Break Galaxies (LBGs or U-dropouts) at $z \sim 2.5$ in the same field. The new IRAC data reaches rest-frame NIR wavelengths, which are crucial in determining the nature of these galaxies. We are able to uniquely identify 3 out of 11 DRGs as old passively evolving systems at $z \sim 2.5$. The others are heavily reddened star-forming galaxies, for which we are now better able to distinguish between the effects of age and dust. Furthermore, the rest-frame NIR data allow more robust estimates of the stellar mass and stellar mass-to-light ratios (M/L_K). We find that in a mass-selected sample DRGs contribute $1.5 - 2\times$ as much as the LBGs to the cosmic stellar mass density at $2 < z < 3.5$. Also, at a given rest-frame K luminosity the red galaxies are twice as massive with average stellar masses $\sim 10^{11} M_\odot$, and their M/L_K mass-to-light ratios exhibit only 1/3 of the scatter compared to the U-dropouts. This is consistent with a picture where DRGs are more massive, more evolved, and have started forming at higher redshift than most LBGs. We find no evidence for a substantial AGN contribution to the observed optical-MIR SEDs.

Ivo Labbé, Jiasheng Huang, Marijn Franx, Gregory Rudnick, P. Barmby,
 Emanuele Daddi, Pieter G. van Dokkum, Giovanni G. Fazio, Natascha M. Förster
 Schreiber, Konrad Kuijken, Alan F. Moorwood, Hans-Walter Rix, Huub
 Röttgering, Ignacio Trujillo, Arjen van der Wel, Paul van der Werf, & Lottje van
 Starkenburg

1 Introduction

MAPPING the properties of massive galaxies as a function of redshift provides strong tests for theories of galaxy formation, as their build-up can be directly probed from high redshift to the present epoch. But while the theories describing the growth of large-scale dark-matter structure are thought to be well-constrained (Freedman et al. 2001; Efstathiou et al. 2002; Spergel et al. 2003), the formation history of the stars inside the dark-matter halos is still poorly understood. Direct observations are essential for progress on this front.

Until recently, it was difficult observe statistically meaningful samples of massive high-redshift galaxies. The best-studied samples are selected on the rest-frame UV light through the Lyman Break technique (LBGs; Steidel et al. 1996a,b, 2003; Madau et al. 1996; Giavalisco & Dickinson 2001), which yielded large numbers of relatively low mass, unobscured, star-forming galaxies at $z > 2$ (Papovich, Dickinson, & Ferguson 2001; Shapley et al. 2001). Recent advances in near-infrared (NIR) capabilities on large telescopes are now making it possible to access the rest-frame optical for large numbers of galaxies to $z \sim 3$. The rest-frame optical is already much less sensitive to dust obscuration and on-going star formation than the rest-frame UV, and is expected to be a better (yet imperfect) tracer of stellar mass.

In this context, we started the Faint Infrared Extragalactic Survey (FIRES; Franx et al. 2000): a deep optical-to-infrared multicolor survey of NIR-selected galaxies in two fields. In the deepest field, the Hubble Deep Field South (HDFS), we spent 102 hours of VLT/ISAAC imaging on one pointing in the J_s , H , and K_s -bands resulting in the deepest ground-based NIR imaging, and the deepest K -band imaging to date, even from space (Labbé et al. 2003). We selected high-redshift galaxies with the simple color criterion $J_s - K_s > 2.3$, designed to isolate galaxies at $2 < z < 4$ with a prominent Balmer- or 4000 Å break (see Franx et al. 2003). We find these Distant Red Galaxies (DRGs) at high surface densities ~ 3 arcmin $^{-1}$ to $K = 22.5$, with space densities about half of that of LBGs selected from ground-based imaging down to $R = 25.5$.

Previous studies using broadband optical-to-NIR SED fitting and optical/NIR spectroscopy have suggested that DRGs, at a given rest-frame optical luminosity, have higher ages, contain more dust, and are more massive than LBGs (Franx et al. 2003; van Dokkum et al. 2004; Förster Schreiber et al. 2004), and they may contribute comparably to the cosmic stellar mass density (Franx et al. 2003; Rudnick et al. 2003). However, the nature of their red colors is still poorly understood, and the masses are somewhat uncertain, raising many questions. Are they all truly old, or are some also very young and very dusty? What is the fraction of passively, evolving “dead” systems? How much do the DRGs contribute to the stellar mass density in a mass-selected sample? And how do they relate to the blue Lyman break galaxies. Finally, what is their role in the formation and evolution of massive galaxies?

To shed light on all of these questions, we need imaging in the rest-frame NIR, including the rest-frame K -band. Here, we present deep MIR imaging with IRAC on the Spitzer Space Telescope of a sample of distant red galaxies in the field of the HDFS. In combination with the wealth of existing deep imaging from HST/WFPC2 and VLT/ISAAC, the IRAC data promises to establish more robust stellar masses and mass-to-light ratios, and may help to reduce the degeneracies between age and dust in modeling of the broadband spectral energy distributions (SEDs) (see, e.g., Papovich, Dickinson, & Ferguson 2001; Shapley et al. 2001). Where necessary, we assume $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70 \text{ km s}^{-1}\text{Mpc}^{-1}$. The magnitudes are given on the Vega system.

2 The Observations, Photometry, and Sample Selection

We have observed the HDFS WFPC2 field with the IRAC camera (Fazio et al. 2004) on the Spitzer Space Telescope, integrating 1 hour each in the 3.6, 4.5, 5.8, and 8 micron bands. The full observations and data reduction are described in Labbé et al. (2004), but we give an abbreviated outline here.

The IRAC images were taken in the 4 broadband MIR filters at 3.6, 4.5, 5.8, and 8 microns over a $5' \times 5'$ field of view. The pixels are $\approx 1.2''$ in size. The observations in the field of the HDFS were taken on May 26 2004 (GTO program 214) and were split into dithered frames of 200s each, except for the 8μ band where the frame time was 50s.

We used the Basic Calibrated Data (BCD) as provided by the Spitzer Science Center pipeline, and refined the astrometry of the individual frames with 2MASS sources. We rejected cosmic rays, corrected for known artifacts such as column pulldown, and muxbleed, and accounted for the “first frame effect” by subtracting a median stacked image from the individual frames. Finally, we corrected the frames for geometric distortion, projected them on the existing ISAAC K -band image (Labbé et al. 2003), and average-combined them. We compared the result to a median combined image to make sure that we excluded all cosmic rays.

The pixel scale of the final images is $0''.36$ per pixel, about $1/3$ of the original IRAC scale and equal to $3\times$ the scale of the ISAAC image. The limiting depths at 3.6, 4.5, 5.8, and 8 micron are 22.2, 21.3, 18.9, and 18.2 mag, respectively (derived from the 5σ effective flux dispersion in 3 arcsec diameter apertures). The image quality ranges from 2 - 2.4 arcseconds FWHM and is best in the 4.5μ band. The positional accuracy with respect to sources in the ISAAC K -band is better than $0''.2$ across the field. From hereon, we only consider the deepest central part of the MIR imaging overlapping the $2.5' \times 2.5'$ ISAAC field.

To achieve consistent photometry across the MIR IRAC bands and the existing NIR catalog¹, we carefully matched the point-spread-function (PSF) of the K_s , 3.6, 4.5, and 5.8 band to the 8μ band, where the image quality was worst. We

¹NIR images and catalogs are available from <http://www.strw.leidenuniv.nl/~fires>

deconvolved a selection of bright stars in the 8 micron image with a selection of bright stars in the other maps, providing us with the required PSF-match kernels. We convolved the maps to the 8 micron PSF, and checked the quality of the match by dividing stellar growth curves, normalized to 1 at the IRAC zero-point radius of $R = 12.2$ arcseconds. The agreement is better than 3% at radii $R > 2.2$ arcseconds; hence we used a diameter of $D = 4''.4$ as our fixed aperture for photometry.

To reduce the effects of confusion, we used the deep ISAAC K -band image to model and subtract neighbouring sources. The image quality in the K -band is $\sim 0''.5$, so confusion is less of an issue. We proceeded by convolving each K -band source in isolation to the IRAC PSF, and fitted them simultaneously to each individual IRAC map, leaving only the fluxes as free parameters. We did not use the model fluxes directly; instead, we use the models to subtract the confusing neighbours next to the sources of interest. We then proceeded with normal aperture photometry in a $D = 4''.4$ diameter. We combined the IRAC fluxes with existing optical/NIR photometry in $D = 2''$ apertures from the catalog published by (Labbé et al. 2003). Thus to obtain consistent colors, we applied an aperture correction to the IRAC fluxes which was the ratio of the original K -band flux in the $D = 2''$ aperture and the K -band flux in the PSF-matched $D = 4''.4$ aperture. Finally, we assumed a minimum photometric error of 10% reflecting various calibration uncertainties. The end result is a fairly homogenous photometric catalog spanning the observed optical-to-MIR in 11 filters.

The primary sample of interest is comprised of the DRGs in the field of the HDFs which were selected on $J_s - K_s > 2.3$ and $K < 22.5$ (see Franx et al. 2003). The sample comprises 14 galaxies in the redshift range $1.9 < z < 3.8$, which were all detected at 3.6, and 4.5 micron. Two galaxies have only upper limits at 5.8 and 8.0 micron. One of these was excluded from further analysis as the deblending procedure was unsuccessful; the source was also confused in the K -band. The final sample contains 13 DRGs.

As a comparison sample, we selected Lyman Break Galaxies in the same field from the WFPC2 imaging (Casertano et al. 2000) to the same limit in K , using the criteria of Madau et al. (1996). We note that the photometric system of the WFPC2 is somewhat different from the one adopted for the selection of ground-based LBGs (Steidel et al. 2003). The blue $F300W$ bandpass causes Lyman break galaxies to enter the selection window already at redshifts $z \gtrsim 1.8$ (Giavalisco & Dickinson 2001). As a result the redshift distribution of space-based “U-dropouts” is a better match to that of the DRGs than the groundbased LBGs, or the more recent “BM/BX” objects at $z \sim 2$ (Steidel et al. 2004).

We kept the 37 LBGs that fell in the same redshift range as the DRGs ($1.9 < z < 3.8$). Because of their higher surface density, source confusion played a more important role. Conservatively, we selected only the most isolated, and hence least affected galaxies out of the 37, leaving a final sample of 23. We checked that the median K -band magnitude and the median $I - K$ color of this subsample was

the same as that of the full sample; hence it should be representative. Whenever calculating global averages, we simply scale the average properties of either sample to the original numbers (14 and 37).

To conclude, the median K -band magnitude of the DRGs is 21.4 and that of the space-based LBGs is 21.5. The redshift distributions are comparable, and both have a median $z_{phot} = 2.5$. Stars were excluded from the LBG sample using a method detailed in Labbé et al. (2003) and Rudnick et al. (2003). The galaxy identifications throughout this paper refer to the published catalog of Labbé et al. (2003), which includes photometric redshifts for all sources.

We determined the rest-frame luminosities and colors by direct interpolation in AB magnitudes between the observed filters. The rest-frame luminosities are sensitive to the uncertainties in the photometric redshifts. We used photometric redshifts based on an algorithm developed by Rudnick et al. (2001, 2003) and spectroscopic redshifts where available. The photometric redshifts are in good agreement with the spectroscopic redshifts, with an average $|z_{phot} - z_{spec}|/(1 + z_{spec}) = 0.05$ for sources at $z > 2$. We checked that the publicly available HYPERZ method (Bolzonella, Miralles, & Pello 2000) gave consistent answers, and we checked that adding the MIR did not change the photometric redshifts significantly. We decided not to use these new photometric redshifts for now, as the differences were small.

3 Mid-Infrared Properties of Red Galaxies at $z > 2$

One of the primary motivations for obtaining the 3.6 – 8 micron IRAC imaging is to extend the spectral energy distribution (SED) of the DRGs into the rest-frame near-infrared. The rest-frame NIR is essential to understand why these galaxies have colors that are much redder than Lyman Break Galaxies. Is it because they are much more obscured, or because their stellar populations are more evolved? The answer to this question is a crucial step in understanding their mutual relation. Some insight has already been obtained from modeling of the observed optical and near-NIR SED (Förster Schreiber et al. 2004), but it has not been possible to uniquely distinguish old, passively evolving galaxies from very dusty, actively star forming systems.

The rest-frame NIR is expected to separate very young and extremely reddened ($A_V > 3$) galaxies from older galaxies with much less extinction ($A_V < 1.5$), as in the first case the spectrum is red throughout the rest-frame UV and optical, and peaks somewhere around the rest-frame J -band, while in the other case the spectrum peaks in the rest-frame optical.

As a first step we can see if the nature of these objects can be elucidated by a specific combination of two colors alone. We inspect the observed $I - K$ versus $K - 4.5\mu$ color-color diagram of the galaxies (see Figure 1). At a redshift of $z=2.5$, these colors correspond to the rest-frame $2200 - V$ versus $V - J$ colors. We examine the three DRGs whose optical-to-NIR SEDs were fit significantly better

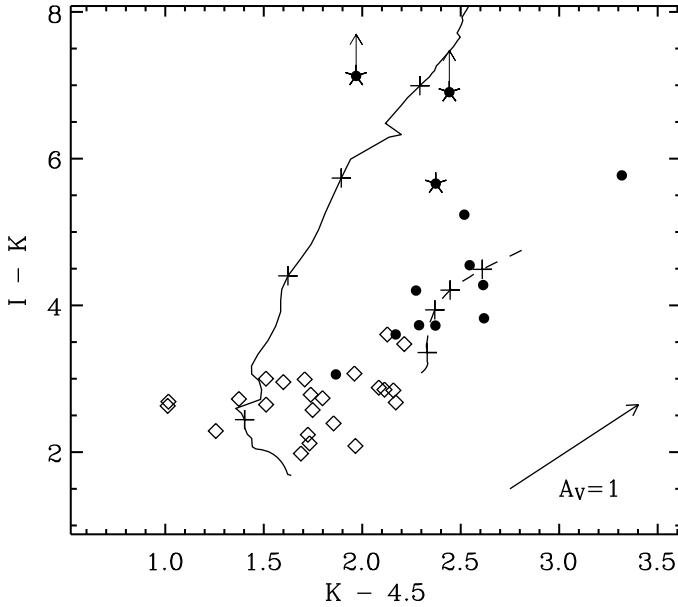


Figure 1 — $I - K$ versus $K - 4.5$ color-color diagram of two samples of $z > 2$ galaxies. Diamonds show Lyman Break galaxies. Filled circles show galaxies selected on the color-criterion $J - K > 2.3$ or Distant Red Galaxies (DRGs). Red galaxies whose optical-to-NIR SED were best fit with unreddened single age bursts are marked with a star. Also shown are time-evolution tracks of Bruzual & Charlot (2003) stellar population models, redshifted to $z = 2.5$. (solid line) A single age burst, (dashed line) a constant star forming model with a reddening of $A_V = 1.5$. The reddening vector assumes a Calzetti et al. (2000) extinction law.

with unreddened single age bursts than with dusty constant star formation models (indicated by stars in Figure 1); these are amongst the reddest in $I - K$ color. For these three galaxies the best fit single age models predict fairly blue $K - 4.5\mu$ colors. On the other hand, the worse fitting dusty constant star formation models for these galaxies predict them to be the reddest in $K - 4.5\mu$, which they are not. Had we only $K - 4.5\mu$ color, and hence lacked the apriori knowledge of the best-fit SFH, this result implies that we would still have been able to crudely distinguish old passively evolving galaxies from those that are heavily reddened and vigorously forming stars.

To look in more detail at the three candidate old “dead” galaxies we show in Figure 2 their broadband SEDs. Remarkably, single burst model predictions based on the optical and NIR data alone, are directly confirmed by the IRAC 3.6 – 8 micron observations. For these galaxies we can definitively rule out reddened

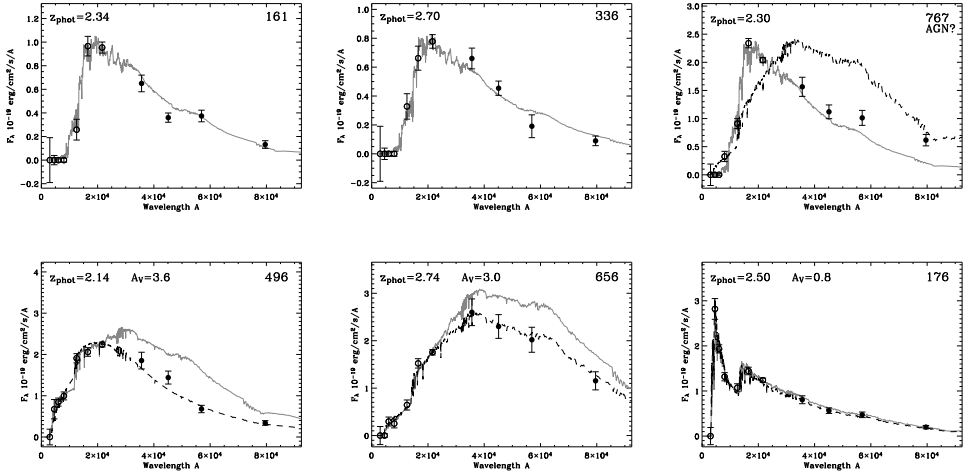


Figure 2 — Comparisons between the predicted IRAC fluxes from fits to the optical-to-NIR SED and the actual IRAC 3.6 – 8 micron observations. The predictions (*gray solid lines*) are based on Bruzual & Charlot (2003) stellar populations fitted to the optical-to-NIR SEDs (*open circles*). The top row shows the 3 DRGs which, excluding the MIR data, were best-fit with an unreddened single age burst. The observed MIR IRAC fluxes (*filled circles*) provide direct confirmation on their nature. Galaxy 767 shows a flux excess at 8 micron, which cannot be produced by constant star formation and reddening (*black dashed lines*); it is possibly related to AGN activity. The bottom row shows the MIR predictions for galaxies using constant star forming models with Calzetti et al. (2000) dust reddening. For heavily reddened galaxies, the MIR observations were often different from the optical-to-NIR based predictions, reflecting the degeneracy between age and dust in the models. For galaxies with lower dust content $A_V \lesssim 1$, IRAC confirms the predictions. Also drawn is the best-fit to the full SED (*black dashed lines*).

models with constant star formation. Before the MIR imaging, we could only marginally exclude these models from the formal χ^2 of the fit (Franx et al. 2003). We also tried less extreme models with declining star formation rates ($\tau = 300 - 500$ Myr) and dust reddening (Calzetti et al. 2000); we find that such models fit significantly worse than the single age bursts.

One of the galaxies clearly has excess flux at 8 micron. Complex stellar populations with, for example, heavily reddened star formation cannot account for such excess, because they also generate $5.8\mu - 8\mu$ colors that are too blue. It may be related to an obscured active galactic nucleus (AGN).

Also in Figure 2 we show the MIR predictions for galaxies whose optical-to-NIR fluxes were better fit with constant star forming models and reddening. In general the observed MIR flux points do not confirm the predictions, reflecting the well-known degeneracy between age and dust in the models which allows a wide variety of predicted values (see, e.g., Papovich, Dickinson, & Ferguson 2001). This effect is particularly severe for heavily reddened galaxies that are red throughout

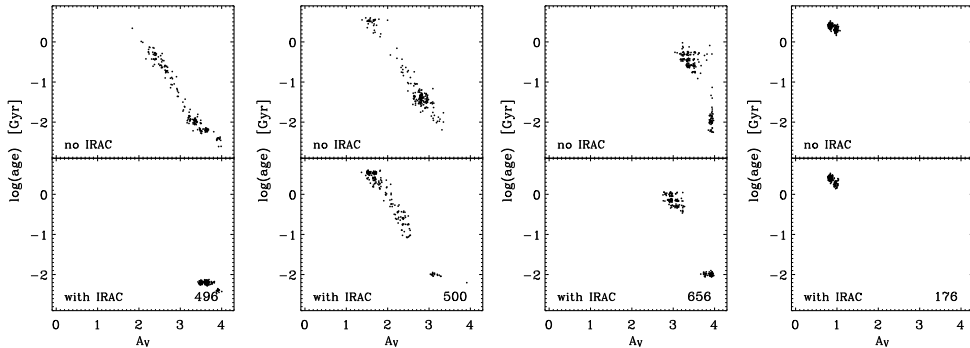


Figure 3 — The range in best-fit extinction and age values of various DRGs, for constant star forming models (Bruzual & Charlot 2003) including dust Calzetti et al (2000). We show separately the solutions obtained excluding, and including the mid-infrared data. The range in values is derived from Monte-Carlo simulations, where the photometry is randomized within the photometric errors, and the fitting procedure is repeated. The improvement of the constraints depends on the nature of the galaxies.

the optical and NIR, but much less so for moderately obscured galaxies that peak at observed $\lambda < 2.2\mu$, e.g. object 176.

The immediate question is how much the MIR fluxes improve the constraints on the models? To investigate this we used the HYPERZ package (Bolzonella, Miralles, & Pello 2000), updated with the latest Bruzual & Charlot (2003) templates, to fit stellar population models to the full SED. We used solar metallicity models with a Salpeter IMF ranging from $0.1 - 100 M_{\odot}$, and we adopted a Calzetti et al. (2000) extinction law. In the fitting we kept the redshifts fixed to the values we derived earlier with a different method (see §2), and we restricted the ages to the age of the Universe at each redshift.

Next, we derived confidence limits on the best-fit values for age and dust through Monte-Carlo simulations. For each galaxy, we generated 200 simulated SEDs by randomizing the photometry within the photometric errors, and we fitted them in the same way as the observed SEDs. We did the simulation twice: one time excluding the MIR fluxes, and the other time including the MIR fluxes. We derived the 1σ limits on each from the central 68% of the distribution. We crudely accounted for systematic errors by multiplying the limits with the square root of the reduced χ^2 of the best-fitting model. We assume that the three free parameters in the model are age, dust, and star formation rate. We show in Figure 3 the simulation output for 4 DRGs spanning a wide range in properties. Three of them were also present in Figure 2.

We can immediately learn several things from the impact of the new MIR data. Firstly, it has become possible to distinguish extremely reddened, very young models ($A_V > 3$) from much older less reddened models ($A_V < 2$) (see galaxy 496 and 500 in Figure 3). Galaxy 496 is now uniquely fit with a very

young, heavily reddened model, implying instant star formation rates of more than $2000 M_{\odot} \text{yr}^{-1}$, in the range of the sub-mm selected “SCUBA” sources (e.g., Ivison et al. 2000; Smail et al. 2000). Galaxy 656 is best fit with a very old, very dusty model in both cases. Here the improvement is modest, although best-fit solution has shifted somewhat. It is clear though that even with deep IRAC imaging the age-dust generacy is not completely resolved for all sources. We note that the uncertainties on the IRAC fluxes are mostly systematic, hence better calibration can improve the situation. But ultimately other observations, such as NIR spectroscopy, are needed to place independent constraints. Finally, as already suggested by Figure 2, constraints on models that are dominated by the blue stellar continuum with moderate extinction levels $A_V < 1$ do not improve particularly from the MIR data: the parameters were already well-constrained from the optical and NIR data alone.

In summary, we find for the whole sample of 13 DRGs that 3 galaxies are uniquely identified as old and passively evolving, 7 galaxies are well-fit by dust-reddened and star forming models with a range in extinctions and ages, 1 galaxy can be fit with either model, and 2 galaxies were badly fit. One of the 2 galaxies with a bad fit (galaxy 66) is known to have strong emission lines.

Current evidence suggests that for about half of the DRG sample the IRAC MIR imaging improves the constraints on the ages and extinction levels. It is therefore imperative to see whether this has changed the median best-fit properties of the galaxies, most importantly estimates of their stellar masses.

4 Comparison to Lyman Break Galaxies

Our previous studies of the broadband SEDs (Franx et al. 2003; Förster Schreiber et al. 2004), and emission lines (van Dokkum et al. 2004) have indicated that at a given rest-frame optical luminosity DRGs are older, more obscured, and more massive than LBGs.

The results are still somewhat preliminary however, as even at optical wavelengths the effects of dust and on-going star formation introduce uncertainties. This may be particularly relevant when comparing the stellar masses of DRGs to those of galaxies with unobscured star formation, such as LBGs. Papovich, Dickinson, & Ferguson (2001) demonstrated that it is in principle possible to hide $5\times$ the stellar mass under the glare of active, unobscured star-formation, in a maximally old stellar population.

The IRAC mid-infrared imaging greatly reduces such uncertainties, as the rest-frame K -band light is a better indicator of the amount of stellar mass. In Figure 4 we present the result of stellar population modeling with the MIR fluxes included. The main result is that the median properties of both samples are unchanged with respect to earlier modeling based on the optical-NIR SEDs (Förster Schreiber et al. 2004). In particular, we find no evidence for an old and massive population

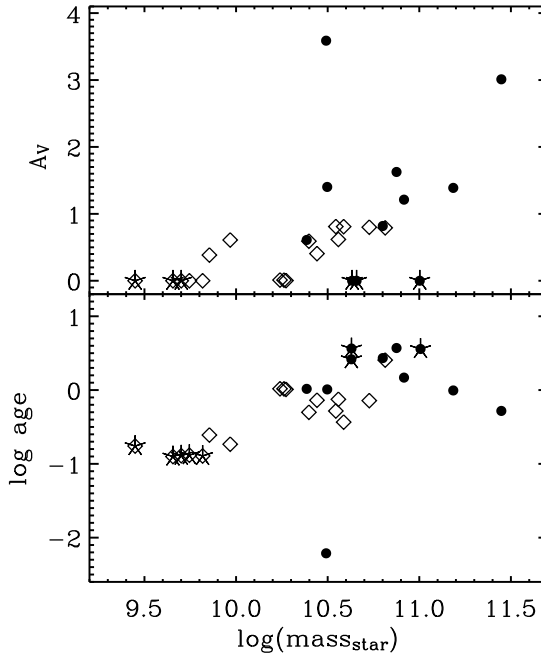


Figure 4 — (a) The best-fit values of Bruzual & Charlot (2003) model fits to the broadband SEDs of DRGs (*filled circles*) and Lyman Break Galaxies (*diamonds*). The stars indicate galaxies that are best-fit with an unreddened single age burst, others were fit with constant-star formation and Calzetti et al. (2000) dust. The DRGs are on average older, dustier, and more massive than LBGs.

hiding in the Lyman Break Galaxies. This is in line with IRAC studies of ground-based LBG samples Barmby et al. (2004), which are at somewhat higher redshift ($z \sim 3$) than our space-based U-dropout galaxies ($z \sim 2.5$).

Using the same fitting method as in §3 with the full complement of data, we find the following average values for the 11 out of 13 DRGs with good fits: a mean stellar mass of $\langle M_* \rangle = 6.8 \times 10^{10} M_\odot$, a mean age of $\langle t \rangle = 1.8$ Gyr, and a mean visual extinction $\langle A_V \rangle = 1.7$, with a large dispersion in all properties. To calculate the averages, we used the biweight estimator (Beers et al. 1990), which is not sensitive to outliers. The 3 out of 11 galaxies that were best-fit with an unreddened single age burst have a median age of $t = 2.6$ Gyr, the maximum age of the universe at $z \sim 2.5$. In contrast, the median properties of LBGs in the HDFs indicate that they are typically 4 times younger, are 4 times less obscured, and have stellar masses that are ~ 4 times lower.

Other systematic uncertainties on the mass estimates remain, such as the shape of the initial mass function (IMF) and the mass-limits, or the effects of metallicity, but these can only be resolved with direct spectroscopic measurements of kinematics and emission line ratios, which is difficult for DRGs (see e.g., van Dokkum et al. 2003, 2004).

Even without relying on SED modeling we can use simple arguments to estimate the relative contribution of two samples to the stellar mass density at $z \sim 3$. In the redshift range $2 < z < 3.5$ we find 11 DRGs and 32 LBGs to the same $K = 22.5$ magnitude limit. The DRGs emit about 60% of the light from LBGs at 8μ , but have redder $K - 8\mu$ colors, $\langle K - 8\mu \rangle = 3.5$ versus $\langle K - 8\mu \rangle = 3.1$ respectively. The $K - 8\mu$ color corresponds to the rest-frame $V - K$ at $z = 2.5$. Bruzual & Charlot (2003) stellar population models indicate that the redder rest-frame $V - K$ colors of the DRGs translate into mass-to-light ratios M/L_K that are twice as high. These values are fairly insensitive to the effects of dust or star formation history.

Hence at the high-mass end the K -selected DRGs contribute somewhat more to the $z \sim 3$ stellar mass density than LBGs, despite their lower number densities. This contribution increases if the sample were selected in the 8μ m band, which is closer to a selection by stellar mass. To a limiting magnitude at 8μ of 18.2, we recover 10 out of 13 DRGs, and 12 out of 23 LBGs, boosting the contribution of DRGs. Hence, in a mass-selected sample at high mass the DRGs contribute $1.5 - 2\times$ more to the stellar-mass density than LBGs.

The major uncertainties in these estimates are the low number statistics and the variations in space density of the red galaxies due to large scale structure. We note that the HDFs is known to contain more objects with red observed $V - H$ colors compared the HDFN (Labbé et al. 2003), although our second deep field MS1054-03 shows a similar surface density in DRGs over a much larger $5' \times 5'$ field (see, e.g., van Dokkum et al. 2003; Förster Schreiber et al. 2004).

5 The rest-frame K -band mass-to-light ratio

We used the stellar population fits to the full SED to obtain estimates of the mass-to-light ratios M/L_K . We show the results in Figure 5. In (a) the M/L_K is plotted against extinction showing that the M/L_K ratios are not sensitive to the effects of dusts. This is expected as dust absorption at rest-frame K is small (cf., the visual mass-to-light ratios, Förster Schreiber et al. 2004).

A fundamental aspect of the rest-frame K -band mass-to-light ratios is that they increase with aging of the stellar population. At the same time, complex star formation histories, such as bursts, cause a spread in the luminosity weighted ages, and thus in the mass-to-light ratios. Interestingly, the DRGs have systematically higher M/L_K than LBGs, as expected from their redder rest-frame $V - K$ color,

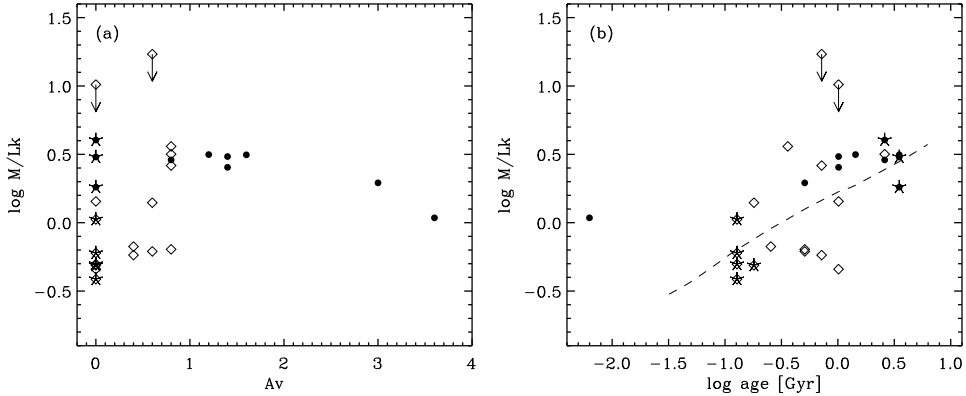


Figure 5 — The mass-to-light ratios in the rest-frame K -band (M/L_K) against visual extinction (a) and against age (b). The ages, masses and extinctions, were derived from Bruzual & Charlot (2003) stellar population fits to the SEDs. We used the best-fitting star formation history (single burst or constant) to calculate the masses. The galaxies best-fit with unreddened single age bursts are marked with a star. The DRGs (filled circles) and Lyman Break Galaxies (diamonds) are indicated. As expected, the observed K -band mass-to-light ratios are not sensitive to extinction effects, but do correlate with age. Overplotted in (b) is the evolution track of a Bruzual & Charlot (2003) stellar population model for constant star formation. Declining star formation histories follow similar tracks. The DRGs have higher mass-to-light ratios and less scatter in the mass-to-light ratios than LBGs.

but more importantly, their scatter is lower. Using the biweight estimator (Beers et al. 1990), we find $\log(M/L_K) = 0.17(\pm 0.15)$ for the DRGs versus $\log(M/L_K) = -0.22(\pm 0.45)$ for the LBGs.

This is consistent with a picture where DRGs are more evolved, and started forming at higher redshifts than LBGs. Similar conclusions were reached from emission line studies of a small sample of DRGs in the field of MS1054-03 (van Dokkum et al. 2004).

6 Discussion and Conclusions

The result presented here offer tantalizing insight into the nature of the red $J_s - K_s > 2.3$ galaxies at $z > 2$. The new mid-IR imaging with IRAC on Spitzer allowed us to uniquely identify “dead” galaxies with old passively evolving stellar populations. We placed better constraints on the relative role of age and dust for the star forming DRGs, breaking the degeneracy for some, and reducing it for many. Furthermore, we obtained more robust stellar masses, and mass-to-light ratios M/L_K in the rest-frame K -band.

We confirm the earlier studies of DRGs that found them to be more evolved, dustier, and more massive than LBGs (van Dokkum et al. 2004; Förster Schreiber et al. 2004). Specifically, we find average masses for the DRGs of $\sim 10^{11} M_{\odot}$, and we excluded the existence of large amounts of stellar mass hiding under the glare of active star formation in our sample of $z \sim 2.5$ U-dropouts (cf., Papovich, Dickinson, & Ferguson 2001; see also Barnby et al. 2004). We find that for samples selected in the observed K -band, the DRGs contribute at least as much to the stellar mass density at $2 < z < 3.5$ as Lyman Break Galaxies. If the galaxies were selected in the rest-frame K -band, as a proxy for selection by stellar mass, then the DRGs would contribute $1.5 - 2\times$ more than LBGs.

The systematically higher mass-to-light ratios of the DRGs and the lower scatter have possibly far-reaching implications for scenarios of their formation and evolution. On the one hand, it strongly suggests that DRGs are more evolved and started forming at higher redshift than LBGs. They may have started out as Lyman Break Galaxies at $z \gtrsim 5$, and then endured a prolonged period of star formation which increased their stellar mass, metallicity, and dust content. This is consistent with studies of their emission line properties (van Dokkum et al. 2004). On the other hand, the comparable M/L_K , masses, and stellar ages of the dusty star forming and “dead” DRGs suggest that they are more closely related to each other than to LBGs.

While it can not be excluded that DRGs undergo a renewed “Lyman Break phase” after the addition of metal poor gas, the higher masses suggest that they are not simply LBGs seen along more obscured lines of sight.

How DRGs relate to lower-redshift galaxies is still unclear. Given their large masses, it is inevitable that they will evolve into massive galaxies locally, as they can never lose appreciable amounts of stellar mass. However, we cannot exclude that they evolve in complicated ways, and change their appearance after dramatic events, such as a gas-rich mergers (see e.g., Steinmetz & Navarro 2002).

Finally, we have found little evidence in the broadband SEDs of the DRGs that AGNs play a major role. We find a flux excess at 8 microns for 1 out of 13 galaxies, possibly related to an obscured AGN, although it did not affect the rest of the SED. For the LBG sample we find evidence for AGN activity at 8 microns in 3 out of 23 galaxies. The high fraction of AGNs found earlier by van Dokkum et al. (2004) in a spectroscopic sample of DRGs might have been a selection effect, or implies that most of the AGNs have low-luminosity. Chandra studies of DRGs in the FIRES MS1054-03 field suggest a luminous AGN fraction of 5% (Rubin et al. 2004), comparable to our findings. In this light it seems fair that we have interpreted our broadband SEDs exclusively in terms of stellar population properties.

The results presented here can clearly benefit from larger samples, and we are undertaking more studies at optical-to-MIR wavelengths towards this goal. In addition, large programs are underway to obtain follow-up optical and NIR

spectroscopy, which are constraining their space densities, and are providing independent ways to measure the extinctions, star formation rates, and masses. While such studies are now very hard, they will benefit tremendously from the arrival of multi-object NIR spectrographs at 8 – 10 meter class telescopes.

Acknowledgments

This research was supported by grants from the Netherlands Foundation for Research (NWO), the Leids Kerkhoven-Bosscha Fonds, the Lorentz Center, and the Smithsonian Institution. GR would like to acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG), SFB 375 (Astroteilchenphysik).

References

- Barmby, P. et al. 2004, ApJS, in press
Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100, 32
Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, A&A, 363, 476
Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000 (BC03)
Casertano, S. et al. AJ, 120, pp. 2747–2824, 2000
Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., & Storchi-Bergmann, T. 2000, ApJ, 533, 682
Efstathiou, G., et al. 2002, MNRAS, 330, L29
Fazio, G.G. et al., ApJS, in press
Franx, M. et al. *The Messenger* **99**, pp. 20–22, 2000
Franx, M. et al. 2003, ApJ, 587, L79
Förster Schreiber et al. 2004, ApJ, accepted (astro-ph/0408077)
Freedman, W. L., et al. 2001, ApJ, 553, 47
Giavalisco, M. & Dickinson, M. 2001, ApJ, 550, 177
Iverson, R. J., Smail, I., Barger, A. J., Kneib, J.-P., Blain, A. W., Owen, F. N., Kerr, T. H., & Cowie, L. L. 2000, MNRAS, 315, 209
Labbé, I., et al. 2003, AJ, 125, 1107
Labbé, I., et al., 2004, in preparation
Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996, MNRAS, 283, 1388
Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, ApJ, 559, 620
Rudnick, G. et al. 2001, AJ, 122, 2205
Rudnick, G., et al. 2003, ApJ, 599, 847
Rubin, K., et al. 2004, ApJ, accepted (astro-ph/0408101)
Smail, I., Iverson, R. J., Owen, F. N., Blain, A. W., & Kneib, J.-P. 2000, ApJ, 528, 612
Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, AJ, 112, 352
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger
Steidel, C. C., Adelberger, K. L., Shapley, A. E., Pettini, M., Dickinson, M., & Giavalisco, M. 2003, ApJ, 592, 728
Steidel, C. C., Shapley, A. E., Pettini, M., Adelberger, K. L., Erb, D. K., Reddy, N. A., & Hunt, M. P. 2004, ApJ, 604, 534
Steinmetz, M. & Navarro, J. F. 2002, New Astronomy, 7, 155
Shapley, A. E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., & Pettini,

- M. 2001, ApJ, 562, 95
Shapley, A. E. et al. 2003, ApJ, accepted (astro-ph/0405187)
Spergel, D. N., et al. 2003, ApJS, 148, 175
van Dokkum, P. G. et al. 2003, ApJ, 587, L83
van Dokkum, P. G. et al. 2004, ApJ, accepted (astro-ph/0404471)

Nederlandse samenvatting

Achtergrond

STERRENSTELSELS zijn de meest majestueuze en imponerende verschijningen aan de hemel. Majestueus, maar tegelijkertijd een onbegrepen fenomeen. Ze herbergen miljarden sterren en vormen lichtbakens in een uitgestrekt heelal dat voornamelijk gevuld is met leegte. Sommigen zijn egaal, roodkleurig, en elipsvormig, anderen zijn blauw, schijfvormig, met prachtige spiraalstructuren. Edwin Hubble was de eerste die sterrenstelsels op basis van deze vormen indeelde. Hoe komt het dat sterren wel in melkwegstelsels ontstaan maar niet daarbuiten? En wat veroorzaakt die verschillen in structuur?

Die vraag heeft diepere gronden dan men zou vermoeden. Het blijkt dat de vorming en evolutie van sterrenstelsels nauw verbonden is met de eigenschappen van het heelal als geheel. De astronoom Fritz Zwicky realiseerde zich al in 1937, dat er iets mysterieus aan de hand was met de bewegingen van sterrenstelsels in clusters. Clusters zijn enorme opeenhopingen van sterrenstelsels die, gevangen in elkaars zwaartekracht, om elkaar heen wentelen. Zwicky mat de snelheden van de sterrenstelsels en rekende uit hoeveel massa vereist was om sterrenstelsels zo te laten bewegen. Dit bleek maar liefst meer dan 100 maal zo veel te zijn als het gewicht van alle sterren in die stelsels bij elkaar. Blijkbaar heeft een sterrenstelsel een hoop te verbergen. Het probleem van de “donkere materie” was geboren.

Tot op de dag van vandaag, bijna 70 jaar later, weten we niet wat donkere materie is. Het enige dat we weten is dat normale lichtgevende materie, het materiaal waar u, ik, en onze Zon uit bestaan, een bijna verwaarloosbare fractie van het heelal vormt, slechts een onbetekenende lichtvervuiling. Opeenhopingen van de onderliggende donkere materie, die langzaam maar zeker groeien door met elkaar te botsen en te versmelten, bepalen uiteindelijk het uiterlijk en de bewegingen van sterrenstelsels. Het positieve is dat het licht van de sterren dus wel aangeeft waar de donkere materie ophoopt. Zodoende geeft het ons een indirecte manier om de donkere materie te bestuderen.

In de afgelopen decennia is onze kennis over de structuur en de inhoud van het heelal op de grootste schaal in een stroomversnelling geraakt. Na een ware stortvloed van waarnemingen uit het diepste heelal, is het nu vrijwel zeker dat dat er ooit een tijd was zonder sterrenstelsels en zonder sterren. De donkere materie was extreem gelijkmatig verdeeld, met slechts minieme rimpels, en voor de rest was er niets dan heet gas. Toen het heelal vervolgens expandeerde en afkoelde,

groeiden de rimpels van donkere materie tot golven. De pieken van deze golven vormden de kiemen van sterrenstelsels, waar afgekoeld gas naar toe stroomde en de eerste sterren vormde. De vreemde situatie in de kosmologie heden ten dage, is dat we meer lijken te weten over het allereerste begin van sterrenstelsels en het eindresultaat, dan over wat er is gebeurd in de tussentijd. Eén van de drijfveren achter dit proefschrift is om daar verandering in te brengen.

Zelfs deze relatief bescheiden taak – er zijn veel lastiger vragen te bedenken – is een formidabele uitdaging. Stof, gas, stervorming, supernova's, zwarte gaten; dit alles speelt een rol in de evolutie van een sterrenstelsel. We weten nog te weinig van sterrenstelsels om ze met formules te beschrijven. We kunnen dus niet simpelweg uitrekenen wat er tussen de oerknal en nu is gebeurd. Daarnaast kunnen we niet met sterrenstelsels experimenteren; we kunnen ze niet opnieuw laten ontstaan.

De redding van de kosmologie is dat de snelheid van het licht niet zo snel is als men denkt. Weliswaar reist het licht in een vingerknip om de aarde en in 8 minuten van hier naar de zon, maar naar de dichtstbijzijnde ster duurt de reis alweer ruim 4 jaar, en van het ene uiteinde van ons sterrenstelsel – de Melkweg – naar het andere, maar liefst 100.000 jaar. De implicatie is dat wanneer men kijkt naar sterrenstelsels op nog veel grotere afstanden, het licht er soms wel miljoenen jaren over heeft gedaan om ons te bereiken. We zien dan hoe het sterrenstelsel er miljoenen jaren geleden uitzag en kijken effectief terug in de tijd.

Dit is de basis van de techniek die in dit proefschrift is toegepast. We maken met de modernste reuzentelescopen de zeer gevoelige afbeeldingen van ver verwijderde sterrenstelsels. Vervolgens bepalen we de afstand tot de sterrenstelsels en zo weten we wanneer het licht verzonden was. Tenslotte zetten we de afbeeldingen in de goede volgorde en het resultaat is een kosmische film die laat zien hoe sterrenstelsels van hun geboorte tot nu zijn geëvolueerd.

Infrarood

In de praktijk is het niet zo eenvoudig. Als men sterrenstelsels vanaf hun jeugd wil zien, moet men meer dan 10 miljard lichtjaar ver weg kijken, dus slechts een paar miljard jaar na de oerknal. Sterrenstelsels op dermate grote afstanden zijn vanaf de aarde bezien absurd zwak. Zelfs met de allerbeste instrumenten op de grootste telescopen moet men tientallen uren licht verzamelen om ze te kunnen zien.

Daarnaast gebeurt er iets met het licht tijdens zijn reis. Door de expansie van het heelal wordt het licht opgerekt en verschuift naar roodere golflengtes, evenredig met de afstand die is afgelegd. Wat op grote afstanden ooit is verzonden op de vertrouwde zichtbare golflengten komt nu bij ons aan in het infrarode deel van het spectrum.

Voor de komst van de huidige generatie infraroodcamera's keek men slechts met instrumenten die gevoelig waren voor zichtbaar licht, zoals met de WFPC2 camera op de Hubble Space Telescope. Door de roodverschuiving keek men in

de werkelijkheid naar het ultraviolette (UV-) licht uit het verre heelal. Tot enkele jaren geleden, grofweg tot het begin van dit proefschrift, werd de meeste informatie over het jonge heelal verkregen via waarnemingen door deze “UV-bril”.

Nu kan het zo zijn dat het heelal door een UV-bril een totaal misvormd, of erger, incompleet beeld geeft. Als we bijvoorbeeld op zo’n manier het nabije heelal zouden bekijken, dan zouden de zwaarste sterrenstelsels, de elliptische stelsels, vrijwel onzichtbaar zijn. Andere reuzenstelsels, zoals onze eigen Melkweg, zouden er totaal anders uitzien.

Dus voordat we toekomen aan de vraag hoe sterrenstelsels zijn ontstaan, moeten we ons afvragen of we wel een compleet beeld hebben van sterrenstelsels in het vroege heelal. Infrarode waarnemingen spelen daarbij een cruciale rol.

Dit proefschrift

Diepste Infrarode Blik in het Heelal

In **hoofdstuk 2** van dit proefschrift presenteren we de diepste en meest gevoelige infraroodfoto’s genomen die ooit zijn gemaakt van de hemel. Gebruik makend van een nieuwste infraroodcamera (ISAAC) op de Very Large Telescope (VLT) van de European Southern Observatory (ESO) te Chili, en slechts observerend onder optimale opstandigheden, bekeken we tussen oktober 1999 en oktober 2000 meer dan 100 uur lang één speciale plek aan de hemel, het Hubble Deep Field South (HDFS).

Het HDFS – een gebiedje aan de hemel dat honderd keer zo klein lijkt als de volle maan – werd in 1998 speciaal geselecteerd omdat het schijnbaar leeg was, dus zonder heldere sterren of sterrenstelsels. Een opname door de Hubble ruimtetelscoop met een extreem lange belichtingstijd (in totaal een week) onthulde toen honderden sterrenstelsels op miljarden lichtjaren afstand. Afgezien van de tegenhanger uit 1995, de Hubble Deep Field North (HDF-N), was tot dan toe nog nooit zo diep in het heelal gekeken.

Verrassend genoeg onthulden onze nieuwe infraroodwaarnemingen tal van sterrenstelsels die bij de vorige observaties verborgen bleven. Van al deze sterrenstelsels bepaalde we vervolgens de precieze helderheden, kleuren, en afstanden. Het bleek dat veel van deze stelsels op zo’n grote afstand staan dat hun licht meer dan 12 miljard jaar nodig heeft gehad om de aarde te bereiken. We zien ze dus in hun prille ontwikkelingsfase, minder dan 2 miljard jaar na de oerknal waaruit het heelal is ontstaan. Desondanks lijken sommige exemplaren dankzij hun rode kleur al verrassend veel op “volwassen” sterrenstelsels, zoals de elliptische stelsels in het lokale heelal. Dit is niet makkelijk te verklaren in gangbare modellen van de vorming van sterrenstelsels.

Het gebruik van de infraroodcamera was essentieel, omdat de ontdekte stelsels het meeste van hun licht uitzenden in het zichtbare licht. Zichtbaar licht, uit-

gezonden door sterren op 12 miljard lichtjaar afstand, wordt door ons op aarde na de roodverschuiving waargenomen als nabij-infrarood licht (golflengte 0,001 - 0,002 millimeter). Dat is ook de voornaamste reden dat de Hubble ruimtetelescoop destijds deze populatie primordiale sterrenstelsels heeft gemist.

In **hoofdstuk 3** beschrijven we andere sterrenstelsels die nog nooit eerder waren gezien zo vroeg in de evolutie van het heelal: spiraalstelsels van reusachtige afmetingen. Het formaat van deze stelsels is praktisch hetzelfde als die van onze eigen Melkweg, of onze naaste buur, het Andromeda stelsel. Theorieën voor de vorming van deze schijfstelsels zeggen dat zulke reuzen pas vele miljarden later ontstaan. Het is nog onbekend waarom we er zo vroeg al zo veel zien.

De Kleur van de Kosmos

Hoofdstuk 4 beschrijft de evolutie van de gemiddelde kosmische kleur, en daarmee de gemiddelde leeftijd van de sterren in het heelal.

Met behulp van de kleuren, helderheden en de afstandsdata konden het roodverschuivingseffect in het licht van de sterrenstelsels ongedaan maken. Tegelijkertijd, gaf de afstand (in lichtjaren) tevens aan hoe ver we naar elk sterrenstelsel terug in de tijd keken. Door het licht van alle sterrenstelsels in een bepaald afstandsinterval te sommeren, kon de ware kleur van het heelal in elk tijdperk berekend worden.

De diepe waarnemingen in het HDFS van het verleden naar heden leverde zo een unieke, historische volkstelling onder sterren op. De waarnemingen tonen aan dat in de meeste sterrenstelsels – net als in vergrijzende westerse landen – de gemiddelde leeftijd van sterren stijgt, aangezien er onvoldoende geboortes plaatsvinden om de overledenen op te volgen. Jonge sterren zijn heet en blauw, terwijl oude sterren koel en rood zijn, zodat ook het heelal als geheel steeds roder wordt. 2,5 Miljard jaar na de oerknal was het heelal nog blauw, maar nu, 11 miljard jaar later, is het beige.

Weliswaar ontstaan in de meeste sterrenstelsels telkens nieuwe sterren uit het gas en stof die de ruimte tussen de sterren vult – op elk moment zal een modaal sterrenstelsel dus sterren van alle leeftijden bevatten – maar gemiddeld over alle sterrenstelsels blijkt er een duidelijke, gelijkmatige verouderingstrend in de sterrenpopulatie op te treden.

Zoals gezegd, is de blauwe kleur van het vroege heelal het gevolg van de jeugdige leeftijden van de sterren die toen de sterrenstelsels bevolkten. Omdat we weten dat jonge, blauwe sterren veel meer licht uitstralen dan oude, rode sterren, kunnen we ook concluderen dat het aantal sterren in het jonge heelal veel lager was dan nu. Immers de totale hoeveelheid licht is ongeveer constant gebleven. Onze resultaten impliceren dat relatief veel sterren vrij laat in de geschiedenis van het heelal gevormd zijn, niet lang voordat ook onze zon is ontstaan. Deze conclusie, die goed aansluit bij wat al bekend was over leeftijden van sterren in ons deel van

het heelal, geeft voor het eerst direct zicht op de evolutie van sterrenpopulaties gedurende een groot deel van de geschiedenis van het heelal.

Cyclische Stervorming

In **hoofdstuk 5** bestuderen we de verdeling van de kleuren van sterrenstelsels met actieve stervorming zo'n 2.5 miljard jaar na de oerknal; hoeveel zijn er blauw, hoeveel zijn er rood? Omdat we er redelijk zeker van zijn dat we geen grote aantallen sterrenstelsels meer over het hoofd zien, kunnen we nu de statistische verdeling van de kleuren gebruiken om meer te leren over de gemiddelde ontstaansgeschiedenis.

De kleurverdeling van sterrenstelsels in het jonge heelal is echter paradoxaal. Aan de ene kant was de spreiding in de kleurverdeling heel klein, wat gewoonlijk duidt op gemiddeld vrij hoge leeftijd in de sterrenpopulaties (meer dan 1.5 miljard jaar), maar de gemiddelde kleur was ook extreem blauw, wat weer duidt op een jonge leeftijd (minder dan 0.5 miljard jaar).

We onderzochten verschillende modellen voor de ontstaansgeschiedenis om deze verdeling te verklaren en zijn tot de conclusie gekomen dat een model met cyclische stervorming de beste oplossing geeft. In dit scenario vormen de stelsels niet continu sterren, maar zijn ze actief gedurende een bepaalde periode om vervolgens een korte tijd inactief te zijn. Na de inactieve periode, hervat de stervorming weer. De inactiviteit mag echter niet te lang duren, want tijdens een inactieve periode wordt de kleur zeer snel roder en we weten inmiddels dat er 2.5 miljard jaar na de oerknal niet heel veel rode stelsels zijn waargenomen.

Het blijkt dat de cyclische stervorming resulteert in gemiddeld zeer blauwe kleuren, zelfs op relatief hoge leeftijden. Zodoende biedt het een oplossing voor de paradox.

Rood, Roder, Roodst

Hoofdstuk 6 besluit met een eerste stap in het vervolgonderzoek van de eerder ontdekte rode stelsels, zoals reeds beschreven in hoofdstuk 2. Onze nabij infrarood data (1 – 2 micrometer) alleen was echter nog niet genoeg om met zekerheid vast te stellen wat de oorzaak is van de rode kleur. Is het daadwerkelijk de oude leeftijd van de sterrenpopulatie? Of speelt een roodkleuring door stofabsorptie een rol? Het antwoord werd gegeven door mid-IR afbeeldingen, verkregen met de IRAC camera aan boord van de gloednieuwe Spitzer ruimtetelescoop.

Onze mid-IR waarnemingen geven zicht op het roodverschoven nabij-IR licht van deze ver verwijderde stelsels. In het intrinsieke nabij-IR licht is het makkelijker om een onderscheid te maken tussen de effecten van stof en leeftijd. We vonden voor het eerst definitieve bewijzen voor zeer oude stelsels in het jonge heelal. Daarnaast bleek dat de door ons gevonden rode stelsels systematisch zwaarder waren dan de sterrenstelsels die voorheen in het zichtbare licht waren ontdekt.

Zelfs al zijn de rode stelsels kleiner in aantal dan de al bekende blauwe stelsels in het jonge heelal, toch herbergen ze de meeste van alle sterren. Zij moeten dus een cruciale rol spelen in de onstaansgeschiedenis van sterrenstelsels. Een belangrijk deel van het toekomstige onderzoek aan sterrenstelsels zal er op gericht zijn om deze stelsels beter in kaart te brengen.

Curriculum vitae

Ik kwam ter wereld op 11 november 1972 te Alphen a/d Rijn en bracht het grootste deel van mijn jeugd door in Leiderdorp. In juni 1991 behaalde ik het VWO diploma aan het Stedelijk Gymnasium te Leiden.

Ik studeerde van 1991 tot en met 1992 Cognitieve Kunstmatige Intelligentie aan de Universiteit van Utrecht. In september 1993 begon ik aan de studie Sterrenkunde aan de Universiteit Leiden. Tijdens mijn doctoraalfase werkte ik met prof. dr. M. Franx aan gravitationeel gelensde afbeeldingen van sterrenstelsels in het verre heelal. Tegelijkertijd was ik freelance programmeur voor verschillende bedrijven. Mijn studie resulteerde op 29 februari 2000 in het behalen van het doctoraalexamen Sterrenkunde.

Op 1 maart 2000 trad ik in dienst als Onderzoeker in Opleiding aan de Universiteit Leiden op een subsidie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek. Onder leiding van prof. dr. M. Franx onderzocht ik de evolutie van sterrenstelsels met behulp van zeer gevoelige infrarood afbeeldingen. Naast mijn promotor waren mijn voornaamste medewerkers dr. G. Rudnick en dr. N.M. Förster Schreiber.

Ik bracht werkbezoeken aan het California Institute of Technology (Pasadena, VS) en het Harvard-Smithsonian Center for Astrophysics (Cambridge, MA, VS). Ik nam deel aan de NOVA herfstschool in 2002, en presenteerde mijn onderzoeksresultaten op conferenties en workshops in Durham, Hawaii, Heidelberg, Manchester, München, Durham, Sydney, en Venetie (2x). Ik heb gedurende mijn onderzoek waarnemingen verricht met de 10 m W.M. Keck Telescope op Hawaii (1 keer) en met de 4.2 m W. Herschel Telescope op La Palma (4 keer). Gedurende 3 semesters assisteerde ik bij het sterrenkundig practicum.

Tenslotte droeg ik bij aan de popularisering van de sterrenkunde, door middel van openbare voordrachten, radio- en kranteninterviews en publicaties in populair-wetenschappelijke tijdschriften. Als muzikant speelde ik jarenlang in een viermans funk-formatie. Ik beoefen verscheidene sporten.

Na mijn promotie zal ik als Carnegie Fellow verbonden zijn aan de Observatories of the Carnegie Institution of Washington (OCTW) in Pasadena (CA, VS).

Nawoord / Acknowledgments

VELE mensen ben ik dank verschuldigd voor de totstandkoming van dit proefschrift. Hier wil ik recht doen aan diegenen, die op welke manier dan ook, zelfs als niet met naam genoemd, onmisbaar zijn geweest.

Voorop wil ik stellen dat ik geen betere plek had kunnen wensen voor deze promotie. Over de hele linie, van het secretariaat en systeembeheer tot aan de staf, heb ik niets dan ondersteuning en aanmoediging ondervonden. Een speciaal woord van dank heb ik voor Kirsten Groen en Jan Lub.

The most important aspect, however, was being part of the FIRES team. Without the enthusiasm, inspiration, and guidance of its members, and the friendship and collegiality of Greg Rudnick and Natascha Förster Schreiber in particular, this work would not have been remotely possible.

I gratefully acknowledge the Leidsche Kerkhoven-Bosscha Fonds, the Nederlandse Organisatie voor Wetenschappelijk Onderzoek, the Leidse Sterrewacht, and the Harvard-Smithsonian Center of Astrophysics for their generous financial support.

Los van het harde werk heb ik de afgelopen jaren genoten van de sociale activiteiten op de Sterrewacht. Het Forza voetbalteam, het jaarlijkse nacht-volleybal toernooi, de vrijdagmiddagborrel: het leven op de Sterrewacht zou grijs zijn geweest zonder deze extracurriculariteiten.

Mijn vrienden hebben door hun afleiding en stelselmatige desinteresse voor de “Astrologie” meer in positieve zin bijgedragen dan zij zich realiseren. Anna, Bas, Flip & Tiets, Lou, Micha, Miesj, Pedro, Superleen en Thomas, zonder jullie was hier niets van terecht gekomen.

Als laatste wil ik mijn ouders en Dymph bedanken voor de liefdevolle wereld waarin ik ben opgegroeid. Met mijn aanstaande reis naar de VS voor de boeg is het onmisbaar te weten dat er altijd een thuis is.