DISSERTATION IN ASTRONOMY

SUBMITTED TO THE
COMBINED FACULTIES OF THE NATURAL SCIENCES AND MATHEMATICS
OF THE RUPERTO-CAROLA-UNIVERSITY OF HEIDELBERG, GERMANY,
FOR THE DEGREE OF
DOCTOR OF NATURAL SCIENCES

PRESENTED BY

M.SC. KASPER BORELLO SCHMIDT
BORN IN NYBORG, DENMARK

ORAL EXAMINATION: JULY 9TH, 2012

FRONTIERS OF GALAXY EVOLUTION

– TIME-DOMAIN OBSERVATIONS AND 3D SPECTROSCOPY –

REFEREES: PROF. DR. HANS-WALTER RIX
PROF. DR. JOACHIM WAMBSGANß

ABSTRACT

Understanding the formation and evolution of galaxies through cosmic time has been a central focus of astrophysics in the last decades: how did the interplay between dark matter structure formation, star formation, galaxy merging, and active galactic nuclei (AGN) give rise to the observed galaxy properties at different redshifts? This thesis presents innovative observational approaches to two aspects of this problem: finding and studying AGN through their variability, and making a first systematic census of galaxy mergers at $z > 1$ through three-dimensional spectroscopy.

First we present a new and simple technique for selecting extensive, complete, and pure quasar samples via their intrinsic variability, parameterizing the single-band light-curve structure function through a power-law to identify quasars among other variable and non-variable sources. Using extensive multi-epoch observations from SDSS Stripe 82 containing ~60 epochs taken over ~8 years, we demonstrate the power of this approach. The presented algorithm identifies quasars with a completeness and purity above 90% at all redshifts. Even for Pan-STARRS 1 mock data of only 6 epochs over 3 years, variability is still an encouragingly efficient quasar classifier. Data on intrinsic quasar variability enable a wide range of astrophysical science. We quantify the *color* variability, confirming and greatly fleshing out previous claims, that quasars become bluer as they brighten. We find a strong redshift dependence of this blueing, which we can attribute to emission lines contributing to the SDSS bands at different redshifts. We find that the color variations of single quasars are much more pronounced than the ranges in color seen in time-averaged ensembles of quasars. This indicates, that the observed color variations cannot be explained by changes in the mean steady state AGN accretion rate, but must arise from accretion disk 'hotspots' or similar phenomena.

In the second, distinct part of the thesis, we present the first large sample of morphologically selected galaxy mergers with three-dimensional spectroscopy at $z \sim 1.5$. With individual masses and star formation rates derived from multi-band photometry, we created emission line maps from the slitless grism spectroscopy of 3D-HST as proxy for star formation maps, providing a comprehensive empirical picture of where star formation takes place in galaxy mergers at the epoch, where the cosmic star formation and merger rates peaked. We find that a broad range of star formation morphologies occur at all redshifts, irrespective of star formation rate and total stellar mass, in these mergers. An initial illustrative comparison to a set of cosmological simulations shows, that simulated mergers with similar mass and gas content show star formation in both merger components far more often than for the observed 3D-HST mergers. This suggests that mergers at $z \sim 1.5$ most commonly happen between galaxies of distinctly different gas fractions.

ZUSAMMENFASSUNG

Das Verständnis der Entstehung und Entwicklung von Galaxien durch die kosmischen Epochen ist seit Jahrzehnten ein zentraler Fokus der Astrophysik: Wie führte das Wechselspiel zwischen der Strukturentstehung der dunklen Materie, Sternentstehung, Galaxienverschmelzung und aktive Galaxiekerne (AGN) zu den beobachteten Eigenschaften der Galaxien bei unterschiedlichen Rotverschiebungen? Diese Dissertation präsentiert neue Beobachtungsherangehensweisen an zwei Aspekte dieses Problems: Das Auffinden und Untersuchen von AGN durch deren Variabilität und die Erstellung der ersten systematischen Untersuchung von Galaxienverschmelzungen bei $z > 1$ durch dreidimensionale Spektroskopie.

Zuerst präsentieren wir eine neue und einfache Technik um große, komplette, und reine Stichproben an Quasaren durch deren intrinsische Variabilität zu bestimmen. Wir parametrisieren die Strukturfunktion deren Einzelbandlichtkurven durch ein Potenzgesetz um Quasare von anderen variablen und nicht-variablen Quellen zu unterscheiden. Mit Hilfe von Multi-Epochen-Beobachtungen des SDSS Stripe 82, welche ~60 Epochen über ~8 Jahre umfassen, können wir die Effizient dieser Suchweise zeigen. Der vorgelegte Algorithmus identifiziert Quasare mit einer Vollständigkeit und Reinheit über 90% bei allen Rotverschiebungen. Sogar für simulierten daten für den Pan-STARRS 1 Survey, mit nur 6 Epochen über 3 Jahre, ist die Variabilität ein ermutigend effizienter Klassifikator für Quasare. Daten zur intrinsischen Variabilität der Quasare ermöglichen einen weiten Bereich von astrophysikalischer Anwendungen. Wir quantifizieren die *Farb*-Variabilität und bestätigen und untermauern unsere vorherigen Hinweise, daß Quasare mit zunehmender Helligkeit blauer werden. Wir finden eine starke Rotverschiebungs-Abhängigkeit dieser Farb-Variabilitaet. Dies können wir Emissionslinien zuschreiben, die zu den SDDS Bändern bei verschiedenen Rotverschiebungen beitragen. Wir finden, daß die Farbvariationen von Einzelquasaren viel ausgeprägter sind als die Farbbereiche in zeitlich gemittelten Quasarensembles. Die zeigt, daß die beobachteten Farbvariationen nicht durch Änderungen in der stationären Durchnittsakkretionsrate der AGN erklärbar sind, sondern von 'hotspots' oder ähnlichen Phänomenen der Akkretionsscheiben herrühren müssen.

In dem zweiten eigenständigen Teil der Dissertation präsentieren wir die erste große, morphologisch selektierte Auswahl von Galaxienverschmelzungen mit dreidimensionaler Spektroskopie bei $z \sim 1.5$. Mit individuellen Massen und Sternentstehungsraten, die von Multiband-Photometrie abgeleitet werden, erzeugen wir Karten der Emissionslinien von der spalt-losen Grismspektroskopie von 3D-HST stellvertretend für Sternentstehungskarten. Diese stellen ein umfassendes, empirisches Bild bereit, das zeigt wo Sternentstehung in Galaxienverschmelzungen in dieser Epoche, in der kosmische Sternentstehung und Verschmelzungsraten einen Höhepunkt erreichen, passiert. Die Untersuchung hat gezeigt, dass ein weiter Bereich von Sternentstehungsmorphologien bei allen Rotverschiebungen passiert, unabhängig von der Sternentstehungsrate und totalen stellaren Masse in diesen Verschmelzungen. Ein illustrativer Anfangsvergleich von einem Satz von kosmischen Simulationen weist daruaf hin, daß Verschmelzungen bei $z \sim 1.5$ am häufigsten zwischen Galaxien mit deutlich unterschiedlichen Gasanteilen passieren.

CONTENTS

1 INTRODUCTION 1

1.1 OUR PLACE IN THE UNIVERSE 2

1.2 THE STANDARD BIG BANG MODEL 3

1.3 STRUCTURE FORMATION & (Λ)CDM BUILDING BLOCKS 5

1.4 BUILDING A GALAXY 10

1.5 GALAXIES' ACTIVE (QUASAR) PHASE 12

1.5.1 QUASARS – AGN AT ITS BEST 16

1.6 OUTLINE OF THIS THESIS 17

2 SELECTING QUASARS VIA VARIABILITY 19

2.1 – PROLOGUE – 20

2.2 IGNITING GALAXIES – THE QUASAR (AGN) ENGINE 22

2.3 VARIABILITY CHARACTERIZATION OF SOURCES AND THE STRUCTURE FUNCTION 27

2.3.1 COLOR SELECTION 27

2.3.2 SOURCE VARIABILITY: POWER-LAW STRUCTURE FUNCTIONS 28

2.4 SDSS STRIPE 82: A TESTBED FOR VARIABILITY STUDIES 31

2.4.1 Spectroscopically confirmed quasars in Stripe 82 32

2.4.2 Stellar locus ‘contaminants’ in Stripe 82 33

2.4.3 UV excess (UVX) objects 33

2.4.4 Non-UV excess (nUVX) objects 34

2.4.5 Quasar Candidate Color Selection without u-band Data . . . 34

2.4.6 Eliminating light curve ‘outliers’ in Stripe 82 35

2.4.7 Down-sampling Stripe 82 light curves to the Pan-STARRS 1 cadence 35

2.5 Power-Law Structure Functions for Sources in Stripe 82 . . . 38

2.6 Results 39

2.6.1 The A–γ Distribution 40

2.6.2 Completeness and Purity 42

2.6.2.1 Quasars in the UVX catalog 44

2.6.2.2 Quasars in the nUVX catalog 45

2.6.2.3 Quasars in the $griz$ Color Box 47

2.7 Discussion 50

2.8 Conclusions 52

2.9 Outlook 54

2.9.1 Probabilistic Combination of Selection Methods 54

2.9.2 Selecting Lensed Quasars with Variability 55

2.10 – Epilogue – 58

3 The Quasar Color Variability 59

3.1 – Prologue – 60

3.2 Probing the Physics of Quasars 62

3.2.1 The (Rest-Frame Optical) Quasar Spectrum 62

3.2.2 EDDINGTON RATIOS AND BLACK HOLE MASSES 64

3.2.3 EMISSION LINE RATIO DIAGNOSTIC – THE BPT DIAGRAM 65

3.2.4 THE VARIABILITY OF QUASARS 67

3.2.5 REVERBERATION MAPPING - MEASURING SIZES OF BLRS 67

3.2.6 SPECTRAL HARDNESS – BRIGHTER MEANS BLUE 69

3.3 SDSS STRIPE 82 DATA 70

3.4 FITTING COLOR VARIABILITY IN MAGNITUDE SPACE 70

3.5 RESULTS 73

3.5.1 COLOR VARIABILITY IN *gr* 73

3.5.2 REPRODUCING THE COLOR VARIABILITY REDSHIFT DEPENDENCE WITH SIMPLE VARIABILITY MODEL 75

3.5.3 COLOR VARIABILITY IN *ui* 77

3.6 DISCUSSION 79

3.6.1 COLOR VARIABILITY AS A FUNCTION OF EDDINGTON LUMINOSITY AND BLACK HOLE MASS 79

3.6.2 THE COLOR VARIABILITY AS A FUNCTION OF THE LIGHT CURVE VARIABILITY CHARACTERISTICS 80

3.6.3 COLOR VARIABILITY AND CHANGES IN THE MEAN ACCRETION RATE . . 85

3.6.3.1 COLOR VARIABILITY OF INDIVIDUAL QUASARS VS. THE COLOR DISTRIBUTION OF QUASAR ENSEMBLES 85

3.6.3.2 COLOR VARIABILITY VS. ACCRETION DISK MODELS . . . 86

3.7 CONCLUSION 89

3.8 OUTLOOK 91

3.8.1 MODELING OF SPECTRAL QUASAR VARIABILITY 91

3.8.2 PHOTOMETRIC REVERBERATION MAPPING 92

3.9 – EPILOGUE – 94

4 STAR FORMATION IN HIGH REDSHIFT GALAXY MERGERS 95

4.1 – PROLOGUE – 96

4.2 STAR FORMATION & GALAXY MERGERS 98

4.2.1 OBSERVING THE STAR FORMATION IN MERGERS 101

4.2.2 MODELING THE STAR FORMATION IN MERGERS 102

4.2.3 THIS STUDY: EXPLOITING NEW DATA 104

4.3 THE 3D-HST SPECTROSCOPIC SURVEY DATA 105

4.3.1 THE 3D-HST GRISM SPECTROSCOPY 106

4.4 SELECTING MERGERS 107

4.4.1 GRISM CATALOG CUTS 107

4.4.2 FITTING SEDS TO PHOTOMETRY 109

4.4.3 VISUAL INSPECTION OF NEAR-INFRARED MORPHOLOGY 109

4.5 EMISSION LINE MAPPING 112

4.5.1 CONTINUUM MODELING AND SUBTRACTION 113

4.5.2 THE EMISSION LINE MAP REDSHIFT, z_{ELMAP} 115

4.6 RESULTS: THE SPATIAL EXTENT OF STAR FORMATION IN HIGH-z MERGERS . 116

4.7 SIMULATING 3D-HST SPECTRA 119

4.7.1 SPH MERGER SIMULATIONS 120

4.7.2 SIMULATING 3D-HST GRISM SPECTRA 121

4.7.3 CREATING STAR FORMATION MAPS OF SIMULATED SPECTRA 122

4.7.4 COMPARING THE 3D-HST DATA WITH SIMULATIONS 122

4.8 CONCLUSION 124

4.9 OUTLOOK 126

4.9.1 EXPANDING THE ANALYSIS TO THE FULL 3D-HST SURVEY 126

4.9.2 IFU FOLLOW-UP OF OBJECTS 127

4.9.3 STAR FORMATION AND MOLECULAR GAS AT HIGH REDSHIFT 128

4.9.4 COMPREHENSIVE AND DETAILED SAMPLING OF THE SIMULATION'S PARAMETER SPACE 128

4.10 – EPILOGUE – 130

5 SUMMARY & CONCLUSION 131

5.1 CONTRIBUTING TO THE GALAXY EVOLUTION PUZZLE 132

5.2 SELECTING QUASARS WITH VARIABILITY 132

5.3 THE COLOR VARIABILITY OF QUASARS 133

5.4 STAR FORMATION IN MERGERS AT $z \sim 1.5$ 135

5.5 OUTLOOK 135

5.6 THE IMMEDIATE FUTURE OF 'YOURS TRULY' 138

A INDIVIDUAL STRUCTURE FUNCTION PARAMETER INFERENCE BY MCMC 141

B FITTING IN COL.-MAG. AND MAG.-MAG. SPACE 145

C SIMPLISTIC SPECTRAL VARIABILITY MODEL 149

D THE 3D-HST EMISSION LINE MAPS 151

LIST OF FIGURES 155

LIST OF TABLES 159

ACKNOWLEDGEMENTS 161

BIBLIOGRAPHY 164

INDEX 177

CHAPTER 1

INTRODUCTION

The aim of this introduction is not to give a broad and general introduction to the vastly different and overwhelmingly many interesting topics of ancient as well as modern astrophysics. The aim of this introduction is rather to set the stage of Chapters 2–4 which contain the bulk material of this thesis. After taking a stroll down astronomy's memory lane, we present the standard cosmological model, which is the foundation of modern astronomy. Within this theoretical framework, we describe the build-up of galaxies including the formation of stars in merging dark matter potential wells. We will focus on two particular epochs of the evolution of galaxies as we set the stage for the main chapters. Firstly, we focus on the build-up of galaxies just 3–4 Gyr after the Big Bang ($z \sim 1.5$), where the vast majority of the present day stars formed within galaxies and in violent collisions of galaxies. Secondly, we focus on the active phase of a galaxies life cycle, where galaxies are bright enough, that they can be observed out to $z = 7$, providing an important probe of the very early Universe, the Universe's large scale structures, the intergalactic medium, as well as the physics of the central regions of individual (active) galaxies.

Apart from in-text references this introduction is based on Longair (1998); Freedman & Kaufmann (2002); Binney & Tremaine (2008) and Mo et al. (2010).

1.1 OUR PLACE IN THE UNIVERSE

Astronomy, being one of the oldest professions in the world, has a long and glorious history, which this short introduction will by no means give justice. Nevertheless, a short description of where the science of astronomy/astrophysics come from is in its place. Astronomy has played a crucial part in the oldest societies known to man, from the Mayan empires over the greek cradle of democracy to the persian and babylonian ancient people. The breathtaking starry night has by many been thought of as godly and/or holy and has inspired people to observe it, describe it and map it.

Up through time the surrounding Universe has not only been amazing and fascinating generations, it has also been a constant cause for disputes. It was long thought that Ptolemy's geocentric second century picture of the Universe, where stars and planets all revolve around the Earth, as illustrated in the two left panels of Figure 1.1, was the correct model. If you proposed otherwise, you were ridiculed or punished. In the renaissance people slowly started to accept that there might be more to the understanding of the Universe than a scheme with the earth in the center and everything else revolving around it. With scientists like Copernicus, Brahe, Kepler, and Galilei significant steps were taken away from the geocentric world picture towards a heliocentric picture as first described by Copernicus, with the planets moving around the Sun on elliptical orbits as described by Kepler (1609). Brahe actually did not believe the heliocentric picture, but his outstanding naked eye(!) observations led Kepler to his discoveries. In the right panel of Figure 1.1 the heliocentric picture is portrayed as Copernicus imagined it. Moving from the geocentric to the heliocentric world picture was based on observations of the night sky, both with and without telescopes, and the eager to understand and describe these observations by mathematics. With Newton's *Philosophiae naturalis principia mathematica* (Newton 1760), where he describes the laws of gravitation and *derives* Kepler's empirical laws of planetary motion for the first time, astronomy entered a new era.

In very much the same way, it is this fascination and urge to describe observations of the Universe by mathematics and physics that drives many astronomers today. Taking part in the observing, detailed theoretical description, and mapping of the magnificent Universe around us is a privilege and what makes this profession so great. With this eager to discover and explore the unknown, we have come a long way since the mayans, the babylonians, the greeks, and the astronomers of the renaissance. Not *despite* of them, but *because* of them, we today believe that we understand how it all started, what the foundation of it all is, and how it evolved into what it is today. As Bernard of Chartres (and later Sir Isaac Newton) puts it:

> *We are like dwarves perched on the shoulders of giants, and thus we are able to see more and farther than the latter.*

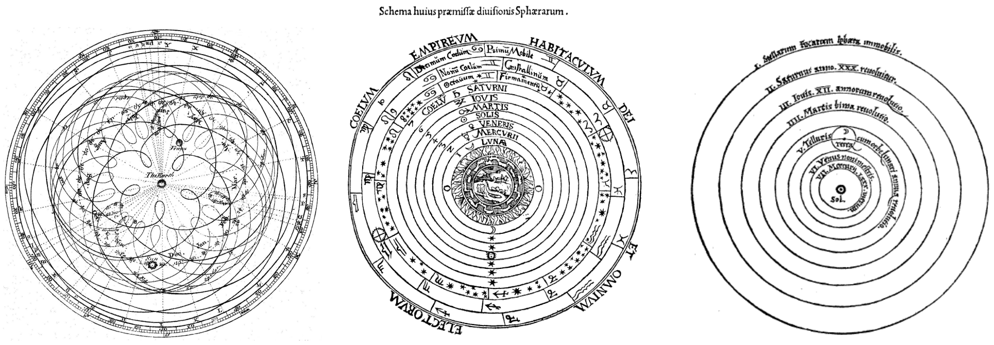
Schema huius præmissæ diuisionis Sphærarum.
EMPIREVM HABITACVLVM DEI ET OMNIVM ELECTORVM
SATVRNI
IOVIS
MARTIS
SOLIS
VENERIS
MERCVRII
LVNÆ
Sol.

FIGURE 1.1: *The traditional geocentric picture of the Universe as described by Ptolemy in his second century Almagest (left) and in a 'modern' sixteenth century version (center). Copernicus' heliocentric model was first presented and illustrated in "On the Revolutions of the Celestial Sphere" (1543) as shown in the right panel. This model changed the view of our place in the Universe – we were no longer the center of attention.*

1.2 THE STANDARD BIG BANG MODEL

People slowly started to acknowledge that the earth was not the center of the solar system, let alone the Milky Way Galaxy or the Universe as a whole, and mathematicians, physicists and astronomers were able to theoretically describe the empirical models that existed of larger and larger portions of the observable Universe. With the Messier catalog (re-published in Messier & Niles 1981) and Herschel's Catalogue of Nebulae (Herschel 1786) in the late eighteenth century and later Dreyer's New General Catalog (Dreyer 1888), it became evident that the sky was filled with objects, that were not 'just' stars. Though it had been speculated by several people that these objects were 'island' Universes similar to our own Milky Way at extreme distances, it was not until the work of Hubble in the early 1920s (less than 100 years ago!), that it was finally shown that they were extragalactic. In 1925 Hubble for the first time used Cepheid's relationship between period and absolute magnitude to show that some of the major known nebulae, i.e., galaxies, were indeed extragalactic (Hubble 1925a,b). In Hubble (1926) he showed that extragalactic objects were spread all over the sky and derived the first empirical relations for these objects. Hubble showed that these extragalactic galaxies were all receding from the Milky Way with a velocity strongly dependent on their distance to us (Hubble 1929). This relationship can be formulated in Hubble's law:

$$v = H_0 \times d \,. \tag{1.1}$$

Here H_0 is the so-called Hubble constant (at present indicated by $_0$), v the receding velocity of the galaxy and d the distance to the galaxy.

Thirteen years earlier Einstein presented his general theory of relativity (Einstein 1916) generalizing the special relativity theory and Newton's laws of gravitation via his field equations (Einstein 1915) providing a unified theory for the behavior of energy, i.e., mass in the spacetime continuum. In Einstein (1917) this theory was put into a cosmological context, and how to obtain stable solution to the field equations was discussed. Einstein created stable solutions by introducing the cosmological constant, Λ, to keep the Universe (artificially) static. Einstein later regretted having introduced Λ, however, as we will describe below, this cosmological constant became an important part of our present understanding of the Universe in the 1980s (though introduced via different reasoning). De Sitter (de Sitter 1917, 1918) and Friedmann (Friedmann 1922, 1924) later showed that stable solutions to Einstein's field equations did indeed exist for expanding as well as stationary models of the Universe, also without introducing the artificial Λ. Hubble was aware of this work in 1929 and therefore noted that the linear distance-velocity relationship from Equation (1.1) might be a first approximation to the actual movement and behavior of the extragalactic nebulae. At cosmological distances Hubble's law is indeed model dependent, but the Hubble expansion in Equation (1.1) is nevertheless a fundamental feature of the local Universe.

Having established both theoretically and observationally in the 1920s that the Universe was expanding, it became obvious that in the very distant past the Universe must have been much denser, than it is today. In fact, with enough time available it would at some point have been considered a physical singularity, i.e., it would have been so dense and hot that all physical models would brake down. This singularity has since the early 1950s been known as the *Big Bang*. The time needed for the Universe to evolve from this singularity into a Universe with the observed expansion rate is simply obtained by inverting Hubble's law in Equation (1.1). This results in an approximate age of the Universe, called the Hubble time

$$t_0 \equiv \frac{1}{H_0} = \frac{d}{v} \sim 13.7 \text{ Gyr} \; . \tag{1.2}$$

Up through the 40s and 50s people like Gamow, Dicke, Alpher & Herman speculated that this Big Bang over-density should be observable as a smooth background black-body radiation, and they estimated it to have a temperature of the order 10s of Kelvin. In 1965 the cosmic microwave background (CMB) was serendipitously observed for the first time as a spurious 3.5 K background as presented in Penzias & Wilson (1965) and Dicke et al. (1965). This was the second observational evidence (Hubble's distance-velocity relation in Equation (1.1) being the first) of the Big Bang model of the Universe. With the Cosmic Background Explorer (COBE; Mather et al. 1990; Smoot et al. 1991; Bennett et al. 1996) and Wilkinson Microwave Anisotropy Probe (WMAP; Hinshaw et al. 2009; Jarosik et al. 2007; Bennett et al. 2003; Jarosik et al. 2011) the CMB has later been observed in amazing detail and has backed up these initial findings. In Figure 1.2 the CMB maps from the COBE 4 years data (left) and the WMAP 5-years data (right) are shown.

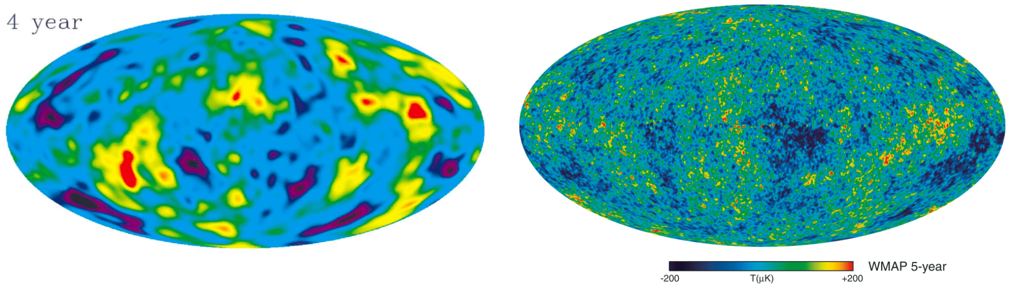

Figure 1.2: *The cosmic microwave background (CMB) from the COBE 4-years data (left; Bennett et al. 1996) and WMAP 5-years data (right; Hinshaw et al. 2009). The color scheme depicts the 13.7 Gyr old temperature (density) fluctuations, which after inflation grew via hierarchical merging and condensation into the galaxies we know and love. This radiation is the left-over from the singularity we call Big Bang and is one of the strongest observational proofs of the standard Big Bang model.*

Previously it had been a challenge to explain the high abundance of metals (primarily He) in the local Universe via Hydrogen burning in stars only, so when Hoyle & Tayler (1964) managed to describe the observed He/H ratio within a factor of two, by assuming that the Universe had been a singularity in the past, the road was paved for the general acceptance of the evolutionary Big Bang model.

1.3 Structure Formation & (Λ)CDM Building Blocks

Apart from the particle physicists' interest in the formation of the first heavy elements, also a more general theoretical interest in describing the Big Bang model grew in the twentieth century. Many such attempts assumed Einsteins's general relativity to be true. One of the preferred models, in the form of Friedmann's solutions to Einstein's field equations mentioned above, made some strong assumptions. The model assumed that the Universe is isotropic and homogenous. These two main assumptions have come to be known as the *cosmological principle*. Even though this was just a cosmological restatement of the thoughts behind Corpenicus' heliocentric picture depicted in Figure 1.1, i.e., there is nothing special about our location in the Universe, it still seemed odd that this should be the case, now that it was clear that the observable Universe was definitely not isotropic and homogenous. Thus, any theoretical model of how the CMB temperature fluctuations over time turn into the observed galaxies, needed to obey these assumptions if Friedmann's equations were to be right.

Already in Jeans (1902) it was described how a stable structure (in this particular case a spherical nebula) can become unstable and collapse under its own gravity. This happens if the size of the initial density perturbations are larger than the Jeans length

$$\lambda_J = \frac{c_s}{\sqrt{\frac{G\rho_0}{\pi}}} , \tag{1.3}$$

where G is Newton's gravitational constant and c_s is the sound speed propagating in the 'medium' of mass density ρ_0. This equation simply states that if the spatial (Jeans) scales are large enough the gravitational attraction can overcome the (sound) pressure and collapse.

Concerned with only spherically symmetric perturbations Lemaître (1931, 1933), Tolman (1934) and later Bondi (1947) applied this line of thought to the expanding Universe, assuming the cosmological principle to be true. It was generalized by Lifshitz (1946), who studied the evolution of the assumed infinitesimal temperature (density) perturbations in the framework of the Friedmann model. From these different studies the authors however arrived to the conclusion that the observed galaxies could not have formed from gravitational collapse, since the Hubble time, t_0 (Equation (1.2)), was simply too short to allow the assumed infinitesimal perturbations to collapse into structures as massive as galaxies.

In the 1960s various authors among those Zel'Dovich & Novikov (1967) and Peebles (1968, 1971) allowed the initial density perturbations to be non-infinitesimal, and developed a framework in which structures could collapse, form filaments of over-dense regions (leaving under-dense voids behind), and evolve into the large scale structure of the Universe, which was observed. In the 70s generally two theories came to dominate: the adiabatic and the isothermal model. The adiabatic picture (e.g., Doroshkevich et al. 1974) describes the perturbations as adiabatic sound waves, i.e., sound waves not exchanging energy with the surroundings, which starts to fragment into smaller and smaller structures, as the density contrast of the surroundings changes. Hence, the adiabatic picture is, so to speak, a 'top-down' picture of structure formation, where large structures form first and then fragment into smaller structures. The isothermal picture, on the other hand, is a 'bottom-up' picture, where small structures form first and then merger to grow into larger and larger structures. It describes how isothermal perturbations in balance with the pressure collapse and form larger structures via hierarchical clustering and merging of over-densities. The discovery of the CMB in 1965 served as a direct observational test of the initial conditions of these theories of structure formation. As the CMB measurements improved it became clear that the size of the observed perturbations were simply too small to have been the primordial seeds of baryonic matter, that developed into the massive galaxies observed in the nearby Universe within the age of the Universe, i.e., within t_0. The main problem was that the baryons needed to cool significantly before they could collapse (or fragment) due to the counteracting radiation pressure in Jeans' theory. However, along with the development of the theories for baryonic structure formation up through the 70s, the concept of *dark matter* was discussed extensively as possibly playing a major role in the picture of structure formation – and indeed, dark matter turned out to be the key for solving the t_0 time-scale issue.

Dark Matter, predicted already in the early 1930s (Oort 1932; Zwicky 1933), consists of non-baryonic particles neither emitting nor reflecting electromagnetic radiation. It was initially introduced into structure formation as neutrinos with non-zero mass, later referred to as *hot* dark matter due to the neutrinos' relatively low masses. It was however realized that several other particles, some more exotic than others, but all weakly interacting and much more massive particles (WIMPs) than the neutrinos, could also serve as non-baryonic dark matter[1]. In the early 1980s the models for this *cold* dark matter (CDM)[2] were presented (e.g., Peebles 1982; Faber 1984; Primack & Blumenthal 1984). Since dark matter is weakly interacting, or in other words collisionless, the gravitational instability of the CDM perturbations is not hampered by a counteracting radiation pressure. This was the problem baryonic structure formation was facing, and it therefore made the dark matter structure formation picture more attractive.

In Press & Schechter (1974) it had several years earlier been described how the mass function of objects experiencing a hierarchical structure formation, i.e., a bottom-up picture similar to the isothermal structure formation from the same period, under the assumption of an initial gaussian distribution of density perturbations, evolves as a function of time. This theory proved extremely useful for describing the buildup of the CDM hierarchical Universe and was later improved by Lacey & Cole (1993) taking detailed observations of the CMB into account[3]. In general, at very early stages the initial gravitational collapse is linear, i.e., the CDM perturbations are slowly collapsing linearly into larger and larger over-densities, which the baryons relentlessly follow. When these over-densities become large enough, the evolution enters a non-linear regime, where the dark matter potential wells collapse to form filaments and voids mapping the CDM 'potential well landscape' that baryons can condense into to form the observable galaxies. In Section 1.4 we will describe this 'condensation' in a bit more detail. Having established the general picture of structure formation, it was no longer as problematic to state that the Universe was homogeneous and isotropic, i.e., to assume the cosmological principle to hold. Since, if all structures form by the collapse of gaussian perturbations in a large uniform 'fluid' or plasma of particles, it is not hard to imagine that a some characteristic scale the cosmological principle is indeed true. This characteristic scale has been shown by extensive observational efforts (e.g., Shectman et al. 1996; Colless et al. 2001; Stoughton et al. 2002) to be of the order $\sim 100 - 200$ Mpc (Longair 1998; Ryden 2003; Binney & Tremaine 2008), i.e., the Universe is indeed homogeneous and isotropic and the cosmological principle therefore holds.

---

[1] In principle brown dwarfs, black holes and neutron stars, all baryonic, are also dark matter 'particles' and are usually referred to as MACHOs (massive compact halo objects). However, these were not present in the very early Universe, so here we let the term dark matter refer to the non-baryonic WIMPs only.

[2] Also *warm* dark matter schemes were introduced, but both the hot and warm dark matter models have proven much less successful than the cold dark matter models, and was mostly popular in the late 70s and the 80s. The terms hot, warm and cold refer to the masses of the dark matter particles (CDM being the heaviest), and the epoch at which the particles became relativistic. For further information on the different types of dark matter see Faber (1984); Ryden (2003); Longair (1998) or Mo et al. (2010).

[3] Lacey & Cole (1993) combined the theory for the collapsing of dark matter structures with the assumption that the CMB fluctuations can be described by a primordial power-spectrum, as confirmed by COBE in 1992 (Smoot et al. 1992) and predicted by the Harrison-Zeldovic power spectrum (Harrison 1970; Zeldovich 1972).

Despite the many successes of the standard Big Bang model, there were still unsolved issues that cosmologists in the 1970s were very puzzled by. To obtain the high degree of isotropy that the CMB predicts, the initial conditions of the Big Bang need to be carefully adjusted, fine-tuned, and arranged such that the Universe evolves into what is observed. The five main unsolved issues or problems with the Big Bang model are often referred to as the flatness, horizon, baryon asymmetry, large scale structures and monopole problem.

The flatness problem concerns the fact that the Universe seems to have an energy density very close to the critical density, preventing the geometry of the Universe to be curved, i.e., the Universe is flat. It appears that the Universe was flat to within a factor of $\sim 10^{-16}$ when it was already 1 second old (Kolb & Turner 1990), which is indeed heavy fine-tuning.

As mentioned the CMB was observed to be very uniform over the full sky. However, by simply rewinding the clock, every point in the sky could not have been in causal contact and therefore did not have the chance to adjust their temperature in the early Universe if only t_0 was available. In other words the *horizon* of each point in the sky is too small for them to have been in causal contact, so how come the full-sky CMB map is so uniform?

The baryon asymmetry problem deals with the fact that there is an asymmetry between baryons and photons in the Universe today. For this asymmetry to survive the hot early phase of the Big Bang model, there must have been a baryon–anti-baryon asymmetry at early times, which insured that enough baryons were left for the electron-positron pair production.

The large scale structure problem questioned whether there was actually enough time for matter to condense and collapse, as described above, to form the observed large scale structures of the Universe within t_0.

Lastly, the so-called *grand unified theory* of particle physics where the electromagnetic, strong, and weak forces were combined in the early Universe, predicts the presence of so-called magnetic monopoles and cosmic strings. Since these have not been observed, their volume density must be approximately 0 at present, which is the essence of the monopole problem. Why are they not here anymore? Thus, all five problems question the initial conditions and consequences of the Big Bang model.

In Guth (1981) and Linde (1982) the inflationary scenario of the early Universe was presented as a solution to these problems. The authors suggested that when the Universe was of the order 10^{-35} seconds old it went through and exponential inflation. Through several e-foldings the geometry of the Universe was driven towards flat, the size, i.e., the event horizon was heavily expanded, the baryon-anti-baryon asymmetry was created, the seeds for the large scale structures were 'frozen out', and the density of magnetic monopoles was diluted to almost 0. The inflation therefore solved the five main problems of the Big Bang models.

The inflationary scenario also predicts that the cosmological constant Λ is significantly different from 0, and thereby re-introduces the cosmological constant into Einstein's field equations; though through different reasoning than Einstein's. In the field equations the terms concerning Λ can be thought of as an energy density, and today the astronomical community refers to the cosmological constant contribution as the *dark energy* of the CDM model. Since the early 1980s inflation has been a crucial part of the modern cosmological picture and the formation history of the Universe, and with the introduction of inflation and dark energy, the CDM model is now referred to as the ΛCDM model. For any further details on the individual problems of the pre-inflation Big Bang model or inflation itself we refer the reader to Kolb & Turner (1990); Longair (1998) and Ryden (2003).

Despite the extensive observational efforts that supported ΛCDM (CMB, the Hubble expansion, the lack of monopoles and cosmic strings, baryon-photon asymmetry, large scale structures, etc.)[4] it wasn't before the late 1990s that the ΛCDM picture was generally accepted. This happened when several authors, using distant supernovae, showed that the expansion of the Universe is indeed accelerating as predicted by the dark energy cosmological constant contribution to the theory (Garnavich et al. 1998; Schmidt et al. 1998; Riess et al. 1998; Perlmutter et al. 1999). Through these and the CMB efforts, among others, it has been established that the global energy densities of the ΛCDM building blocks, i.e., visible baryonic matter, dark matter and dark energy are remarkably different. It appears that the energy density of the Universe consists of only 4% baryons(!), 23% dark matter and 73% dark energy (e.g., Jarosik et al. 2011).

Today ΛCDM with its dark energy component is referred to as the standard model of Big Bang cosmology and structure formation. Hence, today we describe the Universe as having started from a Big Bang (singularity) with perturbations mainly consisting of dark matter, running through a short phase of exponential inflation at early times, after which the dark matter halos started collapsing under their own gravity and grew in a hierarchical manner via merging. This landscape of dark matter potential wells then acted as 'hosts' for the baryons, which condensed into galaxies and started forming the stars, from which we gain the vast majority of our knowledge about the Universe.

---

[4] Some of these observations admittedly supports ΛCDM because the theory was build to fit them, but nevertheless.

1.4 BUILDING A GALAXY

There are two main scenarios through which a galaxy is believed to be able to form from dark matter halos (potential wells) and baryons: through hierarchical merging of halos and the associated baryons, and via a monolithic collapse of the baryons in the potential wells.

The hierarchical merging is a direct consequence of the structure formation picture outlined above, where the merging of dark matter structures drags along the baryons and enable these to merger with them. Already in Toomre & Toomre (1972) is was made clear, that some of the observed peculiar galaxies were well described by a merger scenario. Generally, when galaxies merge the end result is a structure with a broad distribution of internal velocities, such that most orbits around the systems center of gravity are occupied.

The monolithic collapse of the baryonic gas in (isolated) galaxies, on the other hand, often results in a preferred rotation of the system. Since baryonic matter can loose its energy by radiating in ways dark matter cannot (see Section 4.2), the gas residing in the dark matter potential wells will over time loose its energy and slowly sink or condense into the bottom of the potential well (White & Rees 1978; Efstathiou & Silk 1983). Since angular momentum needs to be conserved, this results in a rotating system where the gas often flattens to a disk-like structure with a preferred direction of rotation. Such a fairly equilibrated system can of course be destroyed or stirred up if it encounters the gravitational effects of another galaxy, or in the most extreme case becomes part of a galaxy merger.

Hence, the merging of dark matter halos and their galaxies combined with the condensation of gas in (isolated) dark matter potential wells are the main drivers of building galaxies, and are able to reproduce the observed population of galaxies. Three classes of galaxies are know: the ellipticals, the spirals and the irregulars. These are often divided into subclasses, or Hubble classes, as illustrated in Figure 1.3. These classes address the appearance of the galaxies, like eccentricity of the ellipticals, whether a bar is present or not in the spirals, and how tightly wound the spiral's arms are. The formation of galaxies via hierarchical merging preferentially produces ellipticals, whereas the main end-product of a monolithic collapse is a nicely rotating spiral galaxy. The irregular galaxies consist of everything belonging to neither the elliptical nor the spiral Hubble classes, e.g., merger remnants not yet equilibrated or odd-looking truly irregular galaxies.

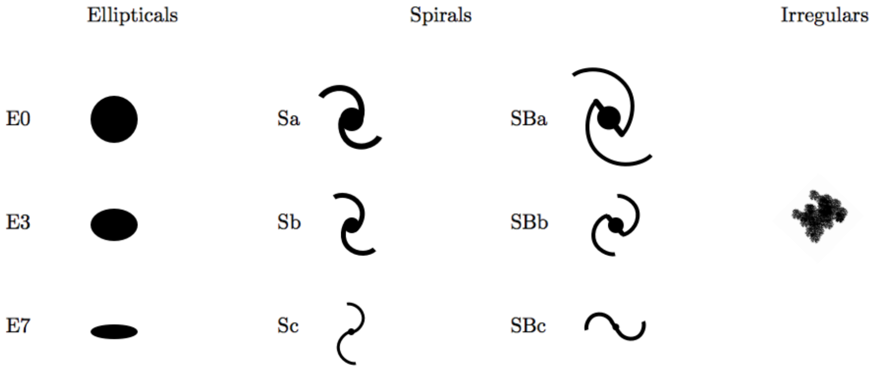

FIGURE 1.3: *A schematic overview of the Hubble galaxy classes. The left panel shows the elliptical classes of galaxies. They are designated by Ex, with x giving the (observed) eccentricity. The two center panels show the sub-division of spiral galaxies. Spirals are designated with Sα where α refers to how tightly wound the spiral arms are, with 'a' referring to the most tightly wound arms. In the case of prominent bars a B is added. The right hand panel represents irregular galaxies, which consist of everything (including recent merger events) that cannot be categorized as either elliptical (Ex), spiral (Sα), or barred spiral (SBα). Ellipticals are the end product of galaxies formed in recent merger events, whereas spirals need an extended epoch of monolithic collapse of gas to conserve the preferred direction of rotation and the spiral pattern.*

It is not only the appearance and the internal velocities of these different galaxies that are significantly different. Also the stellar populations of ellipticals, spirals and irregulars are distinctly different. Since most ellipticals are believed to be the equilibrated end-product of a merging event, most of the stars in the final elliptical were already present prior to the merger, i.e., they are relatively old. On the other hand the condensation of gas in spirals induces new star formation, and hence the age spread of stars in spirals is much broader than in ellipticals. If irregular galaxies are recent mergers that have not yet equilibrated into ellipticals, they might have experienced a recent burst of star formation induced by the collision of gas during merging. Irregulars therefore have a fairly young population of stars. Thus, the population of stars, i.e., the star formation and the merging activity in galaxies is closely linked to each other and are important factors when dealing with galaxy formation and evolution. The epoch where the global star formation rate in the Universe peaked has been estimated to be at $1 < z < 3$ (Ferguson et al. 2000; Heavens et al. 2004). This agrees well with the peak of galaxy merging (Conselice et al. 2008), and therefore supports this picture. We will address the interplay between merging and star formation at $1 < z < 3$ further in Chapter 4 and in particular in Section 4.2.

It is not only bursts of star formation the merging of galaxies seems to trigger. Galaxy mergers are also thought to be able to trigger the most violent phase of a galaxies life, the active (quasar) phase, by enabling large amounts of gas to accrete onto a galaxies central black hole.

1.5 GALAXIES' ACTIVE (QUASAR) PHASE

So far most of the galaxies described have been relaxed equilibrated systems quietly residing in the bottom of cold dark matter potential wells, slowly forming stars from their gas reservoirs. However, this state of peace and quit does not last forever. It is believed that most, if not all, galaxies in their lifetime of several billion years experience one or more short active phases lasting for only $\sim 10^7$ years (e.g., Haiman & Hui 2001; Martini & Weinberg 2001; Hopkins et al. 2005a,b). Galaxies in this phase of their evolution are referred to as active galaxies or 'AGN'. Strictly speaking AGN refers to the cause of the active phase, namely the Active Galactic Nucleus and not the actual galaxy itself. However, it has become standard to refer to active galaxies by this acronym, which we will therefore also do here.

AGN are among the most luminous and powerful objects in the sky. The majority of AGN are known to vary by a few percent over weeks, months and years to the more extreme cases of several percent within a few nights (e.g., Pollock et al. 1979; Webb et al. 1988; Schramm et al. 1994; Giveon et al. 1999; Bauer et al. 2009a or Ulrich et al. 1997 for a review). That some AGN vary on these short time-scales indicate that the variability originates from very small regions of sizes from light-months down to a few light-days(!) across. This early led to the conclusion, that something extremely small and massive must reside at the center of active galaxies (Spitzer & Stone 1967; Spitzer & Saslaw 1966; Lynden-Bell & Wood 1968). It is generally accepted that a supermassive black hole (SMBH) lives at the center of most, if not all, galaxies.

Several classes of AGN exist. Among these are Seyfert 1 and 2 galaxies named after their discoverer (Seyfert 1943), broad and narrow line radio galaxies (BLRG and NLRG respectively), optically violently variable (OVV) galaxies, BL Lac galaxies, Blazars, low-ionization nuclear emission line regions (LINERs), quasi stellar objects (QSOs), and quasars. Some of these names are descriptive for the object class, whereas others are named for historical reasons. Furthermore, the classes are far from clearly distinct. For instance LINERs are thought to be a low-luminosity extension of the Seyfert galaxies, quasar and QSO nowadays refer to the same class, even though they were originally distinct classes, and Blazars are supposed to be the combination of OVVs and BL Lac galaxies. This makes the 'zoology' of AGN a rather complicated matter. Nevertheless, we attempt to summarized the key features of the different kinds of AGN in Table 1.1. This table is by no means exhaustive but gives a good flavor for what characterizes and distinguishes different AGN types.

Table 1.1: *The Main AGN Types and Their Characteristics*

	Seyfert†		Quasar		Radio Galaxy		Blazar	
	1	2	RL	RQ	BLRG	NLRG	OVV	BL Lac
Broad Lines ($\sigma_v > 1000$ km/s)	+	–⋆	+	+	+	–	+	–
Narrow Lines ($\sigma_v \sim 100$ km/s)	+	+	+	+	+	+	+	–
Number Density [Mpc^{-3}]	10^{-4}		10^{-7}		10^{-6}			
M_B (classic definition)	< -21.5 $+5 \log h$		> -21.5 $+5 \log h$		–		–	
Strongly Polarized > 5% of light							+	+
Significant Radio Emission			+		+	+	+	+
Variable > 1%-level	+	+	+	+	+	+	+	+

† *LINERS are believed to be a low-luminosity extension of the Seyferts.*

⋆ *NGC1068 is an example of a classic Seyfert 2 galaxy, where polarized broad lines are actually observed. This might indicate the presence of an obscured/hidden BLR (Tran 1995) supporting the unified model of AGN illustrated in Figure 1.4.*

The sometimes opaque or muddied classification scheme of AGN has given rise to the idea, that the different classes of AGN are in principle the same object perceived differently by the observer due to viewing angle. This unified model for AGN (Antonucci 1993; Urry & Padovani 1995) is still a matter of debate (e.g., Elvis 2000; Almeida et al. 2011; Trump et al. 2011), but has to some extent made the AGN classification more digestible. In Figure 1.4 we illustrate how the different AGN classes from Table 1.1 are supposedly observed according to the unified AGN model.

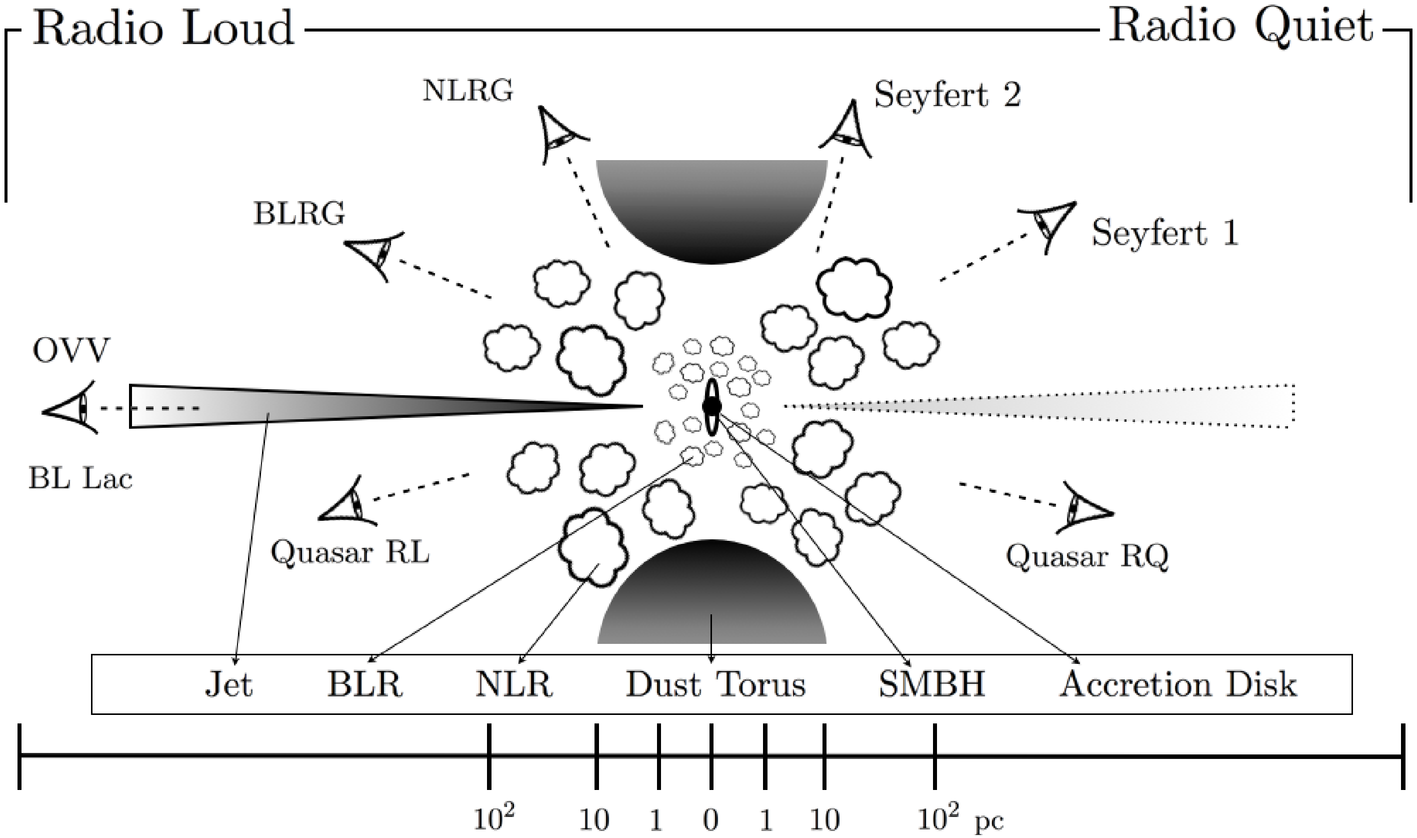
Radio Loud
Radio Quiet
NLRG
Seyfert 2
BLRG
Seyfert 1
OVV
BL Lac
Quasar RL
Quasar RQ
Jet
BLR
NLR
Dust Torus
SMBH
Accretion Disk
10^2
10
1
0
1
10
10^2 pc

FIGURE 1.4: *Illustration of the inner region of an active galaxy. The classes of active galaxies from Table 1.1 are indicated by the viewing angles the unified AGN model predicts them to be observed from (Antonucci 1993; Urry & Padovani 1995). The approximate length-scales are indicated at the bottom. The left hand side contain galaxies with pronounced radio emission, whereas the right hand side galaxies have more modest radio luminosities. Note that the validity of the unified AGN model is still being debated (Elvis 2000; Almeida et al. 2011; Trump et al. 2011).*

AGN are believed to get their power from the SMBH residing in the center of the galaxy. It is believed that the SMBH is swallowing material which radiates strongly, as it falls into the deep potential well of the SMBH from a surrounding accretion disk. This *central engine* (which we will describe in more detail in Section 2.2) is part of the 6 main components of the unified AGN model illustrated in Figure 1.4. These components are:

- **Central supermassive black hole (SMBH)**: The SMBH is assumed to inhabit the bottom of the galaxy's dark matter potential well and be the key part of the central AGN. The SMBH in the (non-AGN) Milky Way has been monitored extensively (Narayan et al. 1995; Schödel et al. 2002; Ghez et al. 2008) and has an estimated mass of $\sim 10^6 M_\odot$. This is two orders of magnitude smaller than the average size of AGN SMBHs of $\sim 10^8 M_\odot$ (e.g., Woo & Urry 2002; McLure & Dunlop 2002; Vestergaard 2002; Häring & Rix 2004).
- **Accretion disk**: A flattened disk of gas and stars rotating around the SMBH. Matter being accreted onto the SMBH from this disk is thought to be the main source of radiation for the extreme AGN luminosities. The accretion disk and the SMBH make up the central AGN engine which is described in Section 2.2.
- **Broad line region (BLR)**: A small inner region believed to be less than a few light-years across, composed of individual high density clouds ($\rho \sim 10^{10}\mathrm{cm}^{-3}$) giving rise to the broad emission line features observed in most AGN spectra. The temperature of the BLR is expected to be $T_{\mathrm{BLR}} \sim 10^4\mathrm{K}$.
- **Narrow line region (NLR)**: A wider ($\sim 50-100\mathrm{pc}$) region giving rise to the narrow lines in the quasar spectra, such as [OIII]. The density of the NLR is expected to be ($\rho \sim 10^6\mathrm{cm}^{-3}$). The temperature in the NLR is the same as in the BLR, i.e., $T_{\mathrm{NLR}} \sim 10^4\mathrm{K}$.
- **Dust Torus**: An obscuring dusty torus believed to block the light from the BLR and NLR for near-edge-on galaxy views.
- **Jet**: Along the rotational axis of the central engine, powerful radio jets (most prominent in the radio loud AGN) form from the synchrotron radiation of the material being accreted onto the SMBH. The most powerful and most variable AGN are seen 'down the barrel' of these jets.

These individual components of the AGN model give rise to some of the characteristic features in AGN spectra (a standard composite AGN (quasar) spectrum is shown in Figure 3.1 on page 63). We will described the specific features of the quasar spectrum and how they can be used to probe the physics governing AGN in Section 3.2. For a more detailed discussion of the individual components of AGN and their spectra please refer to, e.g., Antonucci (1993); Urry & Padovani (1995); Krolik (1999) and Mo et al. (2010).

Despite the extreme differences in length scales of the different components of the AGN (and galaxies in general), from sub-parsec scales of the central SMBH and accretion disk to the several kpc of the bulge and outer regions of the galaxy itself, there seems to be a remarkable tight correlation between the size of the central black hole and the galaxy observables like for example luminosity, the velocity dispersion and the (bulge) mass of the stars, and light concentration (e.g., Kormendy & Richstone 1995; Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Graham et al. 2001; McLure & Dunlop 2002; Tremaine et al. 2002; Marconi & Hunt 2003; Häring & Rix 2004; Jahnke et al. 2009). For example several of these studies find that the black hole mass is tightly correlated with the bulge mass such that $M_{\mathrm{SMBH}} \sim 0.003 \times M_{\mathrm{bulge}}$. In Bandara et al. (2009) they argue that $\log(M_{\mathrm{SMBH}}/M_{\odot}) \sim 8.18 + 1.55 \times (\log(M_{\mathrm{tot}}/M_{\odot}) - 13.0)$ with M_{tot} being the total galaxy (halo) mass, i.e., they argue that a standard AGN SMBH of $10^8 M_{\odot}$ approximately corresponds to $10^{-5} M_{\mathrm{tot}}$. Whether these scaling relations are due to a causal co-evolution of the galaxy and its SMBH across the vastly different length scales or a simple consequence of ΛCDM hierarchical merging is debated (Peng 2007; Jahnke & Maccio 2010). These correlations can be used to constrain evolutionary models of AGN as exemplified in Section 2.2.

1.5.1 QUASARS – AGN AT ITS BEST

One of the most powerful and brightest kind of AGN is the *quasar*. Quasars[5] have extreme luminosities corresponding to absolute magnitudes of $M_r \sim -24$ (Richards et al. 2005) making quasars the second brightest objects in the sky only outshone by a few gamma ray bursts as illustrated in Figure 8 of Bloom et al. (2009). Quasars are observed out to redshifts of 6 (Fan 2006; Fan et al. 2001b; Willott et al. 2010), and even at redshift 7 they have been found (Mortlock et al. 2011b), showing that they had already formed and started shining less than a 1 Gyr year after the Big Bang. This puts strong constraints on theories trying to describe the physics of quasars and AGN as described in Section 2.2.

Even though quasars have been observed at $z \sim 7$ the distribution of quasars peaks at redshifts 2–3 (e.g., Richards et al. 2009). Also the *co-moving* density of quasars peaks a these redshifts (Shaver et al. 1996; Boyle et al. 2000; Fan et al. 2001b; Croom et al. 2009a). The fact that quasars were more common at earlier epochs speaks in favor of the scenario, where many AGN and quasars are triggered by gravitational interaction and merging of galaxies mentioned in Section 1.4. Since the merger rate between galaxies was higher, than it is today, at these redshifts, naturally more quasars would be observed per co-moving volume if they are a consequence of interactions. That the average quasar density drops at higher redshift is explained by the fact that in a ΛCDM Universe there were very few galaxies with masses high enough to host the SMBHs of quasars at $z > 3 - 4$.

---

[5] Referring to a quasar as an object is strictly speaking wrong, since it is not an object in itself but rather a galaxy in its active (quasar) phase of evolution. Nevertheless, as we have referred to AGN as objects we will also in the reminder of this thesis refer to quasars as astronomical objects – but it has to be kept in mind that quasars (and AGN) are strictly speaking *not* objects in themselves, but rather a particular phase of a galaxy's life.

Galaxies in their active quasar phase will be the main topic of Chapters 2 and 3 and will be described in more details in these chapters.

1.6 OUTLINE OF THIS THESIS

This thesis will concern the nature and evolution of quasars and the star formation in mergers at $z \sim 1.5$. The thesis contains three major parts represented by Chapters 2–4. They are presented in chronological order and not (necessarily) in the order the described physics would be important in the formation and evolution of a galaxy, as outlined in this introductory chapter. Chapters 2–4 are based on Schmidt et al. (2010), Schmidt et al. (2012a) and Schmidt et al. (2012b), respectively. At the beginning of each of these chapters a small abstract will summarize the key results and a prologue will set the stage and pose the questions to be addressed in the respective chapter. After the bulk content of each chapter an outlook describes the prospects of the results and give some suggestions on how to take advantage of these in the future. Each chapter is ended by a short epilogue which will try to pick up any loose ends and summarize before the thesis advances to the next chapter. In this way, the hope is that the pro- and epilogues distills the bulk content of this thesis into 6 short, concise and easily digestible sections.

In Chapter 5 the individual chapters, their results and their individual outlooks are summarized and boiled down to the essentials. We also present a short outlook on the topic of galaxy formation and evolution in general before we finish with a short look into the immediate future of ‘yours truly’.

CHAPTER 2

SELECTING QUASARS VIA VARIABILITY

In this Chapter we present a new and simple technique for selecting extensive, complete and pure quasar samples, based on their intrinsic variability. We parametrize the single-band variability by a power-law model for the light-curve structure function, with amplitude A and power-law index γ. We show that quasars can be efficiently separated from other non-variable and variable sources by the location of the individual sources in the A–γ plane. We use ~60 epochs of imaging data, taken over ~8 years, from SDSS Stripe 82, to demonstrate the power of variability as a quasar classifier in multi-epoch surveys. For UV excess selected objects, variability performs just as well as the standard SDSS color selection, identifying quasars with a completeness of 90% and a purity of 95%. In the redshift range $2.5 < z < 3$, where color selection is known to be problematic, variability can select quasars with a completeness of 90% and a purity of 96%. This is a factor of 5-10 times more pure than existing color-selection of quasars in this redshift range. Selecting objects from a broad *griz* color box *without* UV excess information, variability selection in Stripe 82 can afford completeness and purity of 92%, despite a factor of 30 more contaminants than quasars in the color-selected feeder sample. We also show that even with much sparser time sampling, e.g., with just 6 epochs over 3 years as is the case for Pan-STARRS 1, variability is still an encouragingly efficient quasar classifier. Finally, we show that the presented A–γ technique, besides selecting quasars is also efficient at selecting (periodic) variable objects such as RR Lyrae.

This chapter is based on Schmidt et al. (2010).

2.1 – PROLOGUE –

Quasars are some of the most versatile astronomical objects known. First and foremost, they are interesting themselves. For example, samples of quasars provide precise measurements of the evolution and spectral properties of AGN in general and quasars in particular (see Section 3.2 and, e.g., Boyle et al. 2000; Vanden Berk et al. 2001; Richards et al. 2002a, 2004, 2006, 2009; Croom et al. 2009a). As described in Chapter 1, AGN and quasars probe galaxy evolution in general and especially at the epoch where the galaxy merger rate peaked and most stars formed. But not only do quasar samples provide us with insight into the nature (and nurture) of galaxy evolution and map the black hole growth in the centers of galaxies, they are also the key to various other areas of astrophysics. In particular the clustering of quasars is a tracer of mass clustering on both large and small scales (Croom et al. 2005, 2009a; Shen et al. 2007, 2009; Ross et al. 2009) and hence carries information about intergalactic structure formation. Quasar sight lines probe the content of the intergalactic medium as well as individual structures in *absorption* (D'Odorico et al. 2002; Hennawi et al. 2006b; Hennawi & Prochaska 2007). Hence, they provide information on quasar environments and the quasar emission geometry, and shed light on the physical nature of Lyman limit systems and damped Lyα systems (described in Section 3.2.1). If quasars are also gravitationally magnified by a lens, exploration of the dark matter (halo) content of galaxies (Dalal & Kochanek 2002; Bradač et al. 2002; Dobler & Keeton 2006; Macciò 2008), gaining knowledge about the molecular gas content in distant galaxies (Yun et al. 1997; Riechers et al. 2007a,b) and using the integrated Sachs-Wolfe effect by cross correlating quasars with the CMB to obtain estimates of cosmological parameters and the dark energy equation of state (e.g., Scranton et al. 2005; Giannantonio et al. 2008; Xia et al. 2009) becomes possible. In summary: large well-defined quasar samples are a cornerstone of observational cosmology.

Many applications rely on large statistical samples of quasars. Photometric quasar samples have recently grown to nearly a million objects (850,000 actual quasars; Richards et al. 2009). Despite these impressive catalog sizes, which at first glance might seem overwhelming, the number statistics still limits the achievable science in several of the science cases mentioned above; especially those where particular, and hence rare, geometric constellations of quasars are needed. For instance a 3σ detection of a luminosity-dependent quasar bias above $z \gtrsim 1.9$ when analyzing the angular clustering of quasars, needs an estimated sample size of at least 1,200,000 actual quasars (Myers et al. 2007, 2008, 2009). Searches for binary quasars (Hennawi et al. 2006a, 2010; Myers et al. 2008), which provide interesting knowledge about small scale clustering and hence shed light on quasar triggering mechanisms and the nature of quasar progenitors, also needs to be based on samples with $> 10^6$ actual quasars in order to obtain reasonably sized statistical samples of possible quasar pairs. Also quasar-galaxy clustering (e.g., Scranton et al. 2005; Lopez et al. 2008; Padmanabhan et al. 2008; Burbidge & Napier 2009), i.e., exploring the statistics of quasars behind the foreground galaxies, calls for larger (relatively low-z) quasar samples than exist today. Furthermore, exploring the 'transverse proximity effect' in the Lyα forest of quasars, with foreground quasars near the sight line of background quasars (e.g., Hennawi et al. 2006a,b; Hennawi & Prochaska 2007) is presently limited by quasar sample sizes. Obtaining larger photometric quasar catalogs to boost possible candidates for spectroscopic follow-up is needed. The size estimates of the cosmological parameters and the dark energy equation of state will also be improved by larger photometric samples of $1<z<5$ quasars. Last but not least, larger photometric quasar catalogs will enhance the number of known gravitationally lensed quasars (e.g., Oguri & Marshall 2010). At present only ∼100 quasar lenses are known (Oguri et al. 2008; Inada et al. 2010). A larger sample of the rare gravitationally lensed quasar systems will, among other things, improve our knowledge about cosmology, galaxy mass distributions, quasar hosts and the growth of the host's central black holes (see Section 2.9.2 as well as Schneider et al. (2006) and references herein).

Hence, pursuing large well-defined quasar samples is not only crucial for investigating the quasars themselves, but provides the data for a large palette of interesting astrophysical applications. The main purpose of this chapter is to investigate a new approach for obtaining such quasar samples in the future. Previous efforts have focused mainly on purely photometric selection of quasars as we will describe below. We will present an alternate approach to selecting quasars, based on their intrinsic variability. This is an important step towards understanding how quasars are selected in the most efficient and powerful way in future time-domain surveys, where color-selection may not be the preferred selection method. We will illustrate that it is actually feasible to take advantage of the intrinsic variability of quasars to select them, and show how this in practice can be applied to future time-domain surveys focusing mainly on the Pan-STARRS 1 survey (described in Section 2.3.1) which is already well underway. To sum up:

> *Huge well-defined quasar samples serve as a versatile cornerstone of observational cosmology; current samples are still not large enough for many applications.*

2.2 IGNITING GALAXIES – THE QUASAR (AGN) ENGINE

As mentioned in Section 1.5 the central engine of AGN is generally believed to be an accretion disk dumping material onto a super-massive black hole (SMBH). In this section we will describe this picture in a bit more detail, before we take advantage of one of its main characteristics, variability in the produced radiation, to select quasars from time-domain data. The fueling of the central engine is a complicated process, which is probably best illustrated by the schematic overview of the fueling of central black holes and the possible formation of SMBHs from Shlosman et al. (1990) shown in Figure 2.1. We will not attempt to describe every step in the fueling process, but only outline the crucial physical consideration one needs to take into account, and refer to Shlosman et al. (1990); Krolik (1999) and Mo et al. (2010) and their references for a more in-depth description of AGN fueling mechanisms.

Igniting a galaxy, i.e., enabling gas to accrete onto the sub-parsec sized SMBH (region) so it starts radiating, faces some theoretical challenges. First of all the SMBH needs to form and exist in the center of the galaxy before it can start radiating. In the low-redshift Universe there have been plenty of time to merge galaxies and black holes to form SMBHs in the centers of gaseous galaxies, however, as mentioned in Section 1.5.1 AGN are found out to redshift 7. This implies that whatever process forms black holes in the very early Universe, either via the collapse of super-massive isolated stars or by merging and accretion onto neutron stars, needs to be able to produce black hole 'seeds' that can grow to masses of $10^8 M_\odot$ via accretion in gaseous galaxies in less than $\sim 5\times10^8$ years, if SMBH are indeed residing at the centers of AGN. It has been argued that it is possible to form black hole seeds of masses $100-200 M_\odot$ at very high redshifts (Bromm et al. 1999; Madau & Rees 2001; Schneider et al. 2002), possibly leaving just enough time for them to grow into SMBH before $z = 7$. This is under the assumption that both a large enough dark matter halo to host the black hole seed and the gas needed for it to grow, is available. Since the SMBHs usually only correspond to $\sim 10^{-5}$ of the total halo mass (see Section 1.5), a dark matter halo of the order $10^8 M_\odot/10^{-5} = 10^{13} M_\odot$ needs to be available at $z > 7$ to ignite AGN, putting further constraints on the AGN models as well as ΛCDM (see Longair 1998; Mo et al. 2010 for further details).

Another challenge in establishing an AGN is, that when the SMBH is formed in the center of the galaxy an accretion flow onto it needs to be established. One difficulty with this is, for example, that a given black hole mass has a maximum luminosity it can radiate with under the assumption of a spherically symmetric system. This limiting luminosity, known as the Eddington luminosity, $L_{\rm Edd}$, is a balance between the gravitational force the SMBH acts with on the in-falling gas, and the radiation pressure the gas produces as

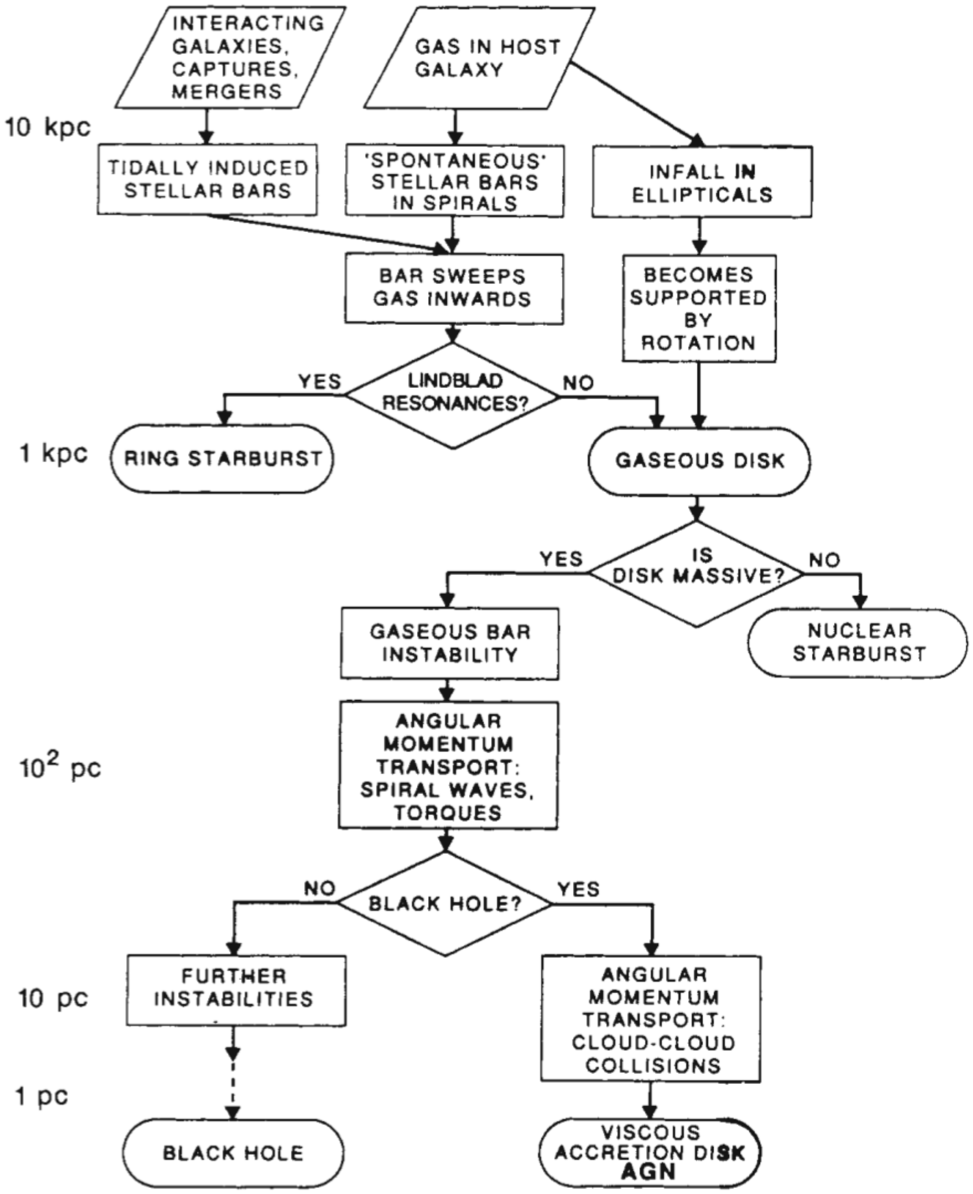
INTERACTING GALAXIES, CAPTURES, MERGERS
GAS IN HOST GALAXY
10 kpc
TIDALLY INDUCED STELLAR BARS
'SPONTANEOUS' STELLAR BARS IN SPIRALS
INFALL IN ELLIPTICALS
BAR SWEEPS GAS INWARDS
BECOMES SUPPORTED BY ROTATION
YES
LINDBLAD RESONANCES?
NO
1 kpc
RING STARBURST
GASEOUS DISK
YES
IS DISK MASSIVE?
NO
GASEOUS BAR INSTABILITY
NUCLEAR STARBURST
10^2 pc
ANGULAR MOMENTUM TRANSPORT: SPIRAL WAVES, TORQUES
NO
BLACK HOLE?
YES
10 pc
FURTHER INSTABILITIES
ANGULAR MOMENTUM TRANSPORT: CLOUD-CLOUD COLLISIONS
1 pc
BLACK HOLE
VISCOUS ACCRETION DISK AGN

FIGURE 2.1: *Schematic overview of the possibly ways of fueling the central black hole of an AGN or quasar illustrating its complicated nature. [Adapted from* Shlosman et al. (1990)*]*

it accretes onto the SMBH. If the radiation pressure is given by

$$P_{\mathrm{rad}}(r) = \frac{L}{4\pi r^2 c} , \tag{2.1}$$

assuming that the gas is ionized and the scattering of photons on the electrons has the Thomson cross section, σ_{T}, the force the radiation acts with is described by

$$F_{\mathrm{rad}} = \sigma_{\mathrm{T}} P_{\mathrm{rad}}(r) n_e(r) , \tag{2.2}$$

with n_e being the electron density. If this force is smaller than the gravitational force on the gas, the infall will be successful. On the other hand if

$$F_{\rm rad} \geq F_{\rm grav} = \frac{GM_{\rm SMBH}\,\rho(r)}{r^2} \, , \tag{2.3}$$

where $\rho(r)$ is the density of the gas cloud surrounding the SMBH, the radiation will prevent more matter from being accreted. This mechanism is 'self-regulating': The amount of radiation will decrease until the gravity is strong enough to accrete matter again, then the radiation pressure will increase until it turns off accretion and radiation then decreases again, and so on. This balance defines the Eddington luminosity and combining Equations (2.1)–(2.3) give that

$$L_{\rm Edd} \equiv \frac{4\pi c G\rho(r)}{\sigma_{\rm T} n_e} M_{\rm SMBH} \approx 1.26 \times 10^{46} \left(\frac{M_{\rm SMBH}}{10^8 M_\odot}\right) \frac{\rm erg}{\rm s} \, , \tag{2.4}$$

where $\rho(r)/n_e$ corresponds to the proton mass, m_p, assuming that the gas consist mostly of ionized hydrogen. This shows that in the case of spherical accretion a bright quasar with a luminosity of $\sim 10^{46} \frac{\rm erg}{\rm s}$ could be powered by a SMBH of $10^8 M_\odot$. The Eddington luminosity hence gives the maximum luminosity an object can obtain via spherical accretion. Super-Eddington luminosities have been observed in several quasars (see for instance Kollmeier et al. (2006) or the left panel of Figure 3.8 on page 80). This is possible because the central accretion in AGN is (obviously) not spherical.

The Eddington luminosity also puts a constraint on the accretion rate needed. Formally luminosity is accretion rate in units of c^2, i.e., $E/t = mc^2/t$. Assuming that the fixed efficiency of converting the mass accreted onto the SMBH into radiation, η, is of the order 0.1 (Krolik 1999; Mo et al. 2010), the highest possible accretion rate in the spherically symmetric model, know as the Eddington accretion rate, becomes:

$$\dot{M}_{\rm Edd} = \frac{L_{\rm Edd}}{c^2\eta} \approx 2.2 \left(\frac{M_{\rm SMBH}}{10^8 M_\odot}\right) \left(\frac{\eta}{0.1}\right)^{-1} \frac{M_\odot}{\rm yr} \, . \tag{2.5}$$

Thus the size of the SMBH regulates the maximum allowed accretion. According to this, an accretion rate of $\sim 2M_\odot/{\rm yr}$ is needed to power a quasar with $L \sim 10^{46}{\rm erg/s}$. Even though the spherical Eddington accretion picture gives a simple theoretical framework for the central engine, it is by no means a sufficient description. It is a big challenge to theoretically obtain continuous (semi)stable accretion rates of this order to power the AGN, and hence, an important step of any AGN model is to collect enough gas to sustain the accretion for the expected lifetime of the AGN. Finding the gas is usually not a big problem, as most galaxies contain several factors more gas in the interstellar medium than needed for the accretion. The real challenge is to transport the gas from the interstellar medium of the galaxy all the way into the SMBH sub-parsec scales. This gas-funneling needs to transport a significant amount of the gas' angular momentum elsewhere, for it to reach these extremely small scales, and to make sure the gas has the right temperature to prevent kinetic gas-pressure from limiting the gas-flow and lowering the accretion rate.

One way of lowering the angular momentum of the gas, sometimes by as much as a factor of 10^4 (Mo et al. 2010), is by interacting gravitationally with other galaxies. As mentioned briefly in Section 1.4 merging of galaxies (both minor and major) are believed to be able to enhance the fueling of AGN, or even trigger AGN. In simulations (e.g., Hopkins et al. 2005a, 2006; Matteo et al. 2005; Cox et al. 2006) it has been investigated whether gravitational interaction of galaxies is capable of bringing the acquired amount of gas into the SMBH region and ignite quasars. Most studies indicate that this is indeed the case, but due to the extreme difference in length scale (from kpc to sub-parsec) it is still debatable whether merging can bring the gas so close to the SMBH that the actual accretion happens. Also observational evidence for merger activity in AGN has been established (Hutchings & Neff 1992; Hutchings et al. 1988; Gehren et al. 1984; Heckman et al. 1984). Whether it is a *necessity* for AGN to have been gravitationally interacting to ignite has been a matter of great debate (Sanders et al. 1988; Canalizo & Stockton 2001; Dunlop et al. 2003; Georgakakis et al. 2009; Cisternas et al. 2011; Schawinski et al. 2011; Kocevski et al. 2012). However, in the most recent studies strong arguments are presented, that mergers are no necessity for galaxies to ignite, even though mergers do seem to be able to trigger AGN in some cases. A second possibility for gas to loose its angular momentum is by spontaneous instability caused by interaction of the accretion disk with other parts of the galaxy, minor interactions or super nova explosions (e.g., Jogee 2006). Hence, it seems to be possible for galaxies to ignite and enter their quasar (AGN) phase both with and without external gravitational interaction, as also shown in the top part of Shlosman's flow-chart in Figure 2.1. For further discussion of the challenges AGN models face see for example Krolik (1999); Kormendy & Kennicutt (2004); Jogee (2006) and Mo et al. (2010).

A natural consequence of the simple model for the central engine described here, is that the accretion rate is not necessarily constant or stable. The temperature of the in-falling gas might change, resulting in an altered spectral shape of the radiation, the Eddington limit might be reached, the amount of gas funneled onto the SMBH by a recent merger might run out, or blobs in the accretion flow could be swallowed irregularly by the SMBH, etc. Hence, it must be expected that the luminosity of AGN varies. As noted in Table 1.1 basically all AGN have indeed been observed to vary. In particular the variability of quasars has been studied in great detail. Quasars are observed to exhibit brightness variations, of typically $\gtrsim 10\%$ over several years (e.g., Giveon et al. 1999; Vanden Berk et al. 2004; Rengstorf et al. 2004; Sesar et al. 2007; MacLeod et al. 2008; Bramich et al. 2008; Wilhite et al. 2008; Kozłowski et al. 2010; Bauer et al. 2009b; Kelly et al. 2009). Even though accretion disk instabilities (Rees 1984; Kawaguchi et al. 1998; Pereyra et al. 2006) has so far been the most prevailing model for the quasar (AGN) variability, several other physical processes are and have been discussed as other important causes of the observed variability. For instance, it has been suggested that large-scale changes in the amount of in-falling material may be important (e.g., Hopkins et al. 2006 and references therein), as well as starbursts in the host galaxy (Aretxaga et al. 1997; Fernandes et al. 1997), micro lensing by the host galaxy and compact dark matter object (Hawkins 1996; Zackrisson et al. 2003), and stochasticity of multiple supernovae (Terlevich et al. 1992). Irrespective of the physics behind the variability, it is an observational fact that quasars vary significantly over periods of years. This variability has been exploited for several purposes, e.g., to estimate Eddington ratios ($L_{\rm bol}/L_{\rm Edd}$) and black hole masses (see, e.g., Bauer et al. 2009b; Wilhite et al. 2008 and Section 3.2). In Chapter 3 we will use the *color* variability to gain further insight into the underlying processes of the quasar engine.

Apart from determining Eddington ratios and black hole masses and constraining the physics of AGN, another potentially very powerful application of the observed quasar variability is to identify quasars (e.g., Scholz et al. 1997; Eyer 2002; Geha et al. 2003; Sumi et al. 2005). With for example the Sloan digital sky survey (SDSS; Abazajian et al. 2009), the quasar equatorial survey team (QUEST; Rengstorf et al. 2004) and the optical gravitational lensing experiment (OGLE; Udalski et al. 1997), large-scale, multi-epoch and multi-band surveys have emerged, and have been used to search for quasars. The Panoramic Survey Telescope & Rapid Response System 1 (Pan-STARRS 1, Kaiser et al. 2002) and 4 (Pan-STARRS 4, Morgan et al. 2008), and the Large Synoptic Survey Telescope (LSST, Ivezic et al. 2008; LSST Collaboration 2009) will take such surveys to the next level. Even though all these surveys have multi-epoch data, the largest quasar samples stem from color and not variability selection (e.g., Richards et al. 2002b, 2004, 2009; Atlee & Gould 2007; D'abrusco et al. 2009). The characteristic so-called 'UV excess' of quasars, their bright blue SDSS $u-g$ color for instance, is capable of separating the quasars from their stellar contaminants in color-color space, allowing for efficient selection of targets for spectroscopic follow-up (e.g., Strauss et al. 2002). Such UV excess color selection is, however, only efficient for low ($z \lesssim 2.5$) and high ($z \gtrsim 3$) redshift quasars, since the quasar and stellar loci overlap in the $u-g$ color for $2.5 < z < 3.0$ objects, causing the selection efficiency (or purity) in that region to drop below 50%. For quasars with $2.6 < z < 2.8$ the UV excess color selection efficiency is close to 10% (Richards et al. 2006). This confusion reigns until the Ly-break of high-z quasars moves into the g-band and again makes for unusual colors (e.g., Fan et al. 2001a; Fan 2006). More details on this color selection will be given in Section 2.3.1.

Moreover, u-band imaging is expensive: the area and depth of an optical imaging survey can be greatly increased by focusing on redder filters, where atmospheric attenuation is lower and detectors more efficient. For example, the Pan-STARRS 1 telescope offers the possibility of creating the largest sample of quasars to date with its multi-epoch 30,000 deg^2 (3/4 of the sky) *grizY* imaging survey named '3π'. For the purpose of identifying quasars in this data set, the question remains, whether we can compensate for the lack of u-band data by exploiting the multi-epoch nature of the imaging instead. With one eye on the potential of Pan-STARRS 1, we therefore explore the possibilities of creating large, complete and pure samples of quasars based on limited color information, but with light curves spanning several years. We use Stripe 82 of SDSS (Abazajian et al. 2009) as a testbed, both for the method in general and for making mock Pan-STARRS 1 data sets.

The rest of this chapter is organized as follows. In Section 2.3 we briefly review the standard color selection and previous attempts to characterize quasar variability in optical imaging surveys, before we introduce the SDSS Stripe 82 data sets in Section 2.4. We introduce our methodology for quantifying the variability of various objects via their individual power-law structure functions in Section 2.5, and show results of selection experiments in Stripe 82 in Section 2.6. After a brief discussion in Section 2.7, we conclude in Section 2.8 and present an outlook in Section 2.9.

2.3 VARIABILITY CHARACTERIZATION OF SOURCES AND THE STRUCTURE FUNCTION

In this section we briefly review the strengths and weaknesses of optical color selection, of particular sources, focusing on quasars. We present our approach for quantifying source variability, which we will then explore as an additional approach to selecting quasars.

2.3.1 COLOR SELECTION

The most common way to generate large samples of optical quasar candidates for follow-up is by specifying a particular region of interest in color space, as it was done in for example SDSS (Richards et al. 2002b, 2006, 2009). For quasars at $z < 3$ the $u - g$ color is crucial in this approach since it enables a photometric separation of the quasar candidates from the stellar locus, reducing the number of contaminating objects to a point where spectroscopic follow-up is feasible. This is illustrated in Figure 2.2, where we have plotted the median color of ~9000 spectroscopically confirmed quasars, as well as an illustrative comparison sample of 5000 F/G and 483 RR Lyrae stars (see Section 2.4), all drawn from the SDSS Stripe 82 photometric catalog DR7 (Abazajian et al. 2009). The top panels and the bottom left panel of Figure 2.2 shows the distribution of the samples in the color-color planes of the SDSS *ugriz* color cube. This clearly shows the power of the $u - g$ color (upper left panel) compared to the $g - r$, $r - i$ and $i - z$ colors in separating the quasars from their contaminants, especially for low-redshift quasars (i.e., $z \lesssim 2.5$ shown as light blue points). The color magnitude diagram in the lower right panel illustrates that a cut in magnitude will also eliminate contaminants. The contours indicate the stellar locus of Stripe 82 point sources with $15 < r < 18$.

For higher redshift objects ($z \gtrsim 2.5$, shown as magenta points in Figure 2.2) the quasars intersect the stellar locus (top left panel). The purity for $2.5 < z < 3.0$ quasar candidate samples is around 10-50% in the color-selected SDSS quasar target sample (Richards et al. 2006). In general the color selection method is efficient for low and high redshift quasars, but for the intermediate redshift objects contamination becomes a severe problem. We refer to Figures 13 and 14 in Richards et al. (2002b) for a more complete version of Figure 2.2.

With Pan-STARRS 1 the contamination problem is even more pronounced when using the color selection method only. Pan-STARRS 1 has a 5-filter system consisting of SDSS-like g, r, i, z bands (albeit with significantly higher red sensitivity) and a Y filter. The crucial $u - g$ color used in the SDSS color selection method is not available: the contamination of a color-selected Pan-STARRS 1 quasar sample will be a problem for $z > 2.5$ as well as for $z < 2.5$. It is therefore necessary to find a way of separating the majority of quasars from the contaminating stellar locus in order to obtain a pure Pan-STARRS 1 quasar sample. The intrinsic variability of the quasars described in Section 2.2 (and their contaminants) is a very promising tool for doing this.

2.3.2 SOURCE VARIABILITY: POWER-LAW STRUCTURE FUNCTIONS

The structure function characterizes the variability of quasars (and the other sources) by quantifying the variability amplitude as a function of the time lag between compared observations (Cristiani et al. 1996; Giveon et al. 1999; Eyer 2002; Vanden Berk et al. 2004; de Vries et al. 2005; Rengstorf et al. 2006). For any object the observables for estimating the structure function are the $\frac{N(N-1)}{2}$ data pairs, assuming N light curve data points, describing the variability as the magnitude difference between two epochs i and j, corrected for measurement errors, i.e.,

$$V_{i,j}(\Delta t_{i,j}) = \Delta m_{i,j} - \sqrt{\sigma_i^2 + \sigma_j^2}\,. \tag{2.6}$$

Here $\Delta m_{i,j}$ is the measured magnitude difference between observation i and j. The σ_i and σ_j are the photometric errors on the measurements and $\Delta t_{i,j}$ is the time difference between the two observations. The quantity V is defined like this so that its average, over a large number of data pairs, is an *estimator for the intrinsic standard deviation of the source magnitude.*

At this point we note that $\Delta t_{i,j}$ usually refers to the time lag in the quasar rest frame. However, computing this requires *a priori* knowledge of the quasar redshift, and when

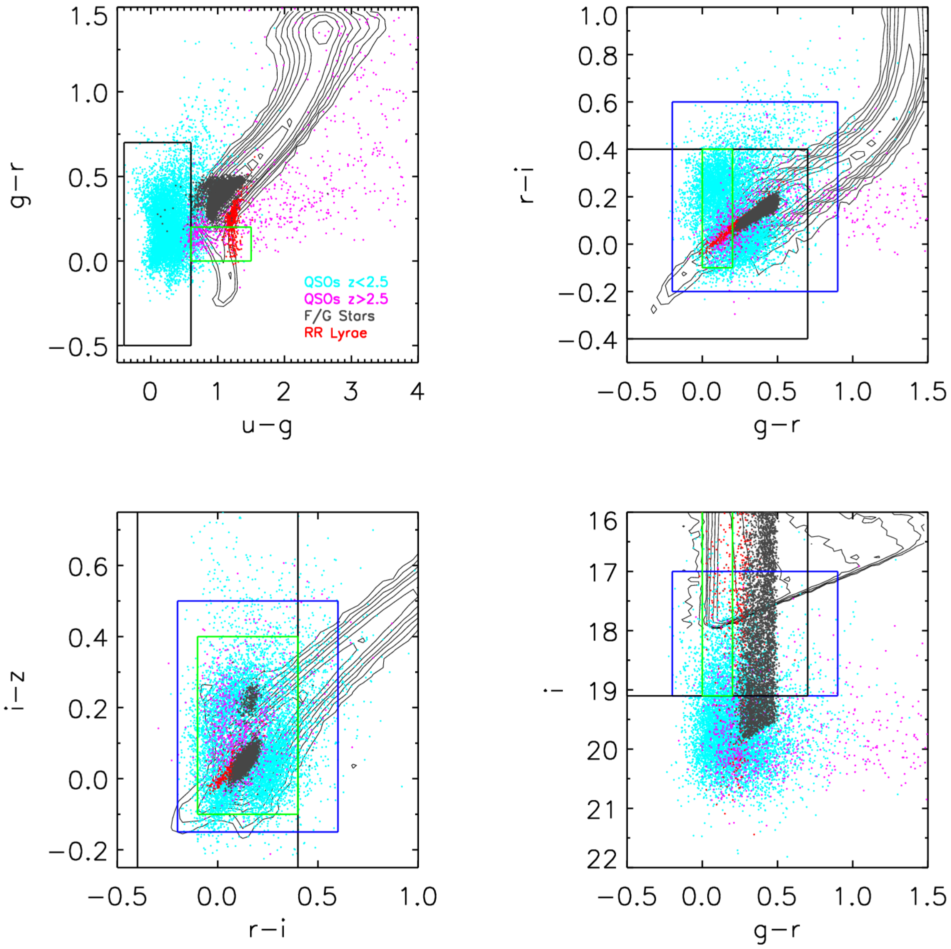

FIGURE 2.2: *Projections of point-source colors from the SDSS Stripe 82 data described in Section 2.4 to the ugriz color space. The light blue (magenta) points show* $z < 2.5$ ($z > 2.5$) *spectroscopically confirmed quasars. Illustrative contaminant point sources are shown as grey (F/G stars) and red (RR Lyrae) points. These panels demonstrate the importance of the u-band data in color selection of quasars, with the* $u - g$ *color allowing the clearest discrimination. This clearly shows the necessity for an alternative way to lower the amount of contamination when the filters are too red. The lower right color magnitude diagram illustrates that a cut in magnitude will also eliminate some contaminants, and is included for comparison purposes with Figures 4, 7, 13 and 14 in Richards et al. (2002b). The black, green and blue boxes correspond to the UVX, nUVX and griz color boxes defined in Tables 2.1, 2.2 and 2.3 respectively. The narrow appearance of the contaminant sample in the upper right panel is due to our contaminants being mostly reference stars with small photometric errors. The stellar locus spanned by all point sources in Stripe 82 with* $15 < r < 18$ *is shown as the black contours (see also figures in Richards et al. 2002b).*

selecting objects in imaging-only surveys, we do not know the object redshift. Therefore, we work with time lags in the *observed frame*. This necessary convention differs from most of the quasar variability literature; we will make some comparisons in Section 2.5.

In previous analyses, the average V has been calculated in n time lag bins using data pairs from many quasars, thus ‘stacking’ the variability signal to allow the properties of the quasar population to be probed. Then

$$V(\Delta t) = \left\langle \sqrt{\frac{\pi}{2}}|\Delta m_{i,j}| - \sqrt{\sigma_i^2 + \sigma_j^2} \right\rangle_{\Delta t} , \tag{2.7}$$

where the average, $\langle\ \rangle_{\Delta t}$, is taken over all epoch pairs i, j, whose lag falls in the bin Δt. The same approach can be taken in estimating the structure function of classes of objects (e.g., quasars in a given luminosity and redshift bin) if, say, only two epochs are available per object, but large samples exist (e.g., Richards et al. 2006, 2009); in that case the $\langle\ \rangle$ in Equation (2.7) becomes an ensemble average. Vanden Berk et al. (2004) and others find that the ensemble average quasar structure function appears to follow an increasing power law with time lag.

On the other hand, for the case where the light curve sampling of each object is high, we can compute the average $V(\Delta t)$ for an *individual object* (Eyer 2002). Binning the $\frac{N(N-1)}{2}$ data pairs from an object's N-point light curve gives an estimate of $V(\Delta t)$ defined at each bin center. This approach is computationally efficient, and provides a free-form view of the object's structure function. However, in the case of relatively sparse sampled data (6 epochs over 3 years in the Pan-STARRS 1 3π survey, see Section 2.4.7), binning the data pairs to obtain the structure function from Equation (2.7) to estimate the variability may not be the optimal approach.

In Equation (2.7), both the noise and the intrinsic photometric variability are assumed (implicitly) to have a Gaussian distribution (Rengstorf et al. 2006). We can then extend this simple model of quasar variability to include a power law increase in variability with time lag. Drawing on the results from Richards et al. (2006) we propose a power-law model for the structure function given by

$$V_{\mathrm{mod}}(\Delta t_{i,j}|A, \gamma) = A \left(\frac{\Delta t_{i,j}}{1\mathrm{yr}} \right)^{\gamma} . \tag{2.8}$$

We can then fit this model to a given set of data, $(\Delta m_{i,j}, \Delta t_{i,j})$, as follows. We write the likelihood for the power law parameters A and γ as

$$\mathcal{L}(A, \gamma) = \prod_{i,j} L_{i,j} , \tag{2.9}$$

assuming a set of independent magnitude differences as our data. Here $L_{i,j}$ is the likelihood of observing one particular magnitude difference $\Delta m_{i,j}$ between two light curve points separated by $\Delta t_{i,j}$. Following the ensemble analyses referred to above, we assume an underlying Gaussian distribution of Δm values and Gaussian photometric errors:

$$L_{ij} = \frac{1}{\sqrt{2\pi \mathrm{V}^2_{\mathrm{eff},ij}}} \exp\left(-\frac{\Delta m^2_{ij}}{2\mathrm{V}^2_{\mathrm{eff},ij}} \right) . \tag{2.10}$$

Here, the effective (observed) variability $\mathrm{V}_{\mathrm{eff}}$ is

$$\mathrm{V}^2_{\mathrm{eff},ij} = V_{\mathrm{mod}}(\Delta t_{ij}|A,\gamma)^2 + \left(\sigma^2_i + \sigma^2_j\right), \tag{2.11}$$

i.e., we propagate the photometric errors σ_i and σ_j by adding them in quadrature to the variability 'error' $V_{\mathrm{mod}}(\Delta t_{ij}|A,\gamma)$.

This approach can yield posterior probability distributions on the two model parameters, A and γ. The amplitude A quantifies the root-mean-square magnitude difference on a 1 year timescale, while γ is the logarithmic gradient of this mean change in magnitude. We assign uninformative priors for the parameters (uniform in the logarithm of A, and uniform in the arctangent of γ – since γ represents the slope of a straight line), and then explore the posterior probability distribution for these two power law parameters via a simple Markov chain Monte Carlo (MCMC) code (Metropolis et al. 1953; Hastings 1970) as described in Appendix A on page 141. All we are doing is replacing n binned structure function parameters (the values of V in each of the n bins) with two parameters that define a power law structure function, and then inferring these parameters instead of constructing estimators for them. We will show in Section 2.5 that using a power law model for the variability actually provides a good fit.

2.4 SDSS Stripe 82: a Testbed for Variability Studies

Anticipating the results of Section 2.5, we note that to detect and quantify intrinsic quasar variability will likely require multi-epoch data spanning several years. Before surveys with facilities such as Pan-STARRS and LSST become fully available, SDSS Stripe 82 (Abazajian et al. 2009) forms an excellent training set and methodological test bed (e.g., Sullivan & Collaboration 2005; Sesar et al. 2007; Bramich et al. 2008; Frieman et al. 2008). In this section we will describe the various Stripe 82 data sets that we have created in order to test and illustrate the prospects of our algorithm.

The SDSS Stripe 82 region covers approximately 300 deg^2, from right ascension around 300° to 60° in a 2.5° wide band on the celestial equator. Over 8 observing seasons it has been repeatedly observed in the fall months, resulting in many epochs (typically ~60) in each of the 5 SDSS bands. As the Pan-STARRS 1 3π survey will contain fewer epochs, we can ‘down-sample’ the Stripe 82 object light curves to simulate observations taken with Pan-STARRS 1 (albeit ones at lower angular resolution and depth).

Relative to Pan-STARRS 1, Stripe 82 does have the advantage of $u-g$ color coverage, and extensive bright object spectroscopy. One can therefore construct quite pure samples of quasars, RR Lyrae and so on, that may serve as ground truth for our variability selection.

In the following subsections we describe the various subsamples used in our study in some detail, and provide a brief overview here. We have selected all the spectroscopically confirmed quasars in Stripe 82 together with a representative set of (stellar locus) contaminants, which contains non-varying (type F/G stars) as well as varying (RR Lyrae) point sources to illustrate our method and algorithm prospects. These objects’ photometry data are plotted in *ugriz* color space in Figure 2.2. To investigate the selection of quasars by their colors, we define three color selection boxes and explore the objects returned by each. One of these mimics the more limited color selection possible with Pan-STARRS 1. Quantifying how the variability information improves the Pan-STARRS 1 quasar selection is one of the main goals of this chapter.

In the following subsections, we describe two preparatory steps for turning the ~ 60 epoch Stripe 82 data into a testbed for color plus variability based quasar selection in SDSS (Stripe 82) and Pan-STARRS 1: first, we describe the definition of various sub-sets of candidate objects; then we describe some technical steps ‘cleaning’ the light curves and down-sampling the Stripe 82 data to mimic Pan-STARRS 1 observations.

2.4.1 SPECTROSCOPICALLY CONFIRMED QUASARS IN STRIPE 82

A key to designing a quasar variability selection algorithm is an understanding of the variability properties of objects that are spectroscopically confirmed to be quasars. We have selected all of these (both point sources and extended objects) published in the SDSS DR5 quasar catalog (Schneider et al. 2007) that fall within Stripe 82. There are 9157 spectroscopically confirmed DR5 quasars in Stripe 82, spanning a redshift range from 0.08 to 5.09. These quasars have $15.4 < i < 22.0$ with a mean of 19.5. See Schneider et al. (2007) for the corresponding numbers for the whole DR5 quasar catalog.

To get the multi-epoch photometry (light curves) for the 9157 quasars we performed an SQL neighbor search in the Stripe 82 DR7 database, choosing a search radius of $0.5''$ to minimize the light curve contamination from misidentified (spatial) neighbors. This search on average yielded 60 epochs per object, after selecting only entries with good `BRIGHT`, `EDGE`, `BLENDED`, `NODEBLEND`, `SATUR`, `PEAKCENTER`, `NOTCHECKED`, `INTERP_CENTER` and `DEBLEND_NOPEAK` flags (of which the first 5 are referred to as fatal and the rest as non-fatal flags by Richards et al. (2002b) – see their Table 2 or Stoughton et al. (2002)

Table 9 for a description of the flags). For consistency we applied these same flag checks on *all* object samples we drew from the Stripe 82 catalog. We describe these other samples below.

2.4.2 Stellar locus 'contaminants' in Stripe 82

To get a sample of typical non-variable stellar contaminants we used the SDSS standard star catalog of 1.01 million non-variable point-source objects in Stripe 82 published in Ivezić et al. (2007). From that we created a set of F/G star colored objects, presumably non-varying, by applying a color-magnitude cut on the standard star catalog so that $0.2 < g - r < 0.48$ and $14.0 < g < 20.2$ for all the objects. This is a suitable cut for F/G stars according to the SEGUE team (Yanny et al. 2009) and makes them potential quasar contaminants because of their $g - r$ color (see Figure 2.2). We took a randomly selected subsample of 5000 objects from this catalog and did a neighbor search in Stripe 82 to get multi-epoch observations of these contaminants. We again used a search radius of $0.5''$ and again made sure that none of the flags listed in Section 2.4.1 were set.

To be able to test whether our algorithm is able to separate quasars from known variable contaminants, we used the largest available sample of securely identified RR Lyrae within Stripe 82 (Sesar et al. 2010), which consists of 483 RR Lyrae.

We will refer to the F/G stars and RR Lyrae catalogs collectively as the 'contaminants' in the remainder of the chapter.

2.4.3 UV excess (UVX) objects

We would also like to test our ability to detect quasars in the absence of spectroscopic data. To this end, we defined three photometrically-selected samples of Stripe 82 objects, whose variability properties we will explore.

The first of these is defined by a three-dimensional *ugri* color box in which the SDSS quasar sample is complete for extinction-corrected *i* magnitudes brighter than 19.1 (Richards et al. 2002b). This color box is given in Table 2.1, and is shown with black lines in Figure 2.2. Note that this selection uses the SDSS *u*-band data. We extracted all point sources within Stripe 82 that obeyed these 'UV excess' (UVX) criteria. This returned a catalog of 2912 UVX point sources.

TABLE 2.1: *The UV excess (UVX) color box.*

-0.4 <	$u-g$	< 0.6
-0.5 <	$g-r$	< 0.7
-0.4 <	$r-i$	< 0.4
	i	< 19.1

2.4.4 NON-UV EXCESS (nUVX) OBJECTS

As a compliment to the UVX object sample defined above, where the color selection is known to efficiently return quasars at high completeness, a catalog of 'non-UVX' (nUVX) objects was created from a region of *ugriz* color space where color selection of quasars is known to have problems. The color box from which we selected these nUVX point sources is given in Table 2.2, and shown as a green box in Figure 2.2 (and also Figure 7 of Richards et al. 2002b). In this particular color box, the quasar locus, containing mostly intermediate redshift ($2.5 < z < 3$) quasars, crosses the stellar locus. The color selection therefore has efficiency as low as 10% in this region of color space (Richards et al. 2006). In the nUVX color box we find 3258 objects in Stripe 82.

TABLE 2.2: *The non-UV excess (nUVX) color box.*

0.6 <	$u-g$	< 1.5
0.0 <	$g-r$	< 0.2
-0.1 <	$r-i$	< 0.4
-0.1 <	$i-z$	< 0.4
	i	< 19.1

2.4.5 QUASAR CANDIDATE COLOR SELECTION WITHOUT *u*-BAND DATA

To simulate approximately the anticipated Pan-STARRS 1 3π survey light curves, we defined a third color box suitable for a first cut of the Pan-STARRS 1 catalog. The main purpose of this selection (where no *u*-band information is used) is to excise the quasar locus as it threads through the three-dimensional *griz* color space. However, part of the stellar locus also lies in this box. We restrict ourselves to right ascensions between 0 and 20 degrees (enclosing a sixth of the Stripe 82 area, ~50 deg^2) in order to return a manageable catalog of 12,714 objects. The *griz* box is indicated by the blue solid lines in Figure 2.2 and is defined in Table 2.3. The magnitude cut of 19.1 is chosen to allow straightforward comparison with the UVX and nUVX boxes.

TABLE 2.3: *The griz color box.*

-0.2 <	$g-r$	< 0.9
-0.2 <	$r-i$	< 0.6
-0.15 <	$i-z$	< 0.5
17 <	i	< 19.1

2.4.6 ELIMINATING LIGHT CURVE ‘OUTLIERS’ IN STRIPE 82

Plotting the complete Stripe 82 multi-epoch photometry output for the various objects revealed some outlying points that were several magnitudes fainter than adjacent flux points (see Figure 2.3); only some of these outliers were found to be caused by image defects. However, we assume that such a significant decrease in magnitude in a single observation must be non-physical. We therefore removed the outliers (irrespective of their origin) by running a median filter on the photometric measurements. Measurements with a residual between the medianized light curve and the photometric data larger than 0.25 magnitudes were removed. In Figure 2.3 we show the g, r and i-band multi-epoch photometric measurements (open symbols indicating the removed measurements) with the corresponding medianized light curves over-plotted for quasar SDSS J203817.37+003029.8. The bottom panel shows the residuals, with the limit of 0.25 magnitudes indicated by the dashed lines. As is the case here, in general, only a small fraction of the epochs is removed.

It is these cleaned multi-epoch measurements, where the outlying observations have been removed (i.e., the filled symbols in Figure 2.3), we use in the determination and exploration of the objects’ variability.

2.4.7 DOWN-SAMPLING STRIPE 82 LIGHT CURVES TO THE PAN-STARRS 1 CADENCE

In order to explore quasar selection in the context of the Pan-STARRS 1 3π survey, we down-sampled the Stripe 82 data to mimic the Pan-STARRS 1 observations (Chambers & Denneau 2008 shown schematically in Table 2.4). We assumed 3 observing seasons for Pan-STARRS 1, with a duration of 155 days (covering all filters) each. Only Stripe 82 objects with more than 7 epochs in each (SDSS) season were passed to the actual down-sampling routine: ~1% of the quasars, < 0.1% of the F/G stars and ~20% of the RR Lyrae did not satisfy this criteria.

We down-sampled the Stripe 82 light curve data by matching each season of observations

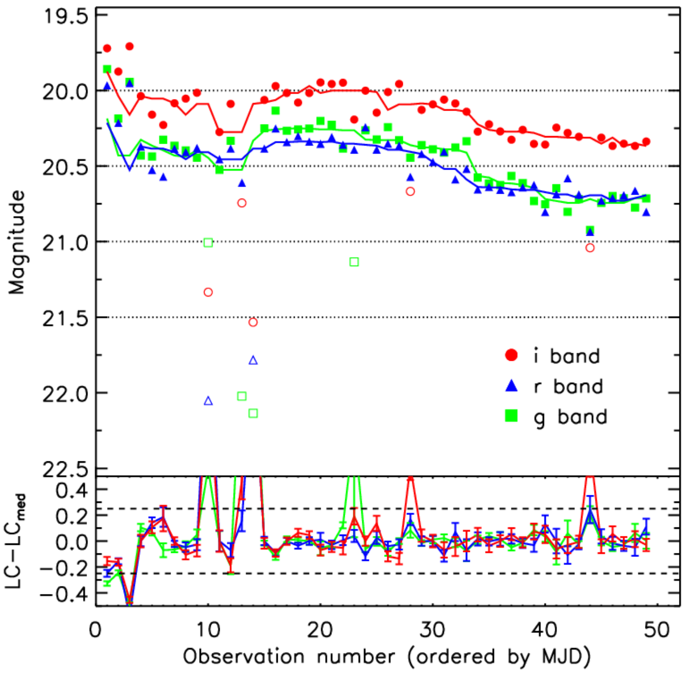

Figure 2.3: *Multi-epoch photometry output from SDSS Stripe 82 for the spectroscopically confirmed quasar SDSS J203817.37+003029.8, shown in the g (green squares), r (blue triangles) and i (red circles) bands. For a handful of epochs, the output magnitudes appear spuriously faint; their inclusion would severely affect the calculation of a variability structure function. Overplotted are the corresponding medianized light curves used to remove the outliers (open symbols) from the raw multi-epoch output. The dotted lines are plotted to guide the eye and are spaced by half a magnitude from 20 to 21.5. In the lower panel the residuals between the medianized light curve and the photometric measurements for the three bands are shown, with the photometric errors over-plotted. The limit used to remove the outliers* ($|LC - LC_{med}| > 0.25$) *is indicated by the horizontal dashed lines.*

for the g, r, i and z-band with the (approximately) correct time intervals between consecutive observations in each band. No color information went into the down-sampling. After identifying 6 suitable Stripe 82 epochs in each band we removed all other observations, providing a set of mock Pan-STARRS 1 data.

Table 2.4: *A schematic overview of the Pan-STARRS 1 3π survey schedule used to create mock Pan-STARRS 1 data.*

Date (days)	-60	.	.	.	.	.	-30	.	.	.	.	.	0	5	10	.	.	.	30	35	40	.	.	.	60	.	.	.	.	.	90
Band	*z*	.	.	.	.	.	*Y*	.	.	.	.	.	*i*	*r*	*g*	.	.	.	*i*	*r*	*g*	.	.	.	*Y*	.	.	.	.	.	*z*
Moon	F	.	.	N	.	.	F	.	.	N	.	.	F	.	.	N	.	.	F	.	.	N	.	.	F	.	.	N	.	.	F

Notes: Date is calculated with respect to the first i-band measurements. Band shows the band observed. Moon indicates whether the moon is full (F) or new (N). The columns are spaced by five days.

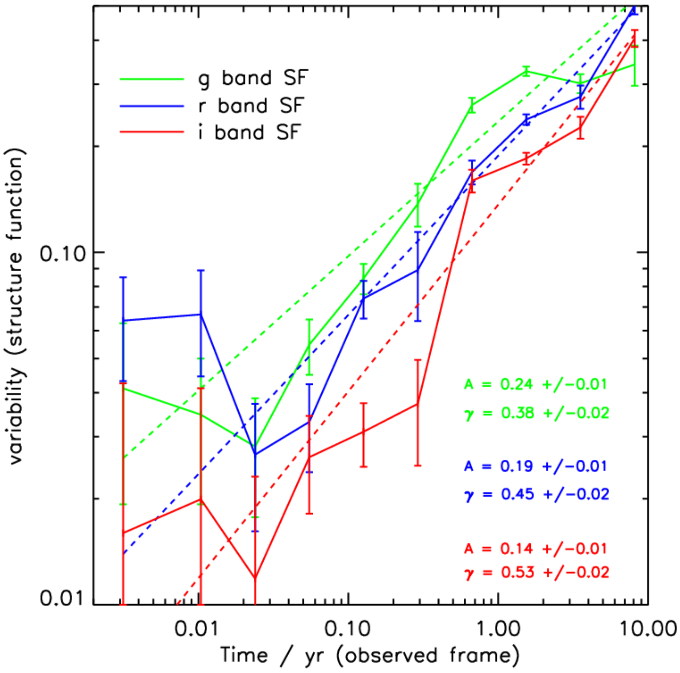

FIGURE 2.4: *Variability structure functions (solid lines) for SDSS J203817.37+003029.8, based on the photometry shown in Figure 2.3 for the g (green), r (blue) and i (red) band, computed from Equation (2.7). The best-fit power law (Equation (2.8)) model for the structure function from the MCMC simulated annealing code (Appendix A on page 141) are over-plotted as dashed lines. The corresponding A and γ parameters of the power law and their estimated errors are quoted in the lower right corner of the plot. Calculating similar structure functions by means of the power-law model for the 9157 Stripe 82 quasars give the median quasar sample structure function shown in the top panel of Figure 2.5.*

2.5 POWER-LAW STRUCTURE FUNCTIONS FOR SOURCES IN STRIPE 82

In Figure 2.4 we show the binned structure functions (Equation (2.7)) from the g, r and i-band light curves of quasar SDSS J203817.37+003029.8 (Figure 2.3). Calculating binned structure functions for individual Stripe 82 objects is a simple way of quantifying the variability of each object if there is a large number of epochs available. However, Figure 2.4 suggests that we might indeed be justified in further compressing the structure function into a two parameters power-law fit, as proposed in Section 2.3.2.

Before doing so, we explore wether this power law behavior is present in an average sense. In Figure 2.5 we show the median sample structure function, created by median-combining all the individual binned structure functions calculated with Equation (2.7) separately for the well sampled Stripe 82 quasars (Section 2.4.1), F/G stars (Section 2.4.2) and RR Lyrae (Section 2.4.2 and Sesar et al. 2010). The shaded regions in Figure 2.5 around the medians enclose 68% and 95% of the individual structure functions. Figure 2.5 shows that the quasar sample median of the binned structure function closely resembles a power law with $A = 0.093 \pm 0.0002$ and $\gamma = 0.43 \pm 0.002$, in agreement with findings elsewhere in the literature (Vanden Berk et al. 2004; Rengstorf et al. 2006; Wilhite et al. 2008; Bauer et al. 2009b). In particular, a value of the slope of the sample median structure function of $\gamma = 0.43$ agrees well with most of the literature (see, e.g., Table 4 in Bauer et al. (2009b) for a brief overview). Rough estimates of the 1-year observed frame power law amplitudes in the literature give amplitudes between 0.10 and 0.14 (depending on the assumed mean redshift of the samples), in good agreement with our estimate for A of 0.093 mag.

Figure 2.5 also shows that the sample structure functions of contaminants are also well described by power laws, but with small values of γ (i.e., they show no long-term growth in their variability). Note that the F/G stars, chosen to be non-variable, have a variability amplitude of ~0.04 mag. The RR Lyrae variability, when sparsely and randomly sampled in Stripe 82, looks like white noise ($|\gamma| \ll 1$) with an amplitude ~0.2 mag. Thus, the use of a power-law model of the form given in Equation (2.8) seems to be a fairly good assumption, and different types of objects may differ both in amplitude and in slope of their structure function.

In Figure 2.4 the power law fits to the three (binned) structure functions are shown as dashed lines. The A and γ with their estimated errors are given in the lower right corner of the plot.

2.6 RESULTS

Having defined different sub-samples of sources, and having shown that we can sensibly quantify their light curve characteristics by a power-law structure function model, we now proceed to characterize each of these sources by their best-fit parameters A and γ.

As opposed to the earlier SDSS analysis (Richards et al. 2002b, 2006, 2009) the improved time sampling of the Stripe 82 (and even Pan-STARRS 1) surveys, enables us to investigate the *distributions* of the A and γ parameters for the individual sources, not only for ensembles.

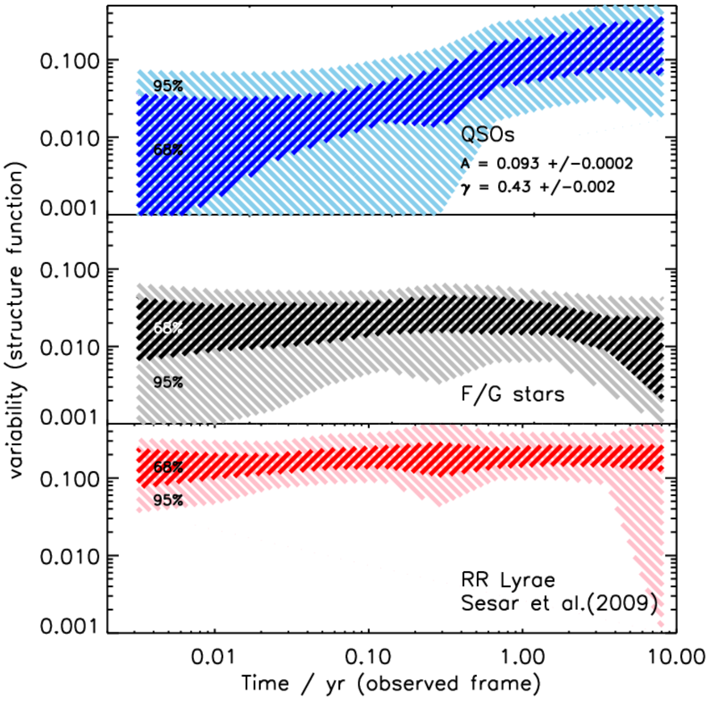

FIGURE 2.5: *The sample-median r-band binned structure function for the quasars (top), F/G stars (center) and 483 RR Lyrae (bottom) in Stripe 82, calculated using Equation (2.7). The power-law nature of the three samples is clearly seen, by the approximate straight lines the structure functions trace. The power law parameters A and γ obtained by fitting the quasar sample structure function with the 2 parameter power law model in Equation (2.8) are shown in the top panel. The shaded regions around each structure function indicate the 68% and 95% scatter around the median value.*

2.6.1 THE A–γ DISTRIBUTION

Figure 2.6 shows the distribution of the variability characteristics, quantified by the best-fit A and γ for all the spectroscopically confirmed quasars, and for the 'contaminant' F/G stars and RR Lyrae, as described in Section 2.4, based on their r-band Stripe 82 light curves. The histograms along the axes of the two-dimensional scatter plot show the projected parameter distribution of the quasars and of the contaminants. Inspection of Figure 2.6 alone shows how well the spectroscopically confirmed quasars separate from the (stellar locus) contaminants in this space, demonstrating that the power-law structure function fit from a single band is an efficient classifier for data of this quality (~60 epochs).

The analogous A–γ distributions for the much sparser Pan-STARRS 1-like sampling of the r-band measurements (Section 2.4.7) are shown in Figure 2.7: these parameter esti-

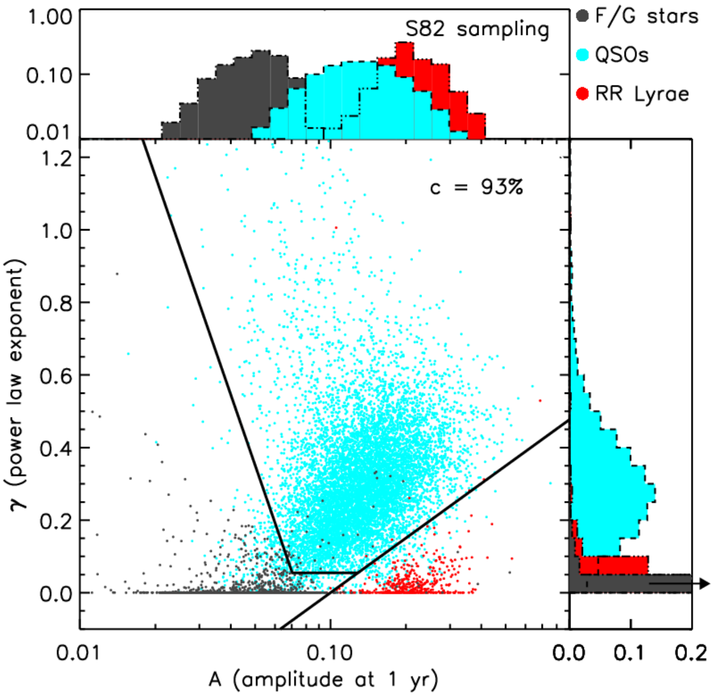

Figure 2.6: *Distribution of the variability structure function parameters A and γ (Equation (2.8)) for ~15,000 individual objects in Stripe 82. The spectroscopically confirmed quasars are shown as light blue points; confirmed RR Lyrae and color-selected F/G stars are shown in red and grey, respectively. A separation of the quasars from the stellar locus contaminants is clearly seen. The three solid lines (Equations (2.12)–(2.14)) define the region in which we estimate the quasar completeness c of our algorithm (which turns out to be 93% in this case, Table 2.5). Along the axes we show the projected A and γ distributions for the sub-samples.*

mates are based on only 6 epochs of photometry over 3 years, rather than the ~60 in the full Stripe 82 survey. The separation of the quasars and the contaminants is less clean with the Pan-STARRS 1 sampling, but one nevertheless clearly sees a quasar-dominated region with rather low contamination. Plots analogous to the ones shown in Figures 2.6 and 2.7, but for the g, i and z-band measurements, show that on average A decreases by 30-50% going from g to z band, whereas γ is unchanged with varying wavelength. Thus, the separation of the quasars from the contaminants via their γ values appears to work comparably well in all 4 bands (with somewhat more scatter in the z-band). In agreement with Kozłowski et al. (2010), no clear difference in the ratio of the amplitudes at different wavelengths between RR Lyrae and quasars is detected.

Since most F/G stars should not vary, but RR Lyrae do, they should have different A distributions. This is seen in Figure 2.6 and 2.7: RR Lyrae have magnitude amplitudes above ~0.1 (e.g., Soszynski et al. 2003; Sesar et al. 2010), while the F/G stars have characteristic values of $A \sim 0.01$. It is therefore clear that our approach can also separate

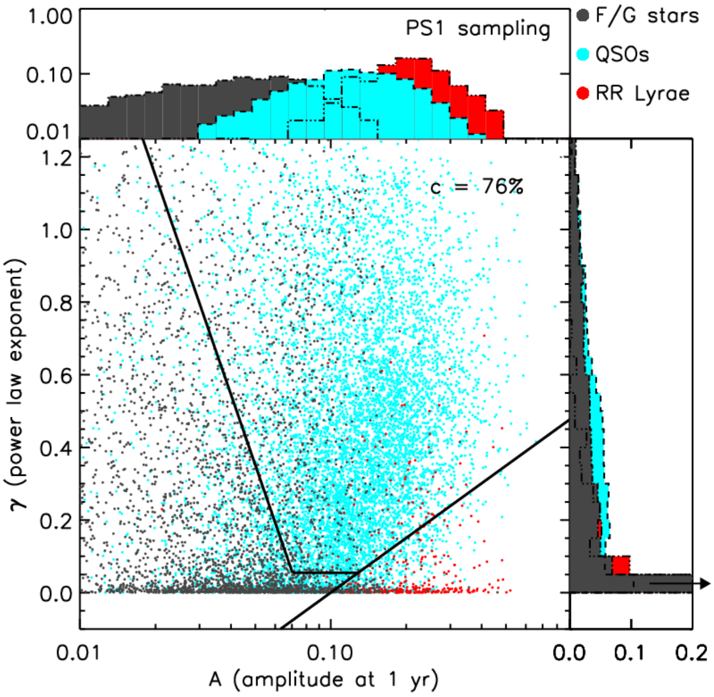

FIGURE 2.7: *Distribution of the variability structure function parameters, similar to Figure 2.6, but after down-sampling to the 6 epochs Pan-STARRS 1 3π survey cadence (Section 2.4.7). A variability separation of quasars and their contaminants is again apparent, albeit not as clearly as in Figure 2.6. The selection region is the same as in Figure 2.6; the completeness is given in the upper right corner of the plot, and in Table 2.5.*

RR Lyrae from stellar (non-varying) contaminants without doing a full fit of a periodic light curve.

2.6.2 COMPLETENESS AND PURITY

To move beyond a merely qualitative assessment of the separability of the quasars from contaminants we now estimate the achievable completeness and purity of the resulting sample. Purity is the more difficult quantity to estimate as it requires appropriate abundances for the contaminants. We do *not* have these for the training set quasar, F/G star and RR Lyrae samples, where the ratio of quasars to contaminants is 2:1, instead of a more realistic ~1:25, and we can therefore only estimate completeness when working with these samples. However, by design the UVX, nUVX and *griz* selection boxes (Sections 2.4.3–2.4.5 and Sections 2.6.2.1–2.6.2.3 below) give parent samples with the correct quasar-contaminants ratio: we use these to explore the purity of our quasar selection algorithm.

For the present we divide the A–γ plane by simple cuts that define a quasar selection box, and then quantify its performance. Specifically, our fiducial quasar selection region is bounded by the following three straight lines:

$$\gamma(A) = 0.5 * \log(A) + 0.50 \tag{2.12}$$

$$\gamma(A) = -2 * \log(A) - 2.25 \tag{2.13}$$

$$\gamma(A) = 0.055\ . \tag{2.14}$$

These cuts are shown as black solid lines in Figures 2.6–2.10.

This can be thought of as a way of providing lower limits on the available completeness and purity, that a more sophisticated selection procedure would improve upon. One could of course tweak the cuts in Equations (2.12)–(2.14) to explore the trade-off between purity and completeness, which we have only done here ‘by eye.’

Applying these cuts to the data leads to the completeness given in Table 2.5 for the ‘QSO+contam’ catalog. The completeness is calculated as the fraction of spectroscopically confirmed quasars in the sample that fall within the cuts. In our simple illustrative setup we have a completeness of 93% for the quasars in the case where the time sampling is equal to Stripe 82 (~60 epochs). In the case of a Pan-STARRS 1-like time sampling the completeness drops to 76%.

TABLE 2.5: *The completeness c and purity p of the SDSS Stripe 82 and mock Pan-STARRS 1 variability selected object catalogs. The δc and δp indicate the poisson error on c and p.*

Parent Cat.	Sampling	c	δc	p	δp	Quasar Reference Cat.
QSO+cont.	Stripe 82	93%	1%	-	-	SDSS
QSO+cont.	Pan-STARRS 1	76%	1%	-	-	SDSS
UVX	Stripe 82	90%	3%	95%	3%	SDSS+2SLAQ
UVX	Pan-STARRS 1	73%	2%	92%	3%	SDSS+2SLAQ
nUVX	Stripe 82	90%	15%	96%	16%	Visual
nUVX	Pan-STARRS 1	65%	12%	32%	5%	Visual
griz box	Stripe 82	92%	6%	92%	6%	Visual & SDSS+2SLAQ
griz box	Pan-STARRS 1	75%	6%	30%	2%	Visual & SDSS+2SLAQ

Of the RR Lyrae 97% and 83% (for the Stripe 82 and Pan-STARRS 1 sampling respectively) fall in the high-A low-γ corner below the line given by Equation (2.12). In this region <0.5% of the 5000 F/G stars lie, illustrating quite a clean separation between RR Lyrae and non-varying stellar contaminants.

Besides calculating the overall completeness, we also split our data into redshift bins. The completeness of the quasars is rather constant as a function of redshift, with a minor loss of about 5-10% for redshifts above 4. Since our training set only contains 52 quasars at $z > 4$, we were not able to investigate this decreased completeness further.

In the next three subsections we proceed to explore the A–γ distributions for the samples of objects that were selected only on the basis of their colors (Table 2.1, 2.2 and 2.3). For those samples, we estimate the completeness (as above) and also the purity, defined as the fraction of known quasars compared to the total number of objects inside the A–γ selection regions.

2.6.2.1 QUASARS IN THE UVX CATALOG

For the UVX object sample (Section 2.4.3) there is enough spectroscopy in Stripe 82 to define a spectroscopically confirmed quasar subsample, a reference catalog of quasars which is complete in Stripe 82. By combining the spectroscopically confirmed quasars in Stripe 82 (Section 2.4.1), with the objects from the 2SLAQ (2-degree field SDSS luminous red galaxies and QSO) survey (Croom et al. 2009b), our final quasar reference catalog contains 11216 individual quasars (9157 from SDSS and 2059 from 2SLAQ). We only selected objects flagged as 'QSO' in the publicly available 2SLAQ data.[1]

Matching this quasar reference catalog with the catalog of UVX Stripe 82 point sources returned 2140 quasars out of the 2912 objects in the UVX catalog. Thus, 73% percent of the point sources in the UVX color box are known quasars. It is not surprising that the fraction is so large, since we used the powerful $u - g$ color in the definition of the color box. This simply re-affirms that the UVX box is a region of color space where the quasars are in the majority, as they are well separated from the stellar locus by the $u - g$ color; it is exactly this separation we are trying to find an alternative to.

We estimated A and γ for the entire UVX catalog, using both the full (Stripe 82) sampling and the sparser Pan-STARRS 1-like version of it. The result is shown in the top row of the A–γ plots in Figure 2.8. Applying our simple variability selection cuts (Equations (2.12)–(2.14)) returned 2033 and 1734 quasar candidates for the Stripe 82 and Pan-STARRS 1 sampling, respectively. Matching these objects with the 2140 know SDSS+2SLAQ quasars in the UVX catalog returned 1935 and 1573 matches. Thus we are able to detect the UVX quasars with a completeness of 90% and a purity of 95% (1935 matches/2033 candidates) when using the Stripe 82 time sampled data. For the Pan-STARRS 1-like sampling of the data we get a completeness of 73% and a purity of 92% (Table 2.5).

---

[1] `http://www.2slaq.info/`

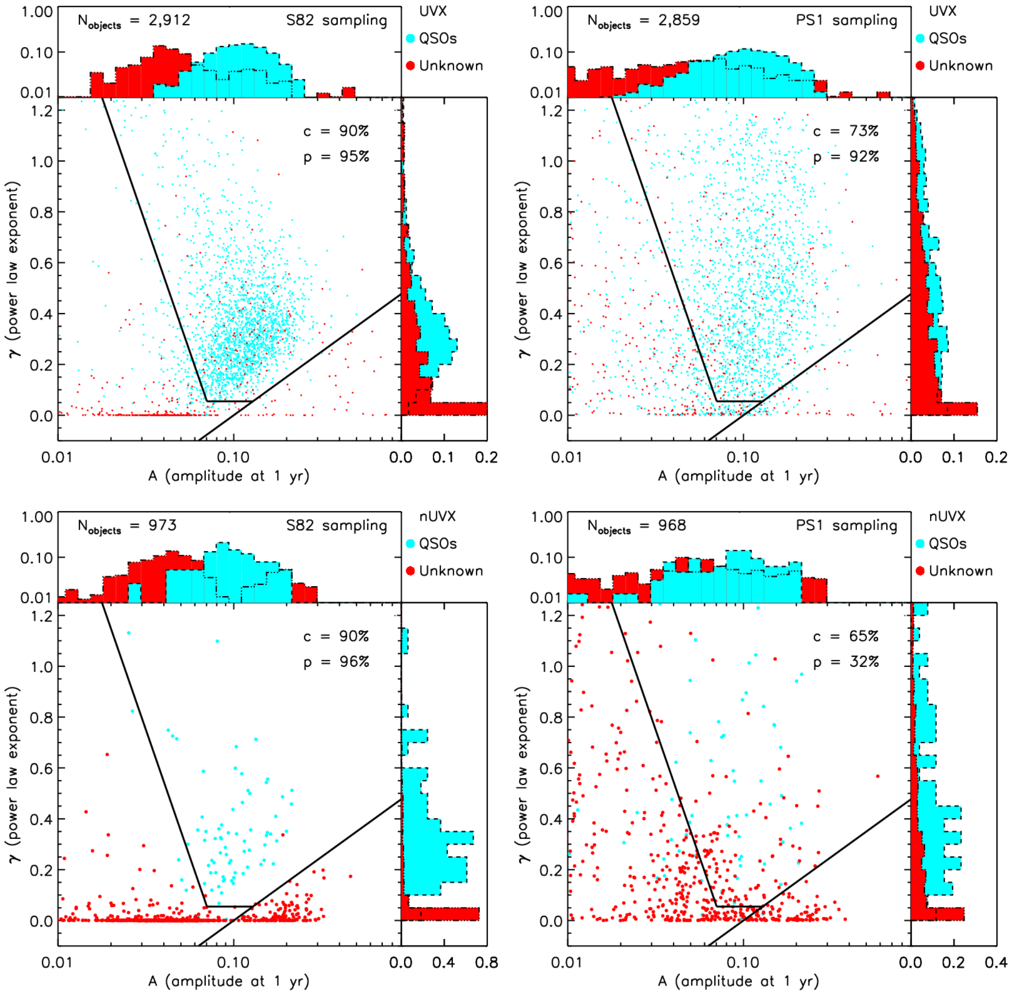

FIGURE 2.8: *Distribution of variability structure function power law parameters A and γ measured for the objects in the UVX and nUVX catalogs of Table 2.5. The left plots corresponds to the catalogs with a Stripe 82 time sampling, while the right plots correspond to the sparser Pan-STARRS 1 time sampling (6 epochs over 3 years). The three solid lines in each scatter plot correspond to the quasar variability selection region defined by Equations (2.12)–(2.14). The completeness and purity estimates are shown in the upper right corner of each scatter plot, and in Table 2.5. In the upper left corner the total number of objects in the catalog is stated. The light blue points indicate the known quasars, and the red points all other objects. Along the axes of the two-dimensional scatter plots the projected probability distributions for A and γ are shown as histograms.*

2.6.2.2 QUASARS IN THE nUVX CATALOG

The catalog of the UVX point sources was deliberately chosen from a region of color space where the color selection already does a superb job finding quasars, as confirmed by the complete SDSS and 2SLAQ quasar catalog and the completeness and purity of our A–γ approach. However, one might argue that in this case (of UVX quasars) a light curve analysis adds little. To explore the A–γ approach further we applied it to the non-UV excess objects described in Section 2.4.4.

In the nUVX color box (Table 2.2) there is no simple way to quantify the completeness of the parent sample of color-selected objects, since we do not know how many quasars were missed in this region of color space during the SDSS survey. To try to quantify this, we extracted the spectra from the 973 objects in this catalog that were targeted for SDSS spectroscopy. By visually inspecting these spectra we were able to compile a catalog of 77 quasars among the Stripe 82 nUVX objects. (Of these 77, 6 quasars are not in the SDSS+2SLAQ catalog). This means that if the SDSS fibers had been allocated to nUVX objects randomly, then the purity of the nUVX quasar sample would be $77/973 = 8\%$. However, in practice the fibers were placed according to a Bayesian ranking that made fuller use of the color information, so that this 8% is likely an upper limit on the purity of the nUVX quasar sample. (The fraction of quasars hidden in the un-targeted nUVX objects is likely lower than the fraction of quasars found in the nUVX objects with spectra.)

Applying our A–γ analysis and our variability selection criteria to the spectroscopic sub-catalog of 973 nUVX objects returned 72 (178) quasar candidates for the Stripe 82 (Pan-STARRS 1-like) time sampling. The nUVX objects' variability parameters are plotted in the bottom row of Figure 2.8. Estimating the completeness and purity in the Stripe 82 sampling case, assuming that the 973 objects with spectra are a random subset of all the nUVX objects, gives that we are 90% complete and 96% pure (Table 2.5). By the same argument as above, the purity is an upper limit on the overall purity of the nUVX quasar sample whereas the completeness is exact. The purity and the completeness stand on their own in quantifying our ability to recover the *spectroscopically confirmed quasars.* Thus, the addition of the variability information enhances the purity to 96% instead of 8%, as is the case for the purely color-selected sample. In the case of the sparser Pan-STARRS 1-like sampling we still have a completeness of 65%, and a purity of 32%. This clearly demonstrates that a variability-based approach is very efficient at selecting nUVX quasars when the data is well sampled in time. Even with the sparse Pan-STARRS 1 sampling, the purity increases by a factor of 4 when variability information is used.

The plot of the Stripe 82-sampled nUVX objects (lower left corner of Figure 2.8) shows a clear bimodality of the A parameter distribution of the unknown (red) objects. As seen with the object training set, this bimodality is a probable separation between the non-varying contaminants and the varying (possible RR Lyrae) contaminants. Thus the nUVX objects have been separated into quasar candidates (high γ; intermediate A), RR Lyrae candidates (low γ; high A) and non-varying stars (low γ; low A).

2.6.2.3 QUASARS IN THE *griz* COLOR BOX

To simulate quasar candidate selection without u-band photometry, we applied our variability analysis to the objects lying in a fairly large multi-color region in *griz* space, the so-called *griz* box defined in Section 2.4.5. This color box fully contains an important part of the stellar locus. Estimating the completeness and purity of our algorithm for the *griz* box is difficult, as no clear estimates of the abundance of quasars exist for such a color cut. We therefore checked the quasar candidates against a catalog of the SDSS+2SLAQ quasars, plus the 6 extra quasars found via the visual inspection of the nUVX spectra. This may fall considerably short of a complete sample of quasars, but at the moment it is the best we can do; the purities calculated in this section are therefore lower limits. Matching this quasar reference catalog to the 12,714 objects in the *griz* box we found 443 known quasars. Estimating A–γ for all these sources, returned 442 (1,118) variability-based quasar candidates, when considering the Stripe 82 (Pan-STARRS 1) sampling and when applying the A–γ cuts of Equations (2.12)–(2.14). Of these candidates 407 (333) were found to be known quasars. Thus, for the broad *griz* color pre-selection, variability selection achieves 92%(75%) completeness and a purity of 92%(30%) for the Stripe 82 (Pan-STARRS 1) sampled data, respectively. The A and γ distributions for the *griz* box selected objects are shown in the top panel of Figures 2.9 (Stripe 82 sampling) and 2.10 (Pan-STARRS 1 sampling).

In the bottom panel of Figure 2.9 and 2.10 we have projected our variability selected quasar candidates back into *ugr* and *gri* color space, in order to understand the color distributions of variability selected quasar candidates. The figures show that 86% and 98% of the not spectroscopically confirmed candidates fall on the *gri* stellar locus, defined as the (blue) contour level containing 95% of the stellar locus objects in Stripe 82. This suggests that many, if not most, of these unconfirmed quasar candidates are stars scattered into our variability selection region. However, there are some possible quasars among the unknowns judging from their colors. For instance, a few of the unknown objects (6% and 1% in the Stripe 82 and Pan-STARRS 1 sampled case respectively) fall in the *ugr* UVX selection box. This illustrates that our purity estimates are lower limits but close to the likely truth. It also shows that the *ugriz* quasar color selection in SDSS (Richards et al. 2002b, 2006, 2009) has done an excellent job, implying that $< 10\%$ of the quasars with $i < 19.1$ are 'hiding' in the stellar locus and have been missed by the SDSS selection.

When defining the *griz* box in Section 2.4.5 we included the stellar locus to achieve as high a completeness as possible and to illustrate the prospects of our approach. However, removing objects falling within the stellar locus, could greatly enhance the purity at a modest reduction of the completeness. Thus, selecting quasar candidates without u-band information can be put into four scenarios:

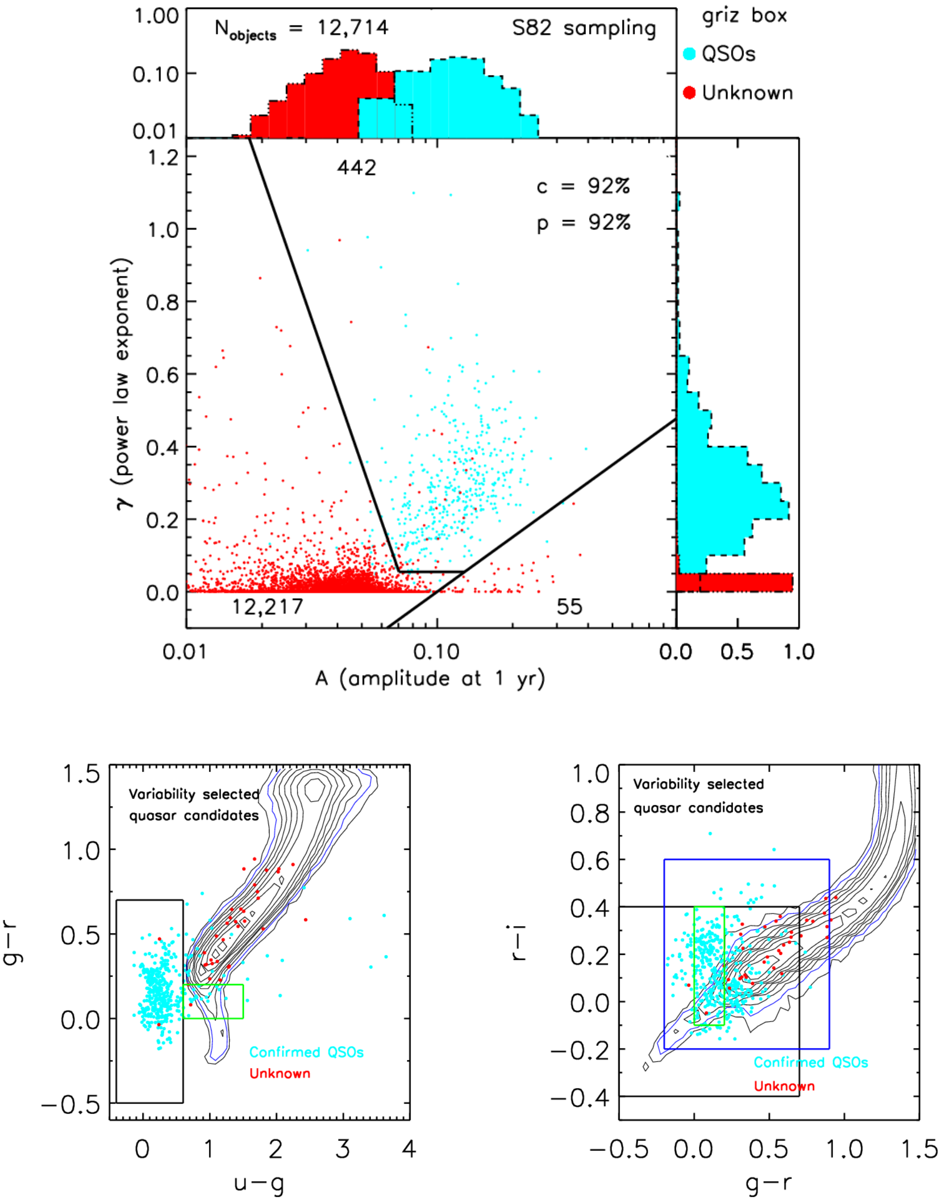

Figure 2.9: *The top panel shows the A and γ power law parameter space of the objects in the griz selection box (Sections 2.4.5 and 2.6.2.3) with Stripe 82 time sampling. The projected probability distributions for A and γ for the confirmed quasars (light blue points) and the unknown objects (red points) are shown as histograms. The solid lines indicate the selection cut defined in Equations (2.12)–(2.14). The numbers 12,217, 442 and 55 in the A–γ plane indicate the number of objects in the given region. The estimated completeness and purity is shown in the upper right corner of the scatter plot. The bottom row shows the 442 variability selected quasar candidates (407 confirmed quasars and 35 unknown candidates) from the A–γ space, projected back into ugr and gri color space. The black, green and blue boxes correspond to the UVX, nUVX and griz selection boxes (Sections 2.4.3–2.4.5 and 2.6.2.1–2.6.2.3). 87% of the confirmed quasars and 6% of the unknown candidates fall in the black UVX box in ugr color space (lower left plot). 5% and 6% of the quasars and unknowns fall in the green nUVX box. The contours in the bottom plots indicate the Stripe 82 stellar locus. 95% of the stellar locus objects are within the blue contour level (the last but one outer contour). Above 80% of the 35 unknown candidates and 41% of the 407 confirmed quasars fall within the blue gri contour, providing an estimate of the number of quasars 'hiding' in the gri stellar locus.*

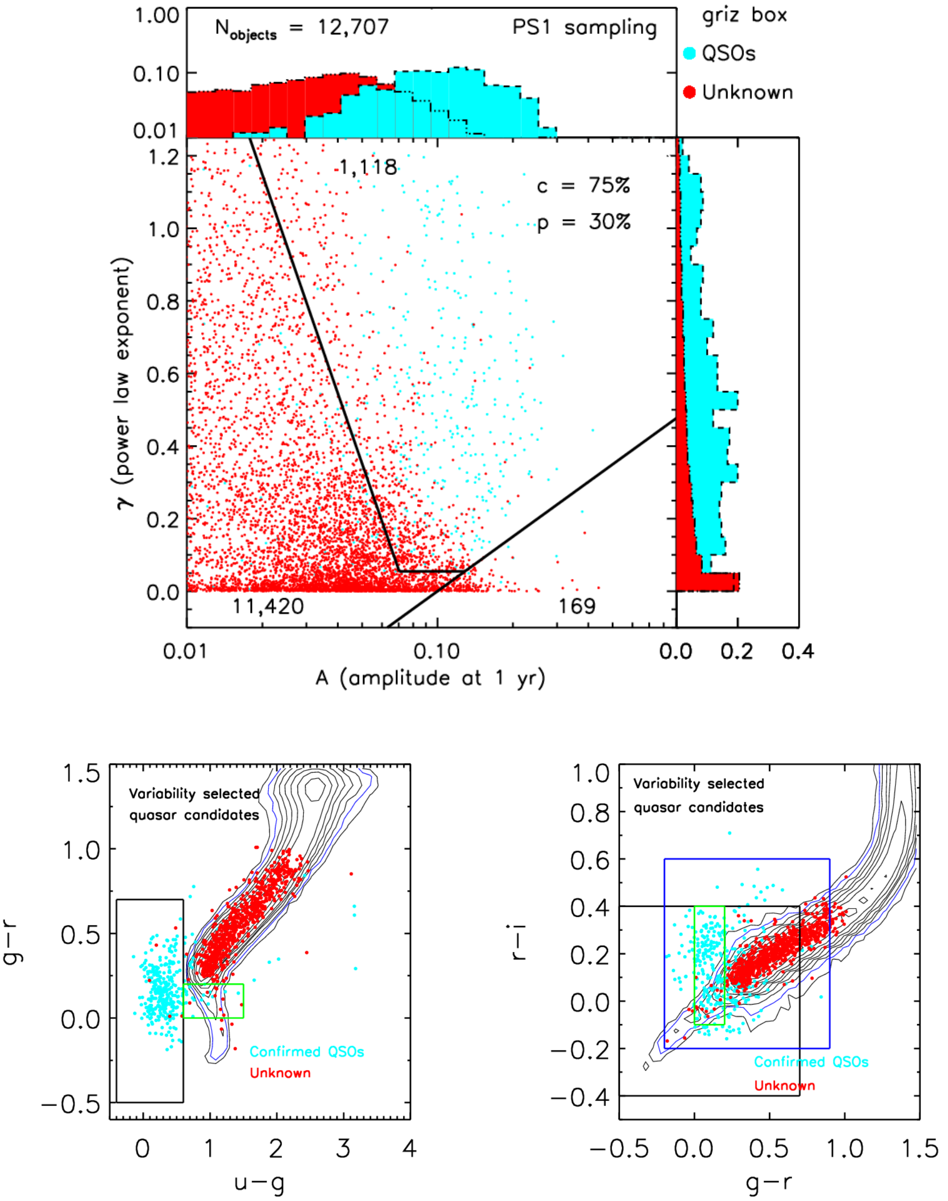

FIGURE 2.10: *Plots similar to those of Figure 2.9 for the griz selection box data down-sampled to a Pan-STARRS 1 cadence (6 epochs over 3 years). The numbers 11,420, 1,118 and 169 in A–γ space (top panel) indicate the number of objects in each of the three regions defined by the black solid lines as defined in Equations (2.12)–(2.14). The estimated completeness and purity of the 1,118 variability selected quasar candidates is shown in the upper right corner of the top panel. Of these 1,118 objects, 42% of the confirmed quasars and 98% of the unknown candidates fall within the blue gri color space (lower right plot) stellar locus contour level, which encloses 95% of the Stripe 82 stellar locus objects. 87% of the confirmed quasars fall in the black UVX selection box in ugr color space (lower left plot), as opposed to only 1% of the unknowns. In the green ugr nUVX box fall 5% and 1% of the quasars and unknown candidates. These percentages indicate that most of the unknowns are stars scattered into our selection region due to the 6 epoch sampling of Pan-STARRS 1.*

1) A 'naive' *griz* box ($i < 19.1$) including the stellar locus and not considering variability information will have a quasar selection completeness above 90%, but a purity of only 4%.

2) Taking a *griz* box color selection, but cutting out the stellar locus (defined by the (blue) contour in Figures 2.9 and 2.10) and still not considering variability information improves the purity to ~48% but lowers the completeness to ~59%. This is what would be easily achievable in a single-epoch *griz* survey.

3) Combining a *griz* color selection including the stellar locus, but considering variability information (illustrated in Figures 2.9 and 2.10) leads to a completeness of 92%(75%) and a purity of 92%(30%) for the Stripe 82 (Pan-STARRS 1) sampled data, respectively. This means that Pan-STARRS 1 can provide a 75% complete quasar sample, if 30% purity were acceptable.

4) Finally, removing the stellar locus from the *griz* selection box and combining this with variability information lowers the completeness to 54%(44%), but returns a purity of 97%(92%) with Stripe 82 (Pan-STARRS 1) time-sampled data. Pan-STARRS 1 can provide a high purity quasar sample that is ~50% complete.

This illustrates how a variability selection cut greatly improves the candidate sample as compared to quasar candidate selection based on colors alone when *u*-band information is not available. Keeping in mind that the purities calculated here are lower limits, these results suggest that the variability selection of quasar candidates in Pan-STARRS 1 or other multi-epoch surveys will produce high quality, extensive samples.

2.7 DISCUSSION

We have explored the selection of quasars by their variability in lieu of their UV excess in the context of the Pan-STARRS 1 3π survey as a test case for various upcoming multi-epoch surveys. Given its extensive, multi-year time sampling, SDSS Stripe 82 is an excellent test bed, that allows us to address variability selection of quasars (and other sources) in general. Besides Pan-STARRS 1 and SDSS, LSST (Ivezic et al. 2008; LSST Collaboration 2009) and Gaia (Perryman et al. 2001) will be able to employ variability selection similar to what we have presented.

For any source, we have characterized its variability by a simple power law model for its light curve structure function, which is similar to (but not the same as) Cristiani et al. (1996); Eyer (2002); Sumi et al. (2005); MacLeod et al. (2008). For each object the amplitude A is the typical variability within 1 year, and the exponent γ describes how the expectation value for the magnitude differences changes with the time between measurements. Applying this simple variability characterization to spectroscopically confirmed quasars, to presumably non-variable sources (F/G stars) and to known RR Lyrae, we have found that the A–γ space separates these source classes nicely. Quasars have $0.07 < A < 0.25$ and $0.15 < \gamma < 0.5$; RR Lyrae have $A \sim 0.2$ and $\gamma \sim 0$, as the rapid periodic variations in their light curves appear as 'white noise' with no secular trend when coarsely sampled on year-long time scales; finally, non-variable sources have $A < 0.05$. The variability properties for quasars derived here for individual objects with ~60-epoch light curves, are consistent with those derived by for instance Vanden Berk et al. (2004) and Bauer et al. (2009b) using ensemble averaging. The results allowed the definition of a simple variability-based quasar selection in multi-epoch data, using cuts in the A–γ plane. As we summarize quantitatively below, this variability selection works very well when considering 60-epoch light curves in Stripe 82, and still works quite well for the sparser sampling expected for Pan-STARRS 1.

The presented algorithm for variability selection is simple enough that it can be applied to large samples. Our IDL code (not optimized for speed) running on two dual-core CPUs takes 2 hours to characterize the ~12,000 objects of the *griz* color box with the Pan-STARRS 1 sampling.

The extensive spectroscopy of quasars in Stripe 82 has enabled stringent completeness and purity estimates for variability selection, at least for UVX quasars. Our analysis has shown that variability-based quasar selection with only 6 epochs over 3 years (as for Pan-STARRS 1) is effective, producing either complete or pure samples. However, something more like the ~60 epochs of Stripe 82 is needed to produce variability selected quasar samples that are both complete and pure (with only weak color pre-selection).

Therefore, it is paramount to eventually combine the variability information in different filters, in order to boost the number of epochs in Pan-STARRS 1. We can attempt to do this by predicting time-offset synthetic r-band magnitudes from the g, i, z and Y bands which could reduce the scatter in the A–γ plane, and so enhance the completeness and the purity of the Pan-STARRS 1 sampled catalogs.

Considerably further in the future, LSST (Ivezic et al. 2008; LSST Collaboration 2009) will start operating and is planned to ultimately have ~200 epochs spread over 10 years. This makes LSST optimal for creating complete and pure variability selected quasar samples. Considering a UVX box like the one used here, a quasar candidate variability selection with the LSST cadence and a 10 year baseline would definitely push both the completeness and purity above the (lower) Stripe 82 limits of 90% and 95% shown in the top left plot of Figure 2.8 and in Table 2.5. Such samples are expected to be obtainable at least down to LSST's 10σ limiting i_{AB} magnitude of around 23.3 (Oguri & Marshall 2010), or even down to 5σ ($i_{\mathrm{AB}} \lesssim 24$) with the 200 epoch variability selection. If we further consider the fact that LSST will have a *u*-band, a LSST color variability quasar selection will presumably be able to create (UVX) quasar catalogs almost as pure and complete as spectroscopic samples. Hence, LSST will be superior to Pan-STARRS 1 in the 20,000 deg^2 LSST will observe. However, until LSST happens, Pan-STARRS 1 is excellent for improving and further developing variability selection of quasars, and it will provide the first large variability (and *grizy*-color) selected quasar samples. The LSST (and Pan-STARRS 1) quasar catalogs will with estimated sizes of 100 quasars per square degree, or even more, continue the exponential growth of quasar catalog sizes seen the last 50 years (Richards et al. 2009), and thereby stress that we are on the verge of an era where confirming the quasar nature of all objects in the catalogs by spectroscopic follow-up will be unfeasible. This makes the purity of the photometric samples crucial for doing statistical analyses of the quasars. As demonstrated variability provides purities above 90% for well sampled data at all redshifts and not only for UVX object and will therefore be an important step in achieving very pure photometric quasar catalogs in the future.

2.8 CONCLUSIONS

We have presented a simple, parameterized variability characterization for sources with many epoch photometry. We do this by fitting a power-law model for the structure function of each source's light curve; the model is specified by an amplitude A and a power-law index γ.

We have applied this approach to understanding variability selection of quasars in multi-epoch, multi-color surveys, either augmenting or supplanting the more common color-selection. Specifically, we have analyzed data in Stripe 82 both to understand variability selection *per se* and as a testbed for ongoing and upcoming surveys, such as Pan-STARRS 1 and LSST. To predict Pan-STARRS 1's ability to identify quasars, we have down-sampled Stripe 82 data to a set of 6 epochs of data, resembling the expected single-band time sampling in the Pan-STARRS 1 3π survey.

For all sources in various sub-samples we calculated the parameters A and γ, and found that for sufficiently many epochs over a multi-year interval, quasars separate well in the A–γ parameter plane from non-variable sources and, e.g., RR Lyrae. The quasar variability as typically characterized by $0.07 < A < 0.25$ and $0.15 < \gamma < 0.5$, is consistent with earlier ensemble analyses.

Drawing on the nearly complete spectral identification of quasars with $i < 19.1$ in Stripe 82, we have explored both the completeness and the purity of single-band, variability selected quasar samples with both Stripe 82 and Pan-STARRS 1 time-sampling, and with or without the benefit of *u*-band photometry (which Pan-STARRS 1 does not have).

Specifically, we found the following:

- Among the complete, spectroscopically confirmed sample of UV excess quasars in Stripe 82, we can identify 90% of them on the basis of ~60 r-band epochs over ~5 years. This variability selected sample only has a 5% contamination from other objects.

- Repeating the same exercise but with the 10-times sparser Pan-STARRS 1 sampling, reduces the completeness to 73% but with still high purity (top right of Figure 2.8).

- In the redshift range $2.5 < z < 3$, where quasars overlap with the stellar locus in color-color space, variability can find 90% (65%) of all spectroscopically confirmed quasars for Stripe 82 (Pan-STARRS 1) sampling. This is a factor of 5-10 more complete than existing color-selection in this particular regime (bottom panel of Figure 2.8).

- To understand our ability to select quasars through their variability when no u-band data are available, we selected all sources in a broad $griz$ color box, known to contain almost all spectroscopically confirmed quasars. In this color box, stars and other contaminants outnumber quasars by a factor of 30 for $17 < i < 19.1$. Nonetheless, variability selection encompasses 92% (75%) of known quasars for Stripe 82 (Pan-STARRS 1) sampling. For Stripe 82 sampling (Figure 2.9) the purity of this sample is still very high (92%), but is somewhat lower with Pan-STARRS 1-like sampling (Figure 2.10), 30%.

- If in the case of Pan-STARRS 1 (6 epochs, no u-band) a more pure quasar sample is desired, this can be done by omitting the stellar locus in the $griz$ box and then looking at the variability of the remaining sources. This yields a completeness of only 44%, but with a purity of 92%.

- The high purity of the variability selected quasar sample in the $griz$ box (with Stripe 82 sampling) confirms that the fraction of overlooked quasars in Stripe 82 must be small ($<$ 10%); this inference is predicted on the assumption that the quasars ‘buried’ in the stellar locus have a variability behavior similar to the others.

- The same A–γ analysis is also very effective at identifying RR Lyrae. With Stripe 82 sampling 97% of the RR Lyrae from Sesar et al. (2010) are found, with Pan-STARRS 1 sampling still 83%; they appear as objects with $(A, \gamma) \approx (0.2, 0)$ (red points in Figure 2.6 and 2.7).

2.9 OUTLOOK

Having shown that the intrinsic quasar variability is indeed an excellent classifier and tool for putting together large, well-defined, complete, and pure samples of quasars, opens new possibilities for building upon or expanding the work presented in this chapter. As already mentioned in the discussion in Section 2.7, an obvious improvement and extension of the selection algorithm presented here, is to expand it from a single to a multi-band treatment to take advantage of the full information of Pan-STARRS 1 and LSST. Two other, maybe less obvious, extensions of the presented algorithm are (i) a combination with existing selection algorithms and (ii) selection of gravitationally lensed quasars via variability. We will describe these in the following.

2.9.1 PROBABILISTIC COMBINATION OF SELECTION METHODS

With surveys like Pan-STARRS (Kaiser et al. 2002), LSST (Ivezic et al. 2008), and Gaia (Perryman et al. 2001), among others, large amounts of data will become available in the (near) future. The science described in Section 2.1 become more feasible with quasar catalogs containing more than a million actual quasars, however, new methods for selecting quasars will also be needed to avoid spectroscopic follow-up confirmation, which becomes unfeasible with such large quasar (candidate) samples. One of these new selection methods could very well be the intrinsic variability of quasars as presented here. That being said, only relying on one method for detecting quasars is far from the optimal. Therefore, a main focus of the quasar selection community should be to combine existing, and possibly new, selection methods for quasars. As argued, quasars are extremely versatile, so the work put into such an effort will definitely pay off.

The above text has mostly focused on the selection of quasars via a structure function parameterization of variability (Schmidt et al. 2010) and the standard SDSS *ugriz* color selection (e.g., Richards et al. 2009, 2002b). However, these are (obviously) not the only methods proposed for selecting quasars in the literature. For instance, it is not only in the optical bands that quasars can be selected convincingly based on their colors. Also near-infrared colors have been shown to aid quasar selection (Hewett et al. 2006; Chiu et al. 2007; Maddox et al. 2008; Wu & Jia 2010). After the results presented in this chapter were published in Schmidt et al. (2010) also several other methods using the intrinsic variability of quasars as a selection tool have been proposed. In MacLeod et al. (2010, 2011); Kozłowski et al. (2010) and Butler & Bloom (2011) they base their color selection on the modeling of quasar light curves as a damped random walk, as presented in Kelly et al. (2009, 2010), by parametrizing the variability by the characteristic damping time-scale, τ, and the asymptotic magnitude of a structure function, SF_∞. In

Palanque-Delabrouille et al. (2011) they base a neural network scheme on the A–γ selection presented in this chapter, and in Kim et al. (2011) a selection method using support vector machines on extracted parameters such as period, variability amplitude, and color, was used to select quasar candidates. Hence, quasar variability selection presently offers several (partially correlated) selection parameters.

Combining all the methods based on colors as well as variability would without a doubt strengthen the selection of quasars. One could imagine representing each of the different selection methods and selection parameters in a probabilistic way, such that they can easily be combined to a 'unified selection' of quasars. A scheme where the overall 'grand' probability, $\mathscr{P}$, of a given object being a quasar is given by something like

$$\mathscr{P}_{\rm QSO} = \frac{1}{N_{\rm methods}} \prod_{i}^{N_{\rm methods}} w_i \, {\rm P}_i \; , \tag{2.15}$$

with i counting the different selection methods, e.g., (A, γ), $(u-g, g-r)$, $(J-K, i-Y)$, $({\rm SF}_\infty, \tau)$, (period,amplitude), etc., ${\rm P}_i$ giving the probability that the object is a quasar according to method i, and w_i representing some clever weighting, would accomplish that. The weighting scheme should be created such that cases where a given selection method does not provide a sensible ${\rm P}_i$ due to for instance lack of epochs (variability) or wavelength coverage (colors), are taken probably into account. How to actually do this in practice is still not clear, but attempts at creating probabilistic selection schemes (so far only based on colors) have been presented (Mortlock et al. 2011a; Kirkpatrick et al. 2011; Bovy et al. 2011).

An example of how combining different selection methods improves the final results is also presented in Wu et al. (2011) (though not in a probabilistic manor like proposed above). Our goal with this study was to find the 'missing' quasars at $z \sim 2.7$ mentioned in Section 2.2 combining the variability selection presented in this chapter with the optical/near-infrared color selection presented in (Wu & Jia 2010). The spectroscopic follow-up of our final candidate sample revealed a 100% success-rate, which speaks for itself and nicely illustrates the power of combining quasar selection methods.

2.9.2 SELECTING LENSED QUASARS WITH VARIABILITY

With new and larger quasar samples another very interesting prospect of the work presented here is the possibility of finding lensed quasars. As mentioned gravitational lenses, and in particular gravitationally lensed quasars, have a wide range of applications. For instance, the anomalous brightnesses of quasar lenses probe the small scale dark matter content of (the lens) galaxies (Dalal & Kochanek 2002; Dobler & Keeton 2006), microlensing of the images by stars in the lens galaxies allows the central AGN

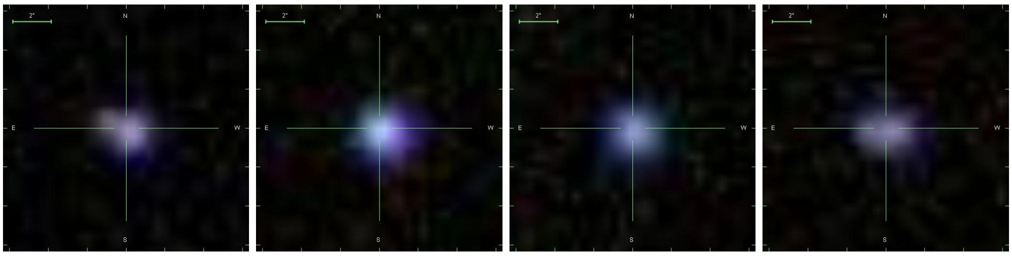

Figure 2.11: *SDSS thumbnails (approximately 12×12 arcsec) of four high priority small-separation quasar lens candidates selected via intrinsic quasar variability as described in this chapter. Confirming any of these as an actual lens will be the first time a lens has been found via variability and will serve as the proof-of-concept for dedicated variability lens searches on large areas of the sky, e.g., the 3π survey of Pan-STARRS 1.*

region of the quasars to be super-resolved (Poindexter et al. 2008), and the lens magnification enables the host galaxy to be studied, i.e., it becomes possible to study the scaling relations between the quasar host and the central black holes mentioned in Section 1.5 (Peng et al. 2006a,b). At present only ∼100 lensed quasar systems are known (Oguri et al. 2008; Inada et al. 2010), and for any given lensing application only a small fraction of these is suitable. Hence, enhancing the number of known quasar lenses will enable a more detailed exploration of many of the applications mentioned above.

In the work presented here, we have mainly focused on point source selection, with the purpose of finding quasar candidates. Such an approach would be directly applicable to a search for wide separation lenses, where (by definition) the multiple images are well-resolved. With the overwhelming sample sizes of the upcoming quasar catalogs the number of wide separation lensed quasars occurring as multiple quasar detection within a relatively small projected distance on the sky will be significant. It will be a relatively straight forward exercise to search the new quasar catalogs, for instance created with a probabilistic ‘unified selection’ as the one described above, for close pairs of quasars. There is a high chance that such pairs turn out to be multiple images of the same quasar, i.e., a lens. In the cases where they turn out to be at different redshifts, these quasar constellations are still valuable and have several interesting applications (e.g., Hennawi et al. 2006a, 2010; Myers et al. 2008).

But also small-separation quasar lenses can potentially be identified through their variability. As these are unresolved for optical intermediate resolution surveys like SDSS, applying for example the A–γ variability selection algorithm presented here to spatially extended (rather than point sources), yet quasar-colored, objects will return a list of small separation quasar lens candidates.

We have initiated such a search on SDSS Stripe 82 with promising preliminary results. By selecting objects with quasar colors, extended morphologies, quasar variability according to A and γ, and suitable photometric (or spectroscopic when available) redshift to be a lens, we have collected a sample of a few hundred small-separation lens candidates. In Figure 2.11 4 of the highest priority lens candidates are shown. We expect that of the order 10 of these candidates are indeed lenses. We have initiated a high resolution photometric follow-up program (either in near-infrared or with space based observatories) in an attempt to resolve the lens-components and possibly the lens galaxy. So far we have obtained La Silla 2.2 m GROND near-infrared photometry for around 40 of our candidates, but have not yet analyzed the data. Furthermore, a proposal for HST follow-up of some of the highest ranked candidates has recently been submitted. In cases where the photometric follow-up reveals particularly interesting (now hopefully resolved) candidates, a second *spectroscopic* follow-up is most probably needed to confirm the lens nature of the candidate. If a candidate is confirmed as a lens, this will be the first time a gravitational lens has been found based on the intrinsic variability of the source.

This pilot study, which we plan to present in a forthcoming paper, will therefore be a proof-of-concept for selecting lenses with variability and a first step in effectuating such a search on larger areas of the sky with multi-epoch data, e.g., the 3π (3/4 of the sky) area of the Pan-STARRS 1 survey, where as many as 2000 lenses are expected (Oguri & Marshall 2010).

2.10 – EPILOGUE –

As presented in this chapter, i.e., in Schmidt et al. (2010), as well as in several more recent studies (e.g., MacLeod et al. 2010, 2011; Palanque-Delabrouille et al. 2011; Butler & Bloom 2011; Kim et al. 2011), the selection of quasars based on their intrinsic variability has proven to be a very powerful tool for finding quasars, and has proven to be more robust and more efficient than the standard color selection alone. In Table 2.5 on page 43 we summarized the results from applying the developed selection algorithm, or A–γ selection, to a set of differently selected sub-samples from SDSS Stripe 82. These results show, that even without information about the UV excess in quasar colors, which is crucial to the success of color selection, quasar variability selection is outstanding. Furthermore, the results show that even in the case of relatively poor time-sampling of the data, quasar variability selection out-performs color selection in areas of color space where standard selection methods face heavy challenges. The high performance of the structure function variability selection presented here confirms that only a few percent of the quasars in SDSS Stripe 82 were overlooked.

We have also shown that besides selecting quasars the A–γ selection very efficiently selects RR Lyrae stars. This propose that a structure function based variability selection can be fine-tuned to select several kinds of variable objects, and not only quasars, taking advantage of the fact that one man's trash is another man's treasure.

That being said, all of this does *not* make the standard color selection obsolete. Abandoning the successes of previous studies because a new and better probe has been discovered, seems thoughtless. A 'unified selection' of the very successful color selection, the variability selection techniques, and any other previous or future quasar selection methods will without a doubt stand stronger against future challenges, than a selection base on just a single method. The strength of combining the different selection methods in this manner, is that *all* information is taken into account when available, clearly strengthening the quasar selection and eventually enabling the community to select quasars from the vast amounts of data with a confidence that eliminates the necessity for spectroscopic follow-up.

CHAPTER 3

THE QUASAR COLOR VARIABILITY

In this chapter we quantify quasar *color* variability using the quasar variability database from Chapter 2, i.e., *ugriz* photometry of ~9000 quasars from SDSS Stripe 82, observed over ~8 years at ~60 epochs each. We confirm previous reports that quasars become bluer when brightening. We find a redshift dependence of this blueing in a given set of bands (e.g., *g* and *r*), but show that it is the result of the flux contribution from less variable or delayed emission lines in the different SDSS bands at different redshifts. After correcting for this effect, quasar color variability is remarkably uniform, and independent not only of redshift, but also of quasar luminosity and black hole mass. The color variations of individual quasars, as they vary in brightness on year timescales, are much more pronounced than the ranges in color seen in samples of quasars across many orders of magnitude in luminosity. This indicates distinct physical mechanisms behind quasar variability and the observed range of quasar luminosities at a given black hole mass – quasar variations cannot be explained by changes in the mean accretion rate. We do find some dependence of the color variability on the characteristics of the flux variations themselves, with fast, low-amplitude, brightness variations producing more color variability. The observed behavior could arise if quasar variability results from flares or ephemeral hot spots in an accretion disk.

This chapter is based on Schmidt et al. (2012a).

3.1 – PROLOGUE –

In Chapter 2 we argued for the necessity of keep looking for quasars and expand the samples, even though they have almost reached sizes of 10^6 actual quasars. We also showed how variability has the potential of being a key tool for the success in doing so. In arguing for the need of larger quasar samples we gave several examples of areas in astrophysics where quasars are crucial (Section 2.1). One of the main reasons large samples of quasars are so interesting is because they offer the possibility to obtain new knowledge about the quasars themselves, and the physical mechanisms that cause them to vary. In particular the intrinsic variability of quasars directly probes the physics of the quasar central engine described in Section 2.2. Nevertheless, the ubiquitous time-variability of the quasar emission is still among the phenomena which are not fully understood and needs to be explained.

At present, the physical cause of quasar variability is not understood in any detail. In Section 2.2 we focused on the idea that the quasar engine, i.e., the central black hole and accretion disk of the host galaxy, gives rise to the variability when the mean steady state accretion rate changes. This seems to be the explanation for the quasar variability that the community has agreed upon. However, strictly speaking there is no evidence that this is actually what gives rise to quasar variability. As mentioned, several other mechanisms, among those starbursts in the host galaxy (Aretxaga et al. 1997; Fernandes et al. 1997), micro lensing by the host galaxy and compact dark matter objects (Hawkins 1996; Zackrisson et al. 2003), and stochasticity of multiple supernovae (Terlevich et al. 1992), have also been proposed as possible origins of optical quasar variability. Many of these suggestions have later been rejected as significant to the quasar variability such that accretion disk instabilities (e.g., Rees 1984; Kawaguchi et al. 1998; Pereyra et al. 2006), and large-scale changes in the amount of in-falling material (e.g., Hopkins et al. 2006 and references therein) are the most probable causes for the main contributors to the observed variability. But the jury is still out!

Through large efforts and dedicated searches (e.g., Schmidt & Green 1983; Croom et al. 2001; Eyer 2002; Richards et al. 2002b, 2004, 2009; Atlee & Gould 2007; D'abrusco et al. 2009; Bovy et al. 2011) large samples of quasars exist and have already been explored in much detail to aid the general understanding of quasars and the physics that governs them. The goal of this chapter is to use one of these data sets, SDSS Stripe 82, to add valuable information to the discussion about the actual cause of the observed variability in quasars and the physics governing these systems.

For this study we will use the change of the quasar color as they vary in luminosity, i.e., the color variability, as a probe of the quasar physics. Several studies have previously tried to exploit color changes to extract information about the underlying physics of quasars and AGN, whether it being through monitoring of the photometry (Giveon et al. 1999; Trèvese et al. 2001; Trèvese & Vagnetti 2002; Geha et al. 2003; Vanden Berk et al. 2004; Wilhite et al. 2005; Sakata et al. 2011) or via reverberation mapping (see Section 3.2.5 and, e.g., Peterson 1993; Kaspi et al. 2005, 2007). In a similar way we will in this chapter, use the outstanding multi-epoch data of the ~9000 quasars in SDSS Stripe 82 described in Section 2.4 to answer important questions about the quasar variability. First of all we will address the reported trend, that quasars get bluer as they brighten, to see if this data agrees with previous studies. We also compare the observations to recent thin accretion disk models to see if simulations are actually capable of reproducing the observations. Maybe most importantly, we will try to address the following crucial question when dealing with the intrinsic variability of quasars: Do the observed (color) variability agree with the general idea that quasars vary because the mean accretion rate of the central engine is changing? We will do this by exploiting the fact that:

> *The intrinsic (color) variability of quasars is a direct probe of the physics governing the central engine of active galaxies.*

3.2 PROBING THE PHYSICS OF QUASARS

The generally accepted model of AGN described in Sections 1.5 and 2.2 has several observational features that makes it possible to probe and extract information about the physics governing this spectacular galaxy evolution phase. The three main observables of AGN and quasars that provide information about the underlying physics are: photometry over most of the electromagnetic spectrum from hard x-rays to radio, the quasar spectra and their broad as well as narrow emission lines (see Section 1.5), and the variability (probably) produced by the accretion disk nature of the AGN, which we used as a selection tool to find quasar candidates in Chapter 2. In this section we will present some of the methods used to probe the physics and characteristics of quasars and their central engine, before we in the reminder of this chapter describe a new study that combines the spectral features and the variability to gain new general knowledge about AGN.

3.2.1 THE (REST-FRAME OPTICAL) QUASAR SPECTRUM

The electromagnetic spectrum is one of the main providers of precise and detailed information about the physics of AGN and quasars in particular, but of all astronomical objects in general – it is popularly said that photometry and imaging is astronomy, whereas spectroscopy is astro*physics*; which is not without a good reason. This will also become evident in the following subsections where many of them are closely linked to the information the quasar spectrum provides.

In Figure 3.1 we show the composite quasar spectrum of Vanden Berk et al. (2001). The quasar spectrum represents the spectra of the general population of AGN (see Table 1.1) and consists of four main components: (i) the absorption blueward of the Lyα emission, (ii) the power-law continuum, (iii) the broad emission lines and (iv) the narrow emission lines.

The pronounced absorption at wavelengths shorter than the Lyα emission at 1216 Å is produced by absorption from intervening columns of neutral hydrogen (HI) in the quasar line of sight, and is present in all distant quasars and AGN as well as 'normal' galaxies. Photons of energy 10.2 eV (corresponding to λ =1216 Å) can excite the neutral hydrogen from its ground state and hence be absorbed. Due to the redshifting of light, caused by the expansion of the Universe while traveling from the distant quasar to the observer, the energy of photons is diluted, i.e., the wavelength becomes longer/redder. This means that at some point all wavelengths initially more energetic than 10.2 eV in the quasar spectrum, i.e., blueward of Lyα will have been diluted to 10.2 eV so it can be absorbed by any neutral hydrogen it might encounter. Hence, the Lyα absorption lines fill up basically all wavelengths blueward of the Lyα emission (which comes from the

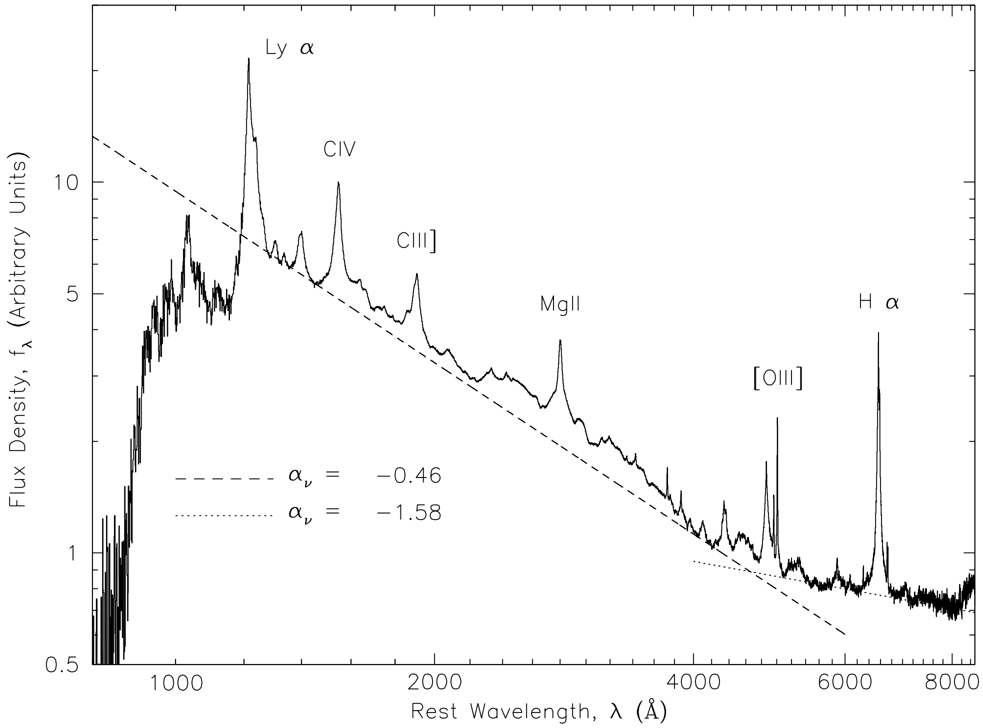

FIGURE 3.1: *Composite quasar spectrum showing the main features of most AGN spectra: the absorption bluewards of Lyα referred to as the Lyα forest, the broken power-law non-thermal continuum, and the broad and narrow emission lines. For a more detailed version of the spectrum see Figure 6 of* Vanden Berk et al. (2001). *[Adapted from* Vanden Berk et al. (2001)*]*

quasar itself) like 'trees' in a forest, and is therefore also referred to as the Lyα forest. The density of the hydrogen the photons encounter determines the amount of absorption, such that very dense regions in the line of sight, e.g., galaxies, will strongly absorb the light and give rise to pronounced broad absorption features. These systems are referred to as damped Lyα systems or Lyman limit systems. The Lyα forest therefore maps the hydrogen column densities in the line of sight to the quasar, which for quasars at high redshift, corresponds to a significant fraction of the Hubble volume. This has been used for several studies both mapping individual galaxies and the intergalactic medium (e.g., Gnedin & Hui 1998; Gnedin 1998; Cowie et al. 1995; Metcalf 2005; Hennawi et al. 2006b; Hennawi & Prochaska 2007). For further details on the Lyα forest and its applications see for example Rauch (1998); Longair (1998) and Mo et al. (2010).

The second important feature of the quasar spectrum is the power-law continuum, which as illustrated in Figure 3.1 and described in for instance Vanden Berk et al. (2001), is actually best approximated by a broken power-law in the rest-frame optical spectral range. The quasar continuum arrises from the radiating accretion disk of the AGN, i.e., it is not the sum of billions of blackbody radiating stars as the spectra of normal galaxies basically are, and is therefore non-thermal. However, the break in the power-law slope, also referred to as the ’near-infrared inflection’, is presumably caused by emission from hot dust and an increased significance of the underlying host-galaxy (stellar blackbody) light (Vanden Berk et al. 2001). Attempts at reproducing this particular quasar continuum shape in models has provided many constraints and large insight into the physics of the central engine. This is very well illustrated by the lengthy discussion of the AGN accretion disk continuum spectrum in for instance Krolik (1999). The clear distinction between the spectra of standard black-body stars (and hence galaxies) is what makes the color selection of quasars described in Section 2.2 so efficient.

The last prominent features of any quasar spectrum are the broad and narrow emission lines. These are produced in the broad and narrow line regions (BLR and NLR) described in Section 1.5 and illustrated in Figure 1.4 on page 14. Several processes contribute to the line emission in AGN, for example the narrow lines might have a significant contribution from the narrow lines of the host galaxy (stellar) spectrum. However, by far the most important process is photoionization of the BLR and NLR clouds by the non-thermal continuum radiation of the accretion disk. The emission lines are thought to be absorption of high energy photons which are then re-emitted at the characteristic wavelengths of the atomic electron-levels. The width of the individual lines are determined by the Doppler-broadening arising from the kinematics of the BLR and NLR clouds. ‘Broad’ lines are usually defined as lines with widths $\sigma_v > 1000$ km/s, whereas narrow lines usually have $\sigma_v \sim 100$ km/s, as also noted in Table 1.1 on page 13. Quasar and AGN emission lines have proven very useful for gaining information about quasar physics as the following subsections will show. For a detailed description of AGN emission lines see Krolik (1999) and Vanden Berk et al. (2001).

To summarize, the characteristics of the continuum as well as the broad and narrow lines are tied to the physics of the different components of the standard AGN model (illustrated in Figure 1.4), whereas the Lyα forest characterizes the HI content between us and the quasar.

3.2.2 EDDINGTON RATIOS AND BLACK HOLE MASSES

Two of the main parameters used in the literature to describe AGN is the mass of the central black hole and the Eddington ratio. The mass of the central black hole is often obtained by inferring the velocity of the BLR (v_{BLR}) from measurements of the velocity broadening of the broad emission lines ($\sigma_{v_{\mathrm{BLR}}}$) in the AGN spectrum, and combining that with simple assumptions about the characteristic size of the BLR (r_{BLR}; or obtained from reverberation mapping as described below) such that

$$M_{\mathrm{vir}} \propto \frac{r_{\mathrm{BLR}} v_{\mathrm{BLR}}^2}{G} \quad . \tag{3.1}$$

To be specific, what is actually obtained is the *virial* mass enclosed by the BLR clouds, which is assumed to approximate the actual mass of the central black hole, M_{BH}. This is done under the assumption that the gravity (and not radiation pressure or magnetic fields) dominates the velocity broadening of the emission lines. Corrections of this assumption have been discussed (e.g., Peterson 2010).

With an estimate of the central black hole mass, the Eddington Luminosity is easily obtained via Equation (2.4) on page 24, and from that the Eddington ratio, which is defined as the ratio between the bolometric luminosity of the AGN and the Eddington Luminosity, i.e., ($L_{\mathrm{bol}}/L_{\mathrm{Edd}}$). The Eddington ratio is also calculated under certain assumptions. Since its definition requires an estimate of the bolometric luminosity, i.e., the luminosity corresponding to the integrated flux of the full AGN spectral energy distribution, a bolometric correction (Shen et al. 2008) is assumed to link observable luminosities to the bolometric luminosity. The Eddington ratio indicates the accretion state of the AGN, as it relates to actual luminosity of the object to the theoretically determined Eddington accretion limit (see Section 2.2).

Several catalogs of AGN and quasar samples with these two key parameters have been published and discussed in the literature (e.g., Peterson et al. 1998; Kaspi et al. 2000; Peterson et al. 2004; Wilhite et al. 2008; Bauer et al. 2009b; Peterson 2010; Shen et al. 2011; Rosa et al. 2011; Shen & Liu 2012). In the study presented in this chapter we will use the catalog of Shen et al. (2011) to obtain these parameters for the Stripe 82 quasar sample as described in Section 3.6.1.

3.2.3 EMISSION LINE RATIO DIAGNOSTIC – THE BPT DIAGRAM

Not only do the (broad) emission lines of quasars and AGN provide information about the physics of their hosts, they can also help distinguishing AGN from 'normal' emission line galaxies, and have been used to divide the AGN population into individual subpopulations (Kewley et al. 2001; Kauffmann et al. 2003; Kewley et al. 2006; Trouille et al. 2011), as was also illustrated in Figure 1.4 and becomes clear when inspecting Table 1.1. One of the preferred methods to make such a distinction is the so-called BPT diagram first presented in Baldwin et al. (1981). In Figure 3.2 we show an example of such a classification scheme from Trouille et al. (2011). Here the line ratios of the prominent [OIII] and Hα emission lines with respect to Hβ and [NII] are plotted. These line ratios are a good proxy for the degree of ionization of the gas in the observed galaxy. The level of ionization reflects the ability of the radiation to actually strip the electrons from the atoms in

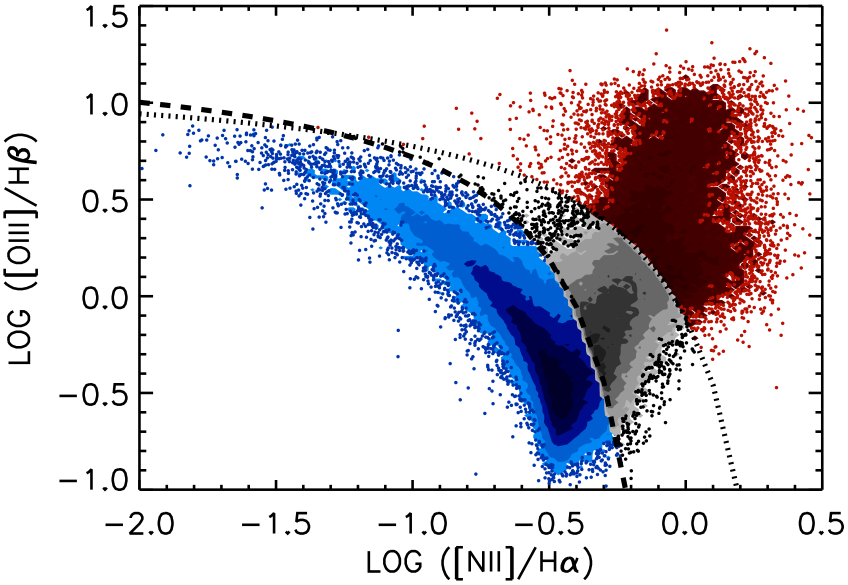

FIGURE 3.2: *The BPT diagram. The plotted emission line ratios [O*III*]/Hβ and [N*II*]/Hα are a good proxy for the level of ionization of the observed gas, and since AGN spectra are generally harder than 'normal' star-bursting emission line galaxies' spectra, the BPT diagram, i.e., the emission line ratios, serve as an excellent AGN classifier. The blue points are star-bursting galaxies, the gray point are composite galaxies (part AGN and part H*II *star-burst galaxy;* Kewley et al. 2006*), and the red points are AGN. The dashed and dotted lines separating these regions are defined in* Kewley et al. (2001) *and* Kauffmann et al. (2003)*. With other line ratios the AGN cloud can be subdivided into different classes of AGN (e.g.,* Kewley et al. 2006*). [Adapted from* Trouille et al. (2011)*].*

the gas via photoionization, and since the quasar/AGN spectrum has a much larger fraction of high-energy (low wavelength) photons than ordinary galaxies, i.e., AGN spectra are harder than normal galaxies', the BPT diagram, or 'line-ratio'-diagram, has proven very useful for distinguishing AGN from normal star-bursting galaxies as well as AGN diagnostics. The applicability of the BPT diagram is of course limited by the availability of the emission lines and has mostly been applied to low-redshift galaxies (from for instance SDSS) where the emission lines fall in the optical pass-bands. However, if the [OIII]–Hβ and Hα–[NII] emission line features can be followed into the near-infrared as the object redshift increases, the BPT diagram diagnostics can also be applied out to higher redshifts. We will illustrate in Chapter 4 that this is in principle possible with for example the 3D-HST survey, though higher resolution follow-up (Section 4.9.2) is needed to obtain confident line ratios, as only [OIII]–Hβ is (marginally) resolved in the 3D-HST spectra.

3.2.4 The Variability of Quasars

The different models of AGN variability mentioned in Section 2.2 have variability on many different timescales. From weeks for changes on thermal timescales in the accretion disk, over months for superpositions of stochastic processes, to several years for viscous changes in the large-scale structures of the accretion disk and for lens crossing times. These 'physical' timescales of AGN (e.g., Webb & Malkan 2000; Collier & Peterson 2001 and references herein) can be compared to the observed AGN variability timescales. The observed variability time-scales span the range from a few hours (Stalin et al. 2004; Gupta & Joshi 2005), possibly the result of processes in a jet (Kelly et al. 2009), to months and years where quasars as mentioned are known to typically vary by $\gtrsim$ 10% (e.g., Giveon et al. 1999; Collier & Peterson 2001; Vanden Berk et al. 2004; Rengstorf et al. 2004; Sesar et al. 2007; Bramich et al. 2008; Wilhite et al. 2008; Bauer et al. 2009b; Kelly et al. 2009; Kozłowski et al. 2010). A longstanding challenge for any AGN theory is to disentangle these processes based on variability-timescales alone, and from that sort out which variability mechanisms are prevalent under what circumstance, and which mechanisms are responsible for which variations.

Moreover, as described in Chapter 2, AGN generally have power-law-shaped structure functions for the temporal variations, i.e., with no particular timescales, and because measurements of 'characteristic variability timescales' are likely dominated, or at least influenced by window functions, interpretations of variability time-scales are always somewhat problematic. All this being said, as described earlier, there seems to be a general consensus in the community, that the most probable scenario for the majority of the observed variability behavior is changes in the accretion disk.

The many observations of these intrinsic variability timescales of quasars and AGN provide a unique tool for probing the physics and the geometry of the central regions of the AGN; for instance via performing reverberation mapping.

3.2.5 Reverberation Mapping - Measuring Sizes of BLRs

Reverberation mapping deals with the monitoring of continuum variations in AGN, and the determination of how the emission lines react to that variation in form of an echo or reverberation, to constrain the geometry of the inner regions of the AGN. Blandford & McKee (1982) outlined the theoretical framework of reverberation mapping (and named it), but also noted that due to the variability timescales expected for a standard AGN, the observational challenges are many. Since 1982 many improvements have been made to the observational campaigns for reverberation mapping, which in 2010 had let to the precise measurements of the size of the BLR in ~45 AGN (Peterson 2010).

The goal of reverberation mapping is to constrain the so-called transfer function, which describes the BLR in velocity–time-lag (v, τ) space (e.g., Blandford & McKee 1982; Peterson 1993; Netzer & Peterson 1997). Hence, the key observations are velocity of the individual BLR clouds, and the time-lag between the variations of the spectrum continuum and the emission lines. As mentioned above, part of the observed variability in AGN is of the order days to months. These requirements, i.e., finely sampled (days) observations over long periods of time (years), are what make spectroscopic reverberation mapping an observational challenge. If the data is available the structure of the BLR can be determined by assuming that the observed time delays are caused by light traveling times between the central continuum source and the BLR, implying that

$$r_{\mathrm{BLR}} = c\,\tau\ . \tag{3.2}$$

This is the simplified version of the transfer function where the velocity is assumed to be the same for all BLR clouds. The size of the time-lag is therefore a proxy for the BLR size. We know from detailed reverberation mapping studies that τ can be of the order years for quasars at $z \sim 1$ (e.g., Kaspi et al. 2005, 2007). This can be done for individual pairs of epochs of observations, and has let to the discovery of a strong correlation between r_{BLR} and the luminosity L (e.g., Kaspi et al. 2000), with $r_{\mathrm{BLR}} \propto L^{0.5}$. This size-luminosity relation and the estimate of the BLR size are used to calibrate the single-epoch mass estimates obtained with Equation (3.1). Hence, with the benefit of the statistic of the many observation epochs in reverberation mapping campaigns, which give reliable (average) emission line widths, reverberation mapping provides reliable *direct* measurements of the black hole mass (via Equation (3.1)) and the size of the BLR (e.g., Kaspi et al. 2005, 2007; Peterson et al. 2004), as opposed to other more indirect estimates, such as scaling relations and AGN line widths (e.g., Vestergaard & Peterson 2006).

To obtain the interplay between the continuum and emission lines, historically reverberation mapping has only been done by monitoring the AGN spectroscopically. However, the results presented in this chapter as well as the work by Haas et al. (2011) and Chelouche & Daniel (2012), illustrate the possibility of performing *photometric* reverberation mapping, making reverberation mapping of large samples of objects feasible, with the potential of improving the number of robustly measured BLR sizes and M_{BH} by orders of magnitude.

For further details on spectroscopic reverberation mapping see Peterson (1993); Netzer & Peterson (1997) and Peterson (2010)

3.2.6 SPECTRAL HARDNESS – BRIGHTER MEANS BLUE

The last diagnostic of AGN and quasar physics we will describe in this short summary, and which will be a main topic of the rest of this chapter, is the characterization of the emitted spectrum described above and illustrated in Figure 3.1. Available multi-wavelength AGN variability data provide clear, albeit somewhat qualitative, evidence that quasars tend to get bluer when they get brighter (e.g., Giveon et al. 1999; Webb & Malkan 2000; Trèvese et al. 2001; Trèvese & Vagnetti 2002; Geha et al. 2003; Vanden Berk et al. 2004; Wilhite et al. 2005; Sakata et al. 2011), i.e., the quasar spectrum gets harder with relatively more high-energy (blue) radiation. Whether this is a consequence of hardening of the variable spectrum, or because the variable blue component becomes more dominant compared to the non-variable red spectral component of the underlying host galaxy, is a matter of debate (Sakata et al. 2010). This correlation between luminosity and spectral shape of the quasar provide information on the variability type, and can be used to further constrain AGN (variability) models.

One practical caveat to almost all the photometric 'brighter makes bluer' claims mentioned above, is that they are based on fitting in flux vs. color space, without accounting for the color-magnitude error correlation in the modeling. Such fitting may lead to spurious, or at least biased, results, as we describe in this chapter and illustrate in Appendix B on page 145. To avoid these error correlations such analyses should be performed on flux-flux space. We will quantify this further below, as we in this chapter develop a better unbiased fitting procedure to quantify whether brighter-makes-bluer on short time-scales (see also Sakata et al. 2011).

Furthermore, we will in the reminder of this chapter, carry out a comprehensive study of color variability in quasars, i.e., we study how flux variability is linked to changes of the (observed) optical colors. We provide both a detailed empirical description of the observed variability, and work on linking it to the physics of the quasar central engine, by quantifying the correlation of color variability with the redshift of the quasars, their $M_{\rm BH}$, $L_{\rm bol}/L_{\rm Edd}$, and the temporal behavior of the flux variations. While the temporal behavior has been studied extensively, there are no comprehensive studies of color variability in large quasar samples with many epochs of data, which as described above, offers great potential as a diagnostic of accretion disk physics.

In Section 3.3 we re-introduce the SDSS Stripe 82 data which was also used in Chapter 2. We describe the development of the unbiased fitting procedure in Section 3.4 and in Section 3.5 we present the color variability results from applying this procedure to the Stripe 82 data. In Section 3.6 we discuss our findings and describe how the color variability depends on the light curve variability properties. Furthermore, we check for $L_{\rm bol}/L_{\rm Edd}$ and $M_{\rm BH}$ dependences, and compare our results to recent accretion disk models, before we conclude in Section 3.7. We finish by presenting a couple of the immediate next steps this work leads to in Section 3.8. The central findings of this chapter are reflected in Figures 3.6, 3.9, and 3.11 on page 77, 81, and 84, respectively.

3.3 SDSS STRIPE 82 DATA

The Sloan Digital Sky Survey's (SDSS's) Stripe 82 is as mentioned in Section 2.4 an equatorial stripe 2.5 degrees wide and about 120 degrees long which has been observed many times in the 5 SDSS bands over more than 8 years. The Stripe 82 data base therefore provides an unprecedented collection of data for variability studies in general (e.g., Ivezić et al. 2007; Bramich et al. 2008; Sesar et al. 2007, 2010; Bhatti et al. 2010) and quasars in particular (Kelly et al. 2009; Schmidt et al. 2010; MacLeod et al. 2010, 2011; Butler & Bloom 2011). The analysis presented here is done on the ~9,000 spectroscopically confirmed quasars in Stripe 82. They have been selected from the SDSS data archive[1] as described in Chapter 2 (Schmidt et al. 2010). Thus, we have a sample of quasars with on average 60 observations spread over a period of roughly 8 years. To obtain further information on each individual quasar, such as for instance estimates of the bolometric luminosity ($L_{\rm bol}$) and black hole mass of the central engine ($M_{\rm BH}$), we cross-matched our list of objects with the catalog of quasar properties presented in Shen et al. (2011). We found a total of 9093 matches which constitute the catalog we will use in the remainder of this chapter, unless noted otherwise. This corresponds to basically the complete catalog of spectroscopically confirmed quasars in Stripe 82, hence, matching to the Shen et al. (2011) catalog did not cut down the sample much.

3.4 FITTING COLOR VARIABILITY IN MAGNITUDE SPACE

In principle optical *color* variability of quasars, i.e., the tendency of changing color, generally becoming bluer when they brighten, has been well established as described in Section 3.2.6. However, as noted this color variability has been established and quantified by fitting data in color-magnitude space, for instance in g vs. $(g-r)$ space, which suffers from co-variances between the color and the magnitude uncertainties that have not been accounted for in the past analyses. As we describe in Appendix B on page 145, we have found that this may lead to severe overestimates of the color variability, especially as the photometric errors are not negligible compared to the intrinsic variability amplitudes. To remedy these biases we fit the color variability in magnitude-magnitude space, and then 'translate' them into color-magnitude relations.

---

[1] `http://casjobs.sdss.org/CasJobs/default.aspx`

The Stripe 82 data presented in Sections 2.4 and 3.3 are unprecedented in their combination of time coverage, number of epochs, filter bands and sample size, which allows us to take the color variability analysis to the next level. Since the quasar variability typically shows modest amplitude (a few tenths of a magnitude or less) it has been characterized in previous work by a linear relation in flux-flux space (e.g., Choloniewski 1981; Winkler 1997; Suganuma et al. 2006; Sakata et al. 2010, 2011). We make the *ansatz* that the photometric measurements of each individual quasar can be represented by a linear relation in *gr*-space (and *ui*-space). The *gr* and *ui* spaces were chosen for high signal-to-noise and broad spectral range respectively (the *uz*-space being too uncertain due to *z*-band measurement errors). By calculating the Pearson correlation coefficient (PCC) for each individual quasar we verified that this approach is indeed sensible. Averaged over the full sample in *gr*-space the PCC = 0.8. A PCC of 1 indicates a linear correlation with basically no scatter. Including an extra parameter in the linear relation to resemble the intrinsic scatter when determining $s_{\rm gr}$ also shows that the scatter of the assumed linear relation is insignificant. Hence, the data do indeed resemble a linear relation quite closely.

When fitting such data a number of factors need to be taken into account: there are comparable errors along both axes, e.g., for g and r magnitudes, the errors vary widely among different data points, and there are 'outliers' (as described in Section 2.4.6 and in Schmidt et al. 2010). In order to take these factors properly into account we have used the linear fitting approach, including outlier pruning, laid out in Hogg et al. (2010) page 29ff. In practice we identify the set of relations that make the data likely outcomes

$$r - \langle r \rangle = s'_{gr}(g - \langle g \rangle) + b \tag{3.3}$$

by a Metropolis-Hastings Markov chain Monte Carlo (MCMC) approach. Thus, we are determining the color variability (slope) s'_{gr} and the offset on the r-axis b (moving each object to its mean g and r value to improve the determination of b) by sampling the parameter space via a MCMC chain. From simple algebraic manipulations of Equation (3.3) we have that

$$\begin{aligned} r &= s'_{gr} g + b' && (3.4)\\ g - r &= -(s'_{gr} - 1)\,(g - \langle g \rangle) + B\ , && (3.5) \end{aligned}$$

where $b' = \langle r \rangle - s'_{gr}\langle g \rangle + b$ and $B = -b + (\langle g \rangle - \langle r \rangle)$ are constants. These equations give the 'transformation' of the fit between magnitude-magnitude space and color-magnitude space. If $(s'_{gr} - 1) < 0$, i.e., if $s'_{gr} < 1$ the quasar gets bluer as it brightens. In the remainder of this chapter we will use

$$s_{gr} = (s'_{gr} - 1) \tag{3.6}$$

as our definition of the color variability. In more general terms this corresponds to $s_{\lambda_1\lambda_2} \equiv \frac{\partial m_{\lambda_2}}{\partial m_{\lambda_1}} - 1$ where the λs refer to the photometric bands. This expression has the intuitive

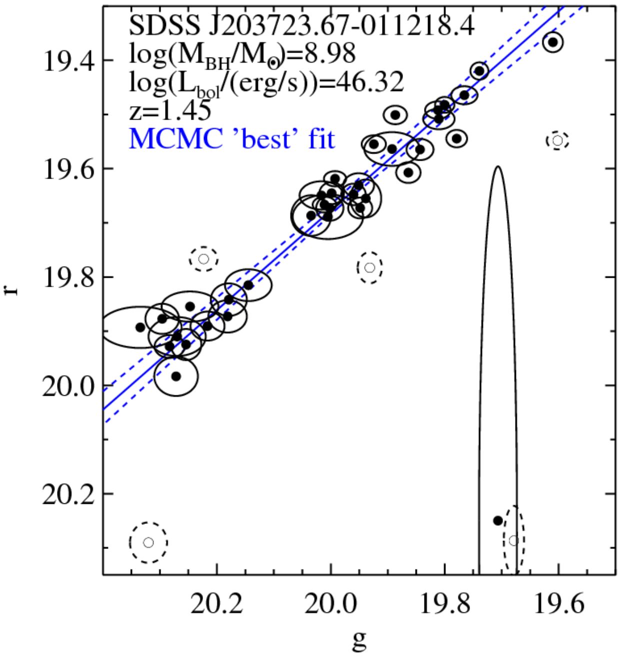

Figure 3.3: *Example of MCMC fitting for the gr color variability, given a set of independently measured g and r light curve points and their uncertainties. These photometric data for the quasar SDSS J2037-0112 are shown as black circles with error ellipses indicating the photometric errors on the measurements. The solid blue line indicates the MCMC 'best' fit (i.e., the parameters for which the data are most likely) of the linear relation in Equation (3.3) to the data that resulted from the MCMC chain, as described in Section 3.4 and in Hogg et al. (2010). The dashed blue lines show the 68% confidence interval of the MCMC. The empty data points with dashed error ellipses have a posterior probability of being outliers to the relation of > 50%, as described in Section 3.4; see also Section 3 in Hogg et al. (2010). Note that the filled point in the lower right corner is not counted as an outlier, due to its large photometric error in r. This fitting accounts for the independence of the g and r measurements, the widely varying error bars, and the fact that even data as well-calibrated as SDSS Stripe 82 have 'outliers'. In the upper left corner we note the mass of the central black hole and the bolometric luminosity taken from the value-added quasar catalog presented in Shen et al. (2011), and the spectroscopic SDSS redshift of the object. In this manner the gr and ui color variability of each of the 9093 Stripe 82 quasars in the sample were determined.*

interpretation that $s_{gr} = 0$ means no color variability, i.e., brightness variation at constant color, whereas $s_{gr} < 0$ accounts for the bluer equals brighter trend (the most common trend in the data) and $s_{gr} > 0$ implies objects that become redder when they brighten.

In Figure 3.3 an example of such a MCMC fit to constrain the color variability is shown for the quasar SDSS J2037-0112 in gr space. The filled circles represent the individual photometric epochs from Stripe 82, with the ellipses indicating the photometric errors in g and r. The 'best' fit from Equation (3.3) is shown as the blue solid line. The blue dashed lines show the 68% confidence interval given by the 16th–84th inter-percentile range of the MCMC 'cloud' of possible fits. The described fitting procedure also allows estimating the probability that a given data point is an outlier to the obtained relation, i.e., an estimate of the posterior probability that each individual observation 'belongs' to the obtained relation (see Section 3 in Hogg et al. 2010). Such outliers can be due to for instance weather, bad calibration, image defects, etc. The data points represented by the open circles with the dashed error ellipses in Figure 3.3 have a posterior probability of being outliers to the shown MCMC fit which is larger than 50%. On average 8% and 19% of the observed epochs were counted as outliers to the obtained relations when fitting in gr and ui-space respectively. Similar fits in gr (and ui) space were performed for all 9093 quasars in the Stripe 82 sample, each resulting in estimates for the gr and ui color variability for each object. Note that the temporal ordering of the flux points plays no role in this analysis.

3.5 RESULTS

In the following subsections we will present the results from the investigation of the color variability, s_{gr} and s_{ui}, for the 9093 spectroscopic Stripe 82 quasars.

3.5.1 COLOR VARIABILITY IN gr

Figure 3.4 shows the immediate result of the MCMC fitting procedure; the directly observed gr color variability for the 9093 spectroscopically confirmed quasars from Stripe 82 as a function of their spectroscopic redshifts (left panel). It is clear that the vast majority of the quasars show color variability $s_{gr} < 0$ (i.e., they get bluer when they brighten) represented by the shaded region. The right panel shows the analogous plot for the ui color variability, which we consider further in Section 3.5.3.

Figure 3.4 shows that there is a very significant redshift dependence on the mean observed color variability, $\langle s_{gr/ui} \rangle(z)$. Spectra (Wilhite et al. 2005) and other information suggest that the $\langle s_{gr/ui} \rangle(z)$ behavior seen in Figure 3.4 arises from a general trend of bluer continuum color in higher flux states modified by the redshift-dependent influence of emission lines in a given observed bandpass. As we will show in detail below, such a description is consistent with trends in the Stripe 82 data.

The colors of quasars in general are known to have a pronounced redshift dependence resembling that seen in Figure 3.4 because emission lines and the continuum affect static

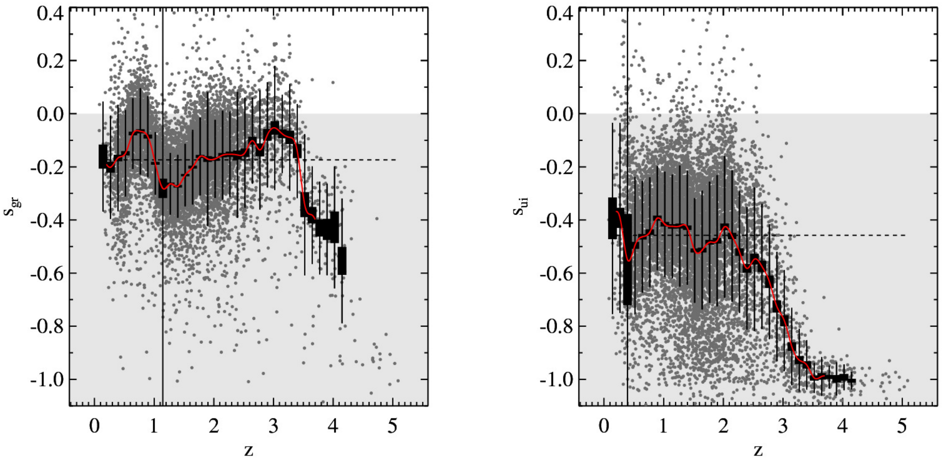

FIGURE 3.4: *The color variability, $s_{\lambda_1\lambda_2}$, in gr (left) and ui-space (right) for the full sample of 9093 Stripe 82 quasars (dark gray dots), as a function of redshift. The black rectangles show the sample mean and its uncertainties in redshift bins of $\Delta z = 0.125$, containing at least 5 data points. The width (σ) of the color variability distribution is indicated by the thin error bars. The red curve gives the interpolated mean redshift trend, $\langle s_{gr}\rangle(z)$, and the black dashed line indicates the sample median color variability of -0.17 (-0.46) in gr (ui). The shaded region indicates where s_{gr} and $s_{ui} < 0$, i.e., where bluer means brighter. The complex redshift-dependence of color variability is due to the influence of various emission lines (see discussion in Section 3.5.1 and 3.5.2).*

or single epoch colors (e.g., Richards et al. 2001; Wilhite et al. 2005; Wu & Jia 2010; Meusinger et al. 2011). For instance, the strong drop in $\langle s_{gr}\rangle(z)$ at $z \sim 0.95$ in Figure 3.4 corresponds exactly to the redshift where the MgII line moves from the g to the r band. This is illustrated in Figure 3.5, where the left panel shows the Vanden Berk et al. (2001) composite quasar spectrum from Figure 3.1 with the 5 SDSS bands as they would be positioned if the quasar was at $z = 0.95$. The right panel shows the ratio between the emission line flux (F_{line}; the composite spectrum minus the estimated continuum flux) and the estimated continuum flux (F_{cont}; modeled as a simple power-law) as a function of redshift. The shift of the MgII line from the g (green) to the r (yellow) band is marked. Likewise the dips and bumps in $\langle s_{gr}\rangle(z)$ at $z \sim 1.85, 2.8$ and 3.5 in the left panel of Figure 3.4 are attributable to the CIII], CIV and Lyα lines (see Figure 3.1) shifting between the g and r bands respectively. In the right panel of Figure 3.5 only $z \lesssim 2.4$ is shown since this is the region where a simple single-power-law approximation of the continuum is valid. For higher redshift the g and r bands move blue-ward of the Lyα line.

In order to isolate the continuum color variability for comparison with other quantities such as $L_{\text{bol}}/L_{\text{Edd}}$ and M_{BH} (Section 3.6.1), we need to eliminate the source redshift dependence induced by the emission lines. This is done by ‘emission line correcting’ the

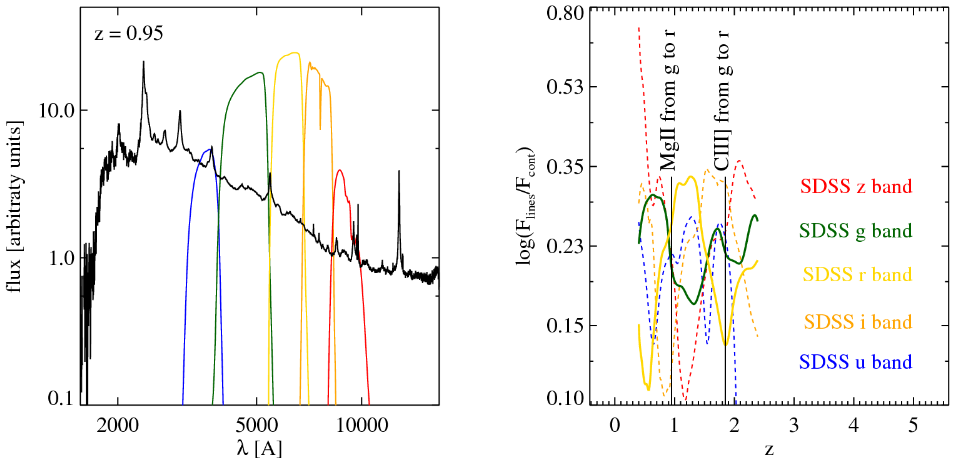

FIGURE 3.5: *The role of emission lines in the color variability. On the left, the* Vanden Berk et al. (2001) *composite SDSS spectrum from Figure* 3.1 *is shown with the 5 SDSS filter's response curves (ugriz from left to right; arbitrarily scaled for visibility) as they would fall if the quasar was at* $z = 0.95$*. In the right panel the expected ratio between the line flux* F_{line} *and the continuum flux* F_{cont} *for the same spectrum is shown as a function of redshift with the SDSS g (green) and r (yellow) bands shown as solid curves and the u (blue), i (orange) and z (red) bands shown as dashed curves. The redshift at which the Mg*II *and C*III*] lines move from the g to the r band (*$z \sim 0.95$ *and* $z \sim 1.85$ *respectively) has been indicated. This illustrates the redshift dependence of the color variability seen in Figure* 3.4 *and described in the text. The redshift range in the right panel has been set equal to the one used in Figure* 3.4 *to ease comparison.*

individual values of s_{gr} and s_{ui} by the quantity

$$\langle s \rangle - \langle s_k \rangle(z) \quad . \tag{3.7}$$

Here the first term is the mean color variability of -0.17 (-0.46) in gr (ui) which is our stand-in for the emission line free color variability (dashed line(s) in Figure 3.4). The second term is the mean color variability for the sample capturing the mean redshift dependence of the sample as depicted by the red line in Figure 3.4 for each of the individual quasars, k.

3.5.2 REPRODUCING THE COLOR VARIABILITY REDSHIFT DEPENDENCE WITH SIMPLE VARIABILITY MODEL

We now quantify to which extent a simple spectral variability model can reproduce the observed redshift trends in $\langle s_{gr} \rangle(z)$. We do this by integrating a time-varying sequence of mock spectra created from the Vanden Berk et al. (2001) composite quasar spectrum over the SDSS g an r filters as illustrated in the left panel of Figure 3.5. After decomposing the Vanden Berk et al. (2001) spectrum in a continuum and line component, by subtracting the estimated power-law continuum from Vanden Berk et al. (2001) illustrated by the long-dashed line in Figure 3.1, we varied both the continuum and line flux to create a mock time sequence of spectra for which we obtained g and r light curves and then s_{gr}. By changing the slope of the continuum (with a pivot-point in the infrared to ensure $s_{gr} < 0$) and scaling the line response by a given amount, a sequence of spectra could be created to simulate a variable quasar. The line response was characterized by the ratio between the total integrated change in continuum flux and the total change in line flux over the modeled wavelength range

$$\alpha = \frac{\delta F_{\text{line}}}{\delta F_{\text{cont}}} \tag{3.8}$$

and could be set free (both lines and continuum can vary freely) or be fixed. Several setups for creating the sequence of variable spectra were inspected. Among those setups were fixed line contribution with changing continuum slope and both continuum and lines changing in various ways. For given α the emission lines are assumed to respond instantly to the continuum variation; i.e., in this simplistic approach we ignore any of the existing reverberation time-delay between the continuum and the lines mentioned in Section 3.2.5. An exploration of this effect to carry out reverberation mapping (e.g., Peterson et al. 2004; Kaspi et al. 2005, 2007; Haas et al. 2011; Chelouche & Daniel 2012) using the broad-band light curves seems promising in light of Figure 3.4, but is beyond the scope of this study (see Section 3.8.2). Details on the simple spectral variability models are given in Appendix C on page 149.

The predictions of the spectral variability models are shown in Figure 3.6 together with the estimated values of s_{gr}, shaded regions, and the mean redshift dependence, $\langle s_{gr}\rangle(z)$, from the left panel of Figure 3.4. This Figure shows that $\langle s_{gr}\rangle(z)$ is best matched if the (implicitly instant) line response is very sub-linear: $\alpha = 0.1$ (purple line in Figure 3.6) is a much better fit than the model with $\alpha = 1$ (red line in Figure 3.6). It is seen that for emission lines that vary in lockstep with the continuum by $\alpha > 25\%$ the redshift features in $\langle s_{gr}\rangle(z)$ are 'inverted'. Actually, unresponsive line fluxes (i.e., $\alpha = 0$), lead to the best match in this model context (black dashed curve in Figure 3.6). Overall, Figure 3.6 tells us that the redshift dependence of the gr color variability is nicely reproduced by a simple spectral variability model where the continuum of the spectrum is hardened, i.e., its power-law slope is changed so brighter makes bluer, and the emission line fluxes in the two bands are (instantly) unresponsive. As mentioned in Section 3.2.5 we know from detailed reverberation studies that emission lines do respond (on $\sim$1/2 year timescales Kaspi et al. 2005, 2007) for quasars at $z \sim 1$. The explanation for $\alpha \approx 0$ may be that the lags in the lines are long enough to introduce a phase offset whereby the line is sometimes stronger, sometimes weaker than predicted from a tight correlation and in the net the correlation gets lost. As noted in Appendix C on page 149 the used variability

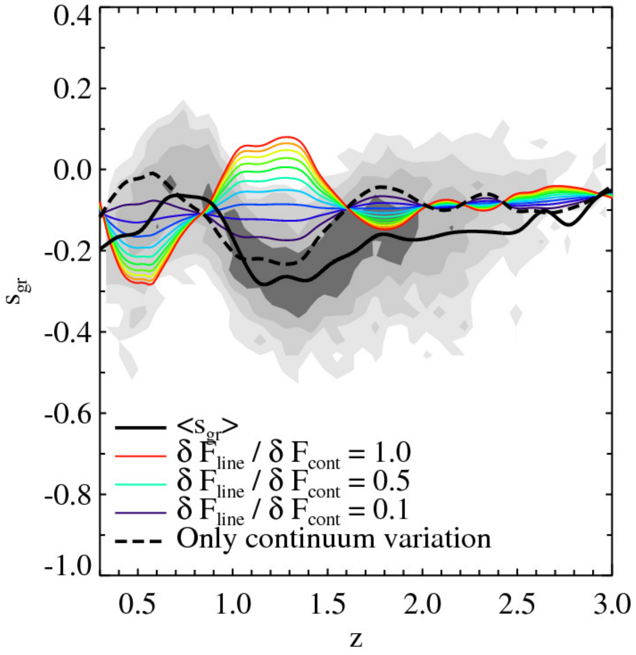

FIGURE 3.6: *Comparison of the observed redshift dependence of* s_{gr} *with our simple spectral variability models. The mean* $\langle s_{gr}\rangle(z)$ *is shown by the solid black line (cf. Figure 3.4). The rainbow colored curves show the color variability models with fixed ratios* α *(see Equation (3.8)) between the line variability and continuum variability from a ratio of 0.1 (purple) to a ratio of* $\alpha = 1.0$ *(red). The variability model, where only the continuum varies (*$\alpha = 0$*) is shown as the dashed black line, providing the best match to the observations. Reverberation mapping time-lags, (*$\tau \sim 0.3 - 1$ *year* $\times$ $(1 + z)$ *in quasars, e.g., Kaspi et al. 2007) are not taken into account.*

model leads to the same conclusions and predicts the redshift behavior in Figure 3.4 equally well for s_{ui}.

Improved modeling, including a broken power-law continuum, explicit treatment of line reverberation and lack of variability at higher rest wavelengths as shown in Vanden Berk et al. (2004) and Wilhite et al. (2005), would be fruitful to carry out (see Section 3.8.1 for details), but is beyond the scope of the present work.

3.5.3 COLOR VARIABILITY IN *ui*

The SDSS Stripe 82 data offer the opportunity to extend this analysis beyond the relatively short spectral range covered by g and r, 4770 Å to 6231 Å in the observed frame. We do so by exploring the color variability in the u vs. i magnitude-magnitude space,

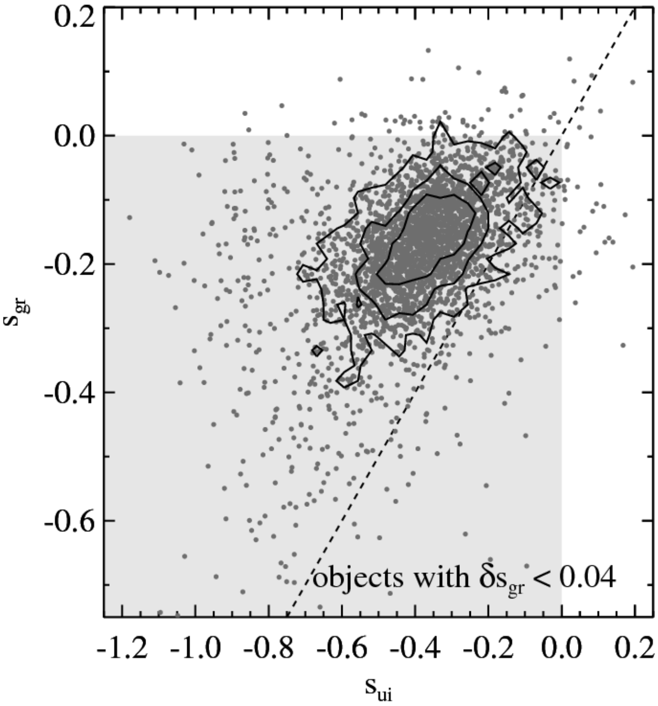

FIGURE 3.7: *The color variability in gr vs. ui color variability space, after correcting for the line-induced redshift dependence (as described in Section 3.5.1). The gray shaded region shows where* $s_{\lambda_1\lambda_2} < 0$, *i.e., where bluer means brighter. The dashed line indicates* $s_{gr} = s_{ui}$ *for reference. The ui color variability is more pronounced on average than the gr color variability. Only the 3111 objects of the sample with* $\delta s_{gr} < 0.04$ *are shown.*

which covers a spectral range from 3543 Å to 7625 Å. We chose to use the *i*-band instead of *z* because of the significantly smaller photometric uncertainties in the *i*-band. The fitting procedure was exactly analogous to the case of *gr* color variability as described in Section 3.4. The right panel of Figure 3.4 shows the estimated *ui* color variability for the Stripe 82 quasar sample. Despite the larger scatter in s_{ui} at any given redshift we see similar features such as a distinct redshift dependence in $\langle s_{ui}\rangle(z)$ superimposed on quite dramatic overall color variability of -0.46. It is clear that the *ui* color variability is more pronounced than the *gr* color variability; $\langle s_{ui}\rangle(z) < \langle s_{gr}\rangle(z)$. This holds true for the ensemble properties as well as for individual objects, as illustrated in Figure 3.7, where we plot the emission line corrected (as described in Section 3.5.1) color variability in *gr* and *ui*. The fact that $s_{ui} < s_{gr}$ for almost all objects, implies that there is a relatively stronger blueing over the *ui* spectral range than over the smaller *gr* range. The same conclusion is reached when accounting for the difference in the wavelength baselines between *gr* and *ui* by normalizing s_{gr} and s_{ui} with the corresponding wavelength ratio.

3.6 Discussion

We now proceed to put the color variability into the context of other physical parameters that describe the quasar phase and the time dependence of variability, exploring to which physical processes color variability may be linked.

3.6.1 Color Variability as a Function of Eddington Luminosity and Black Hole Mass

All 9093 spectroscopically confirmed quasars have matches in the quasar catalog presented in Shen et al. (2011), of which 99.9% (9088) have an estimate of the bolometric luminosity and 84.1% (7615) have an estimated black hole mass derived from Mg II (see Shen et al. 2011 for details). This allows us to normalize the bolometric luminosity to the Eddington luminosity ($L_{\rm Edd}$; Equation (2.4)) to obtain the Eddington ratio.

If we plot the emission line corrected (as described in Section 3.5.1) *gr* color variability against $L_{\rm bol}/L_{\rm Edd}$ and $M_{\rm BH}$, as shown in Figure 3.8, it is evident that there is no detectable relation between the color variability s_{gr} and the $L_{\rm bol}/L_{\rm Edd}$ or $M_{\rm BH}$. This is also illustrated in Figure 3.9 where a 2D histogram of $L_{\rm bol}/L_{\rm Edd}$ and $M_{\rm BH}$, with the bins color coded according to the median s_{gr}, is shown: across the well sampled range in $L_{\rm bol}/L_{\rm Edd}$ and $M_{\rm BH}$, the median s_{gr} varies by no more than 0.01 as a function of these two variables about the mean value of -0.17. The 2D histogram has been smoothed by a 2D gaussian to reflect the uncertainty in luminosity and mass, with the full width at half maximum of the smoothing kernel (represented by the ellipse in the bottom left of Figure 3.9) FWHM $= [\mathrm{FWHM}(M_{\rm BH}), \mathrm{FWHM}(L_{\rm bol}/L_{\rm Edd})] = [0.35, 0.24]$ corresponding to $[\sigma(M_{\rm BH}), \sigma(L_{\rm bol}/L_{\rm Edd})] = [0.15\ \mathrm{dex}, 0.1\ \mathrm{dex}]$. Plots similar to the ones shown in Figure 3.8 and 3.9 for the *ui* color variability show no significant $L_{\rm bol}/L_{\rm Edd}$ or $M_{\rm BH}$ dependence, either. In Figure 3.8 the full sample, i.e., all masses and redshifts are shown. Inspecting smaller sub-samples in z (and $M_{\rm BH}$) space does not change the picture. Hence, we find no correlation between the color variability in *gr* (and *ui*) with $L_{\rm bol}/L_{\rm Edd}$ or $M_{\rm BH}$. More broadly, this seems to imply that the overall state of the quasar (characterized by $L_{\rm bol}/L_{\rm Edd}$ and $M_{\rm BH}$) plays no significant role in determining the color variability.

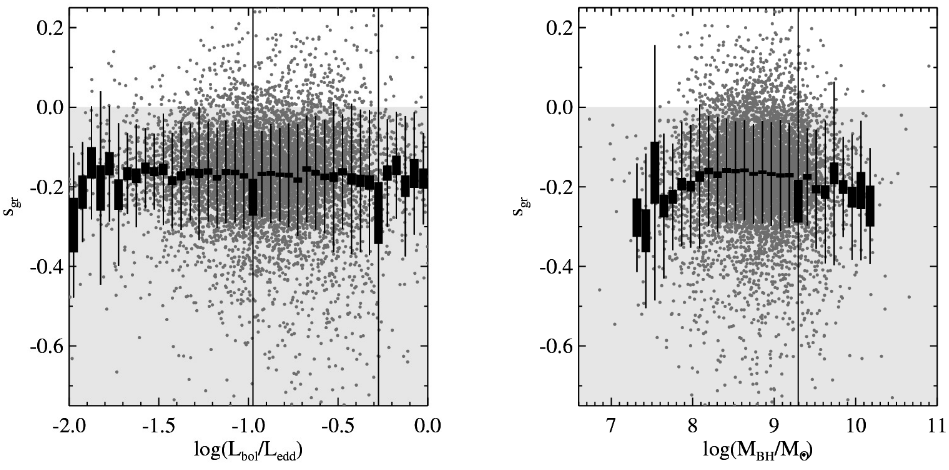

FIGURE 3.8: *The emission line corrected gr color variability as a function of L_{bol}/L_{Edd} (left) and $M_{BH}/M_{\odot}$ (right). The black rectangles indicate the mean s_{gr} and its error in each L and M_{BH} bin. The thin error bars show the width of the s_{gr} distribution similar to in Figure 3.4. No correlation between the color variability and L_{bol}/L_{Edd} and $M_{BH}/M_{\odot}$ is detected. This implies that the color variability is not a function of black hole mass or the overall (Eddington) accretion rate (Equation (2.5)).*

3.6.2 THE COLOR VARIABILITY AS A FUNCTION OF THE LIGHT CURVE VARIABILITY CHARACTERISTICS

In Chapter 2 (Schmidt et al. 2010) we characterized the r-band variability of all the 9093 quasars through a structure function with an amplitude parameter A and the light curve stochasticity, γ. The structure function variability of each individual quasar was modeled by a simple power-law (Equation (2.8))

$$\mathrm{SF_{mod}}(\Delta t_{\mathrm{obs}}|A,\gamma) = A\left(\frac{\Delta t_{\mathrm{obs}}}{1\mathrm{yr}}\right)^{\gamma} , \tag{3.9}$$

with Δt_{obs} being the time between the observation of two individual photometric epochs and $\mathrm{SF_{mod}} = \sqrt{\langle (m(t_1) - m(t_2))^2 \rangle}$. The structure function of a periodically varying object or one varying like white noise will as mentioned have a flat structure function and hence a small power-law exponent γ (see Figure 2.5). Thus a large γ indicates a secularly varying object or an object with a random walk like variability. The latter has been shown to describe quasar variability well in Kelly et al. (2009) and MacLeod et al. (2011). The amplitude A corresponds to the average variability on a 1 year timescale. In Chapter 2 (Schmidt et al. 2010) all calculations were done in the observed frame to mimic quasar identification with no prior information such as redshift. However, the spectroscopic redshift of each of the 9093 Stripe 82 quasars is known, and the amplitude can be corrected for time-dilation. The rest-frame variability amplitude A' is defined to be $A(1+z)^{\gamma}$ such

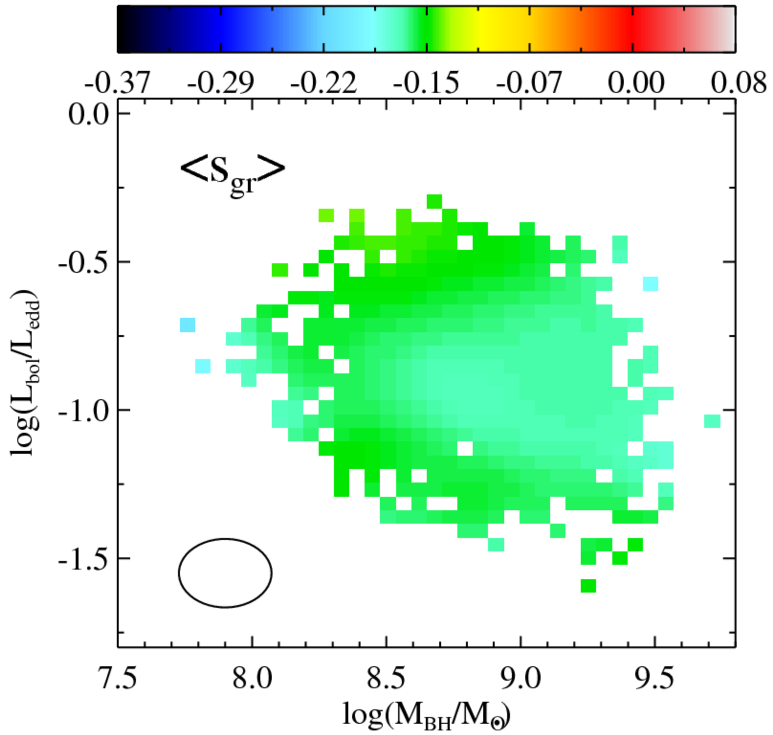

FIGURE 3.9: *The color variability s_{gr} as a function of L_{bol}/L_{Edd} and $M_{BH}/M_{\odot}$, as in Figure 3.8, again showing no significant trends. The color coding indicates the emission line corrected median gr color variability of the objects in the bin. To reflect the uncertainty in luminosity and mass, the distribution has been smoothed by a 2D gaussian (represented by the ellipse in the bottom left corner) with full width at half maximum of* $FWHM = [FWHM(M_{BH}), FWHM(L_{bol}/L_{Edd})] = [0.35, 0.24]$ *corresponding to* $[\sigma(M_{BH}), \sigma(L_{bol}/L_{Edd})] = [0.15\ dex, 0.1\ dex]$.

that

$$\mathrm{SF}_{\mathrm{mod}}(\Delta t_{\mathrm{rest}}|A', \gamma) = A' \left(\frac{\Delta t_{\mathrm{rest}}}{1\mathrm{yr}}\right)^{\gamma} , \tag{3.10}$$

where Δt_{rest} is now the difference between observations in the quasar rest-frame. All quoted A's are estimated from the robust r-band measurements as described in Chapter 2 (Schmidt et al. 2010). The variability amplitudes are independent of redshift in agreement with Giveon et al. (1999) and the majority of the previous studies listed in their Table 1.

In the following, however, the structure function parameters A and γ have been obtained somewhat differently from in Chapter 2 (Schmidt et al. 2010). Rather than fitting the structure function directly to the magnitude differences we fit a Gaussian Process model (Rasmussen & Williams 2006) defined by the structure function to the magnitudes directly. This properly includes all of the correlations between data points. This Gaussian Process model consists of an n-dimensional Gaussian distribution (for n epochs) with a constant mean m and $n \times n$ variance matrix V. The elements of this variance matrix are given by

$$V_{ij} \equiv V(|t_i - t_j|) = V(\Delta t_{ij}) = \frac{1}{2}\left[\mathrm{SF}^2_{ij}(\infty) - \mathrm{SF}^2_{ij}(\Delta t_{ij})\right] \tag{3.11}$$

for data points at epochs t_i and t_j. Here the structure function SF_{ij} is given by

$$\mathrm{SF}_{ij} = \sqrt{\langle (m(t_i) - m(t_j))^2 \rangle} \quad . \tag{3.12}$$

The photometric uncertainty variances are added to the diagonal elements of V. For the power-law structure function we cut off the power-law at 10 years such that $\mathrm{SF}(\infty)$ is finite. As all data span less than 10 years this cut-off does not influence the fit. This type of fit is similar to the Ornstein-Uhlenbeck process describing quasar variability as a damped random walk (e.g., Kozłowski et al. 2010; Butler & Bloom 2011; MacLeod et al. 2011).

We can now look at the emission line corrected s_{gr} and s_{ui} as a function of A' and γ for all quasars. Figure 3.10 shows that s_{gr} and s_{ui} seemingly vary both with A' and with γ. However, the limit of little variability (small A') requires particular care, both because outliers play a bigger role and because A' and γ starts to be degenerate. We estimated the gr color variability of 500 color-selected non-varying F/G stars (see Chapter 2 for further details) and of the 483 Stripe 82 RR Lyrae stars from Sesar et al. (2010). These are over-plotted in the top panel of Figure 3.10 as the blue and red points, respectively. As expected the RR Lyrae have a well defined color variability, whereas the inferred color variability of the non-varying F/G stars span a much wider range of s_{gr}. Interestingly, the majority of the non-varying F/G stars have color variability estimates of $s_{gr} < 0$ like the quasars and the RR Lyrae stars. This seems to be caused by the outliers in g being relatively larger than the outliers in r, hence affecting the initial guess of the MCMC in a bluer-brighter direction. In the case of the RR Lyrae the well defined mean color variability in gr is expected, as RR Lyrae change their effective temperature and luminosity during their pulsation. By creating a sequence of black body spectra with temperatures from 6200K to 7200K, estimating the flux received in the g and r bands for each spectrum, and using that as a simple model for a variable RR Lyrae star, a color variability of $s_{gr} \sim -0.23$ is obtained, in very good agreement with the observations (Figure 3.10, top panel, red dots). Thus, in general the s_{gr} for the F/G and RR Lyrae stars look as expected.

A more direct way to estimate the fidelity of color variability estimates at low A' values is to recover s_{gr} estimates for objects of known (simulated) color variability. We induce such simulated variability into the 500 F/G stars by generating data of a certain 1 year amplitude (A') from the original F/G star g and r photometry. We do that by generating new u, g, r and i magnitudes for the individual epochs j via the expression

$$p_{j,\mathrm{sim}} = p_j + A'_{\mathrm{sim}} \frac{\Delta \mathrm{MJD}_j}{365.25} \quad , \tag{3.13}$$

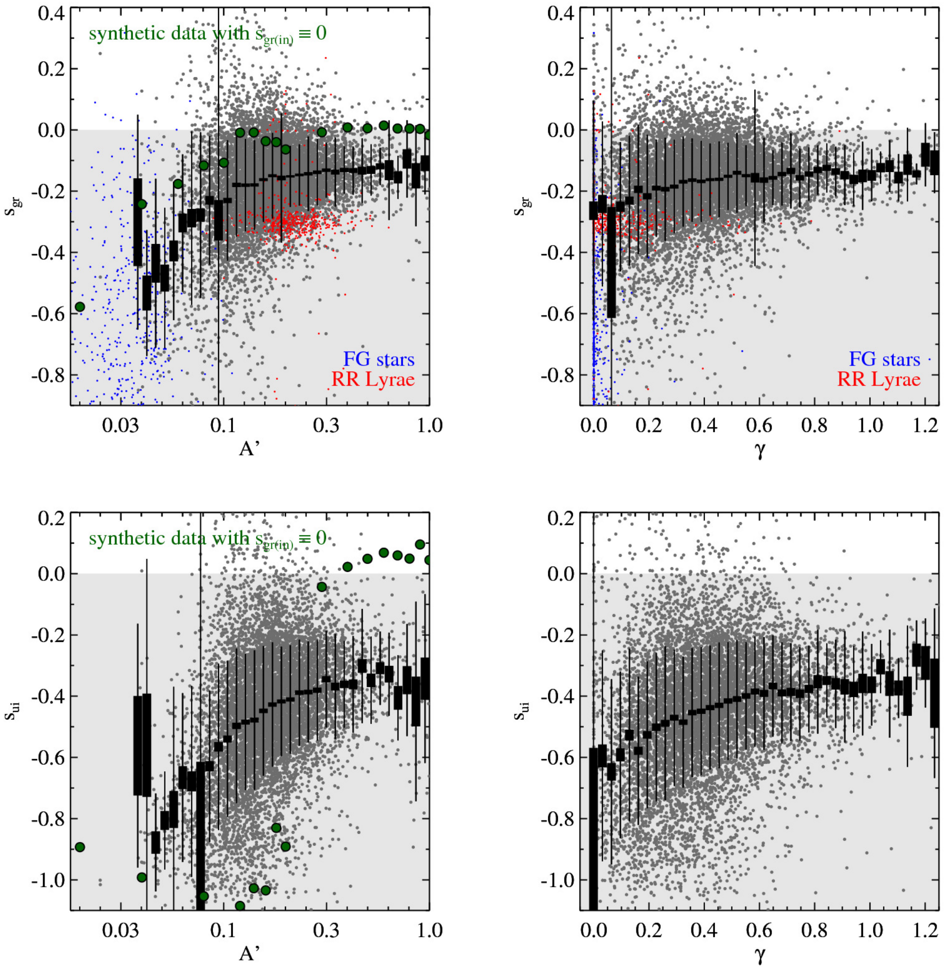

FIGURE 3.10: *Correlation between the light curve (temporal variability) structure function and color variability. The emission line corrected color variability in gr (top) and ui (bottom) are plotted against A' (left) and γ (right). A' indicates the mean level of variability within one year (rest-frame), and γ, the power law exponent in the structure function, indicates how random (low γ) or secular (high γ) the light curve variations are. Gray dots (data), black rectangles and error bars are analogous to Figure 3.8. In the gr plots (top) 500 non-varying F/G stars and the 483 RR Lyrae from Sesar et al. (2010) are shown on top of the 9093 Stripe 82 quasars (gray dots) as blue and red points respectively. In the left column the recovered average s_{gr} (s_{ui}) for 50 F/G stars is shown as green filled circles, where synthetic brightness variations with $s_{gr}(in) \equiv 0$ and different variability amplitudes A'_{sim} have been created. As described in the text, this illustrates that only gr (ui) trends for $A' \gtrsim 0.1$ ($A' \gtrsim 0.25$) can and should be trusted. It shows that objects with large A', i.e., with large variability amplitudes, have a color variability close(r) to 0, i.e., less blueing when brightening, than do objects with small A'. The trends in the two right hand plots are dominated by low A' objects and is therefore not trustworthy. Figure 3.11 shows a cleaned version of his figure.*

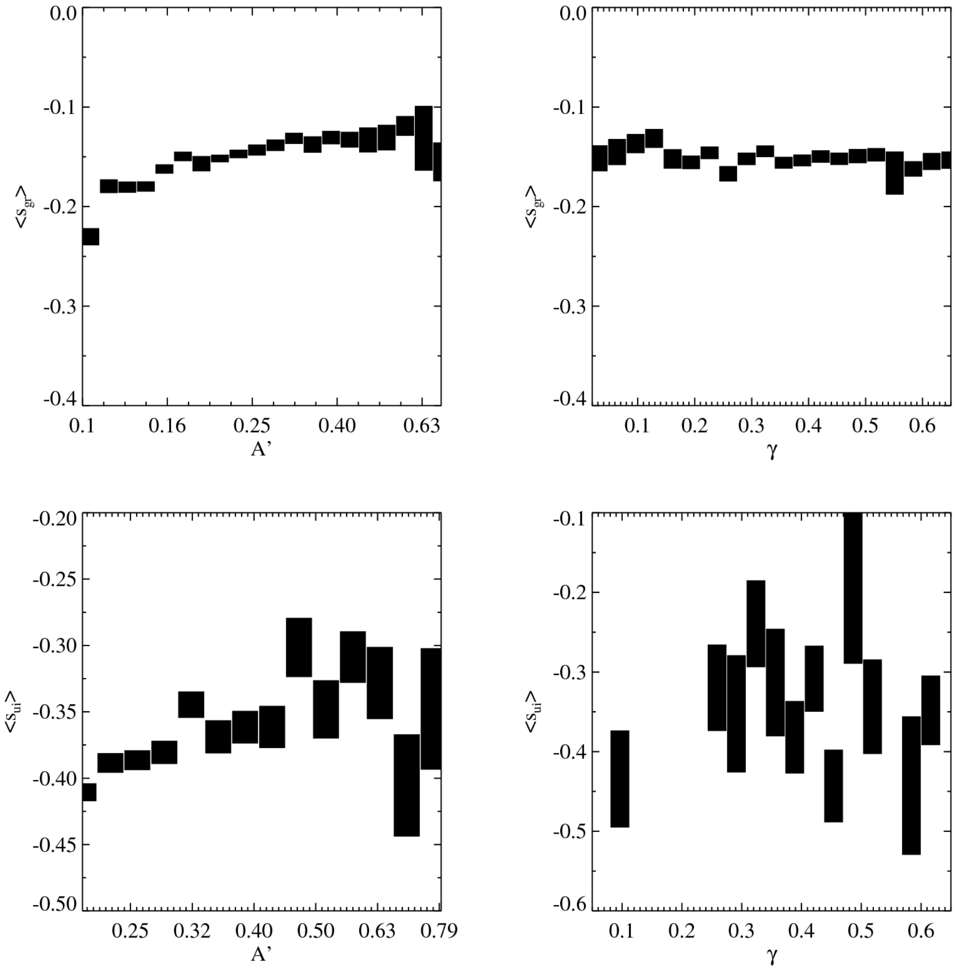

FIGURE 3.11: *The estimated correlations between mean color variability and structure function parameters. The four panels show the portion of parameter space in Figure 3.10 that is well populated by the sample. The rectangles in the left hand plots correspond to the ones in Figure 3.10. In the right hand plots the rectangles are only estimated from objects with* $A' > 0.1$ (0.25). *A clear trend is seen, that the color variability of objects with high variability amplitude are closer to 0 than objects with small variability amplitude (left panel). This indicates that strongly varying quasars get less blue as they brighten on average, than do moderate varying quasars. The color variability is independent of the power law index* γ.

which by construction is a data set with $s = 0$. Here p represents the photometric measurements in a given band, j runs over the individual epochs, and $\Delta\mathrm{MJD}_j$ refers to the observation time of the jth epoch with respect to the first observation. In this way many of the aspects of the real data (i.e., the outliers and realistic photometric errors) are included in the simulated data. In the two left panels of Figure 3.10 this recovered mean average (over 50 randomly chosen F/G stars) color variability is shown for a sequence of variability amplitudes, $A'_{\rm sim}$, as large filled green circles. This shows that the recovered gr (ui) color variability has some systematic errors below $A' \lesssim 0.1$ ($A' \lesssim 0.25$). When ignoring quasars with variability amplitudes smaller than 0.1 (0.25) the correlation between s_{gr} and s_{ui} and A' is still present and significant. The trustworthy part of the relations, $\langle s_{gr}\rangle(A, \gamma)$ and $\langle s_{ui}\rangle(A, \gamma)$, in Figure 3.10 is shown in Figure 3.11. Again the black rectangles represent the uncertainty on the estimated mean color variability. The left hand plots correspond directly to Figure 3.10, whereas for $\langle s_{gr}\rangle(\gamma)$ and $\langle s_{ui}\rangle(\gamma)$ the black rectangles are estimated only from objects with $A' > 0.1$ ($A' > 0.25$). Figure 3.11 clearly illustrates the trend that objects with larger variability amplitude have a smaller color variability (meaning less blueing when brightening) than for low A'. On the other hand the color variability is independent of γ.

As mentioned, Vanden Berk et al. (2004) and Wilhite et al. (2005) showed that there is a lack of variability at high rest wavelengths. Furthermore, it is known that quasar variability is anti-correlated with luminosity (e.g., Hook et al. 1994; Cristiani et al. 1996; Vanden Berk et al. 2004). This might lead to the suspicion that the trend presented in Figure 3.10 and 3.11 is nothing more than a redshift effect. If this was the case the relation should be due mainly to low-z objects, since the most variable quasars are supposedly low luminosity quasars, i.e., necessarily only observed at low z, and should therefore disappear at high redshift. Estimating the relation between s_{gr} and A' in various redshift bins (also split in luminosity) shows that the trend is equally strong for all redshifts and all luminosities. Hence, the presented relation appears to be of a physical origin and not merely a redshift effect.

3.6.3 COLOR VARIABILITY AND CHANGES IN THE MEAN ACCRETION RATE

In this section we carry out a cursory exploration as to the physical origin of the observed color variability.

3.6.3.1 COLOR VARIABILITY OF INDIVIDUAL QUASARS VS. THE COLOR DISTRIBUTION OF QUASAR ENSEMBLES

The colors of quasars at a given redshift are known to depend only weakly on their mean accretion luminosity or accretion rate (Davis et al. 2007), while we find that individual quasars become considerably bluer when they brighten on year time-scales. This suggests different physical mechanisms creating the accretion luminosity range in ensembles and the luminosity variations in individual quasars.

In Figure 3.12 the emission line corrected color variability of a sub-sample of the Stripe

82 quasars is shown in *gr* and *ui*-space. This sub-sample represents the 'average' quasars, i.e., the combined sample of the 33rd–66th percentile of masses and the 25th–75th percentile of redshifts for the quasar sample. The color variability of each individual quasar is depicted as a short solid gray line showing s'_{gr} (s'_{ui}) from Equation (3.3) for each quasar centered on $[\langle g\rangle, \langle r\rangle]$ ($[\langle u\rangle, \langle i\rangle]$) for that particular quasar. Only every 10th object of the sub-sample is actually shown to keep the individual gray lines visible. The length of the lines resemble the change in the photometric g-band (u-band) data of the quasar. The Figure compares the average s'_{gr} (s'_{ui}) of all the individual quasars in the sub-sample (red solid line) with a fit to the time-averaged color distribution of the sub-sample (black solid line) where each data point corresponds to $[\langle g\rangle, \langle r\rangle]_k$ ($[\langle u\rangle, \langle i\rangle]_k$) with k counting the quasars. Figure 3.12 reveals that indeed the mean color variability for individual quasars is much more pronounced than the equivalent quantity for the ensemble, $s_{\mathrm{sample}} = \frac{d\langle m_r\rangle}{d\langle m_g\rangle}$. For the given sub-sample $s_{gr} \sim -0.18$ and $s_{ui} \sim -0.48$ on average (as opposed to $s_{gr} \sim -0.17$ and $s_{ui} \sim -0.46$ for the full sample) compared to $s_{gr} = -0.01$ and $s_{ui} \sim -0.08$ for the corresponding time-averaged sub-sample color distribution. The difference is highly significant in both cases, with the *ui* color variability difference formally larger, because of the broader spectral range. The exact same trends are found for plots containing the full quasar sample.

This result shows that (temporal) color variability of individual quasars is considerably stronger than the color range of ensembles of quasars at similar redshifts and with similar black hole masses, that presumably differ in $L_{\mathrm{bol}}/L_{\mathrm{Edd}}$.

3.6.3.2 COLOR VARIABILITY VS. ACCRETION DISK MODELS

Explaining quasar spectral energy distributions, and in particular the UV/optical continuum through steady-state accretion disk models has an established history (e.g., Shakura & Sunyaev 1973; Bonning et al. 2007; Davis et al. 2007). However, comparing the observed color variability of large samples of quasars with the predicted colors of model sequences of varying accretion rate has not been done yet. Such a comparison could tell us whether it is sensible to think of the quasar variability on scales of years as changes in the mean accretion rate. The superb Stripe 82 data enables us to perform such a comparison, by comparing the observed color and color variability of the Stripe 82 quasars with sequences of accretion disk models presented in Davis et al. (2007).

Davis et al. (2007) presented three different thin accretion disk models that describe the spectral slope of quasars as a function of $L_{\mathrm{bol}}/L_{\mathrm{Edd}}$ and M_{BH}. We took these three models and worked out predictions for the observed g and r band for models of a given M_{BH} but varying accretion rates. The three models presented in Davis et al. (2007) and the color we adopt for their graphical representation are:

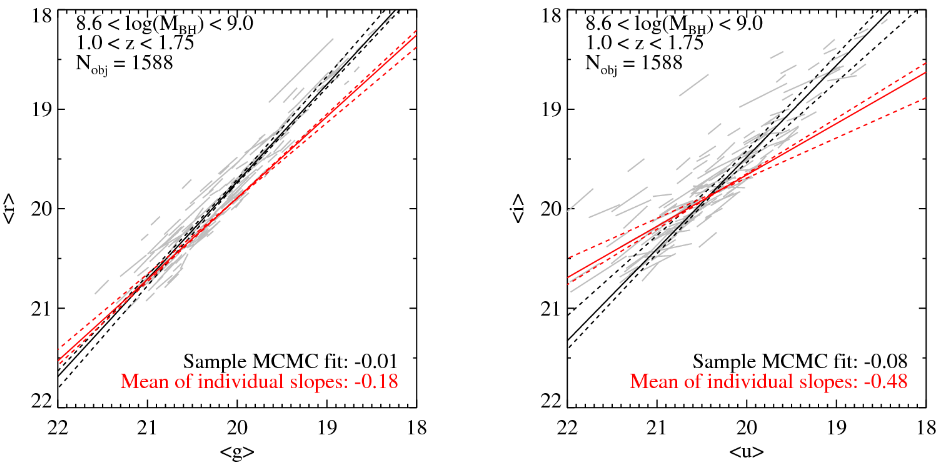

Figure 3.12: *The color variability in individual quasars (red) vs. the color-luminosity relation in the ensemble of quasars (black), drawing on the sub-sample of 33rd–66th percentile of masses and the 25th–75th percentile of redshifts. The mass and redshift ranges are shown in each panel. Each individual quasar is indicated by a short solid gray line corresponding to the fitted and emission line corrected gr (ui) relation, with a length reflecting the standard deviation of the g (u) light curve. Only every 10th object of the sub-sample is shown to keep the individual objects visible. The red solid lines show the average trend within this sub-sample of* $s_{gr} \sim -0.18$ *and* $s_{ui} \sim -0.48$, *with the dashed red lines indicating the bootstrapping uncertainty. The black lines show a MCMC fit to the time-averaged fluxes for the sample, i.e., the fit to* $(\langle g \rangle, \langle r \rangle) \pm (stdev(g), stdev(r))$ *for each quasar, corresponding to* $s_{gr} \sim -0.01$ *and* $s_{ui} \sim -0.08$. *The dashed black lines show the 68% confidence interval of the MCMC fit. In other words, the red line is the ensemble mean color variability while the black line is a fit to the ensemble of time-averaged mean magnitudes. Hence, it is clear that the average color variability of the individual quasars deviate significantly from the time averaged sample color variability.*

1) A relativistic model of accretion onto a Schwarzschild black hole with a spin parameter of 0. The emission is based on Non-LTE atmosphere calculations (green).
2) A relativistic model of accretion onto a Schwarzschild black hole with a spin parameter of 0. The disk is emitting as a black body (red).
3) A Model of accretion onto a spinning black hole (spin parameter of 0.9) with emission based on Non-LTE atmosphere calculations (orange).

For further details on the models we refer to Davis et al. (2007).

We can compare the models to the data in two respects: (i) do they predict the right color (which has been done before) and (ii) do they predict the right change of color with changing accretion rate or luminosity? In Figure 3.13 the quasar from Figure 3.3 is shown in g-r-$(g-r)$ space (without error ellipses) together with its best fit color variability

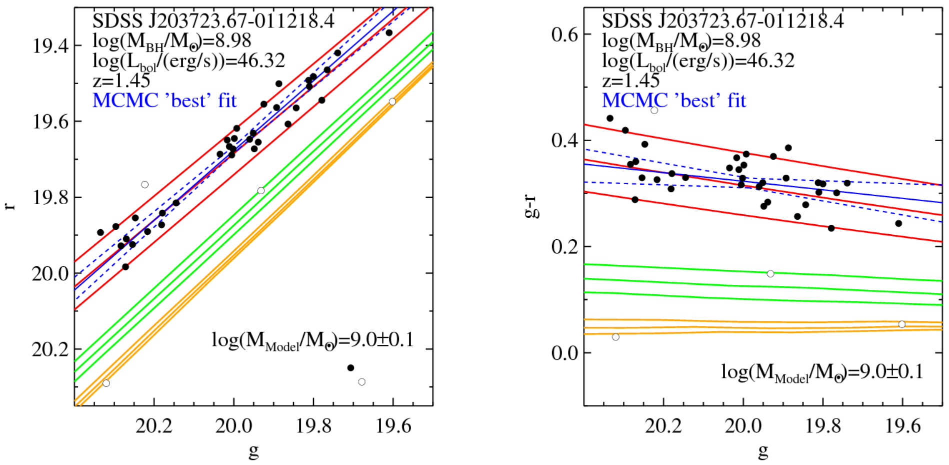

FIGURE 3.13: *Comparison of the observed color variability with sequences of steady-state accretion disk models (Davis et al. 2007) of different accretion rates. This comparison is illustrated using the quasar SDSS J2037-0112 from Figure 3.3, shown in r vs. g (left) and $g - r$ vs. g (right) space. The black symbols are the individual photometric measurements. The blue solid line shows the MCMC fit in gr-space to these measurements and the blue dashed lines indicate the 68% confidence interval of that fit. The open symbols denote likely outliers (see Figure 3.3). The Davis et al. (2007) models described in Section 3.6.3.2 are shown in green (model 1), red (model 2) and orange (model 3) with L_{bol}/L_{Edd} changing along the lines. Each set of 3 lines denotes models for differing black hole masses near the value determined from the MgII line width (Shen et al. 2011). From top to bottom each set of lines (models) correspond to 9.1, 9.0 and 8.9* $\log(M_{model}/M_{\odot})$. *For this particular object the match to model 2 is good; see Figure 3.14 for an ensemble comparison.*

(blue solid line). The three accretion disk models are shown as solid lines in bundles of three, where each of the three lines corresponds to a different black hole mass, as noted in the bottom right corner of each panel. In this particular case model 2) matches the data well both in color and in the change of color with changing luminosity. However, such a good match is not representative for the ensemble. We quantify this for the whole sample by estimating the 'goodness' of the models as:

$$\mathbf{D}_y = \left(\sum_j^{N_k} \frac{1}{\delta y_j} \right)^{-1} \sum_j^{N_k} \frac{y_j - y(x)_{j,\mathrm{model}}}{\delta y_j} \qquad (3.14)$$

$$\Delta s = s_k - s_{k,\mathrm{model}} \ , \qquad (3.15)$$

where $y = (g - r)$ and $x = g$. The index j runs over the N_k epochs for each individual quasar k. The model prediction is the color at a given luminosity (or accretion rate), for a fixed black hole mass $y(x)_{j,\mathrm{model}}$. The photometric error on the color for the jth measurement is denoted by δy_j. Since the error on the M_{BH} estimates based on MgII (Shen et al. 2011) is ~ 0.4 dex (Vestergaard & Peterson 2006; Rosa et al. 2011), we have chosen to show three values for M_{BH}, leading to three model prediction lines, for each model in Figure 3.13.

The 'goodness' parameters $\mathbf{D}_y$ and Δs defined in Equations (3.14) and (3.15) therefore describe how well the observed values of color and color variability are predicted by the Davis et al. (2007) models. $\mathbf{D}_y$ can be seen as the standard χ^2 measure of comparison between model and data *before* squaring, i.e., it estimates the difference between the model color and the observed color averaged over all epochs for each quasar. The Δs is simply the difference between model color variability and observed color variability of each quasar.

Figure 3.14 summarizes the model–data comparison for a subset of quasars with $z \sim 1.5$: the quantities from Equations (3.14) and (3.15). All models predict a color variability – as a function of changes in the mean accretion rate – that is weaker than observed. On average model 2) shown in red matches the observed *gr* color variability the best. Furthermore, the $\mathbf{D}_{g-r}$ values indicate that the $g - r$ color is on average overestimated by model 1) and 3), whereas the distribution of model 2) has a mean very close to the dashed perfect agreement line. Creating similar plots for other mass and redshift ranges as well as for the results in *ui*-space show the exact same trends. Thus, of the three Davis et al. (2007) accretion disk models considered here, model 2) matches the observed color and the obtained *gr* and *ui* color variability the best.

3.7 CONCLUSION

In this chapter we determined and analyzed the color variability of 9093 spectroscopically confirmed quasars from SDSS Stripe 82, to understand to which extent and why quasars get bluer (redder) if they brighten (dim), by fitting linear relations between the SDSS *g* and *r* bands as well as between the *u* and *i* bands in magnitude-magnitude space. The connection of various quasar properties to the color variability were inspected before the results were compared to models of accretion disks with varying accretion rates from Davis et al. (2007). Our main results can be summarized as follows:

- We showed that quasar color variability, $s_{\lambda_1\lambda_2} \equiv \frac{\partial m_{\lambda_2}}{\partial m_{\lambda_1}} - 1$, is best determined by fitting data in the statistically independent magnitude-magnitude space, rather than in color-magnitude space as many studies have done. Unless care is taken to account for the data correlations, the latter approach may lead to spurious or biased estimates of color variability. The *gr* and *ui* color variability for the vast majority of the 9093 quasars s_{gr} and s_{ui} are both < 0, confirming that quasars get bluer when they get brighter.

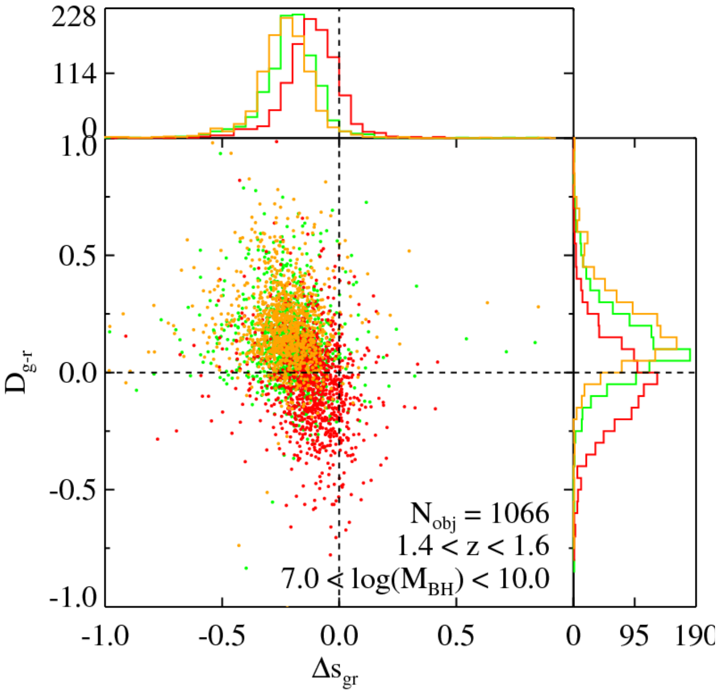

FIGURE 3.14: *Comparison of the observed color variability with steady-state accretion disk models with changing L_{bol}/L_{Edd}. Each individual point reflects the data vs. model discrepancy in the mean color $\boldsymbol{D}_{g\text{-}r}$ (Equation (3.14)) and Δs_{gr} (Equation (3.15)) for one individual quasar in the parameter space illustrated in Figure 3.13. The plot shows the sub-sample of 1066 quasars with* $1.4 < z < 1.6$. *The color of the points correspond to model 1) (green), 2) (red) and 3) (orange) from Davis et al. (2007) converted into g and r magnitudes. The dashed lines show perfect agreement between model and data in color and color variability, respectively. The projections of the distributions are shown as histograms along the axes. While the different sets of models can reproduce (by design) the mean quasar colors (i.e., $\boldsymbol{D}_{g\text{-}r}$), the observed color variability is far too strong to be interpreted as changes in accretion rate in a steady-state model context.*

- The color variability as measured in *gr* and *ui* space exhibits a distinct redshift dependence, which we could clearly attribute to the effect of emission lines exiting/entering the photometric SDSS bands. From a set of simple models of spectral quasar color variability, a model in which the line and continuum vary in phase but with the line amplitude fixed to 10% or less of that of the continuum, is able to reproduce the observed redshift trends in the *gr* (and *ui*) color variability as well as the observed amount of color variability.

- The fact that we clearly see the impact of the emission line fluxes on the broad band photometry through the redshift-dependence of the color variability implies that broad-band reverberation mapping should be possible with the data set at hand.

- Correcting for the emission lines leaves us with a sample mean (continuum) color variability of $s_{gr} \equiv \frac{\partial m_r}{\partial m_g} - 1 = -0.17$ and, analogously, $s_{ui} = -0.46$.

- We found that the emission line corrected color variability is independent of $L_{\rm bol}/L_{\rm Edd}$ and $M_{\rm BH}$ in both gr and ui: there is no correlation between the mass and luminosity of quasars and their color variability.
- The color variability, however, does depend on the light curve variability properties (described by a power-law structure function as in Chapter 2 and Schmidt et al. 2010). We found that quasars with large variability amplitudes (A') tend to have less color variability, as compared to quasars with small variability amplitudes.
- We found that the characteristic color variability on timescales of years of the individual quasars is larger than the dependence of the typical quasar colors on their overall accretion state (i.e., $L_{\rm bol}/L_{\rm Edd}$). This implies that changes in the overall accretion rate cannot explain the observed color variability. Ephemeral hot spots may however be a plausible explanation for the observed color variability. This picture is confirmed by our comparison of the observed color variability to sequences of steady-state accretion disk models by Davis et al. (2007) with varying accretion rates, which also exhibit much less color variability as a function of accretion rate.

Our analysis provides a clear indication that on time-scales of years, quasar variability does not reflect changes in the mean accretion rate. Some other mechanism must be at work; presumably some disk instability. What mechanism match the existing data, certainly warrants further modeling. The current study can also be viewed as an initial foray into the realm of multi-band, multi-epoch panoptic photometry that the Pan-STARRS and LSST surveys can bring to full fruition.

3.8 OUTLOOK

As illustrated above, the intrinsic color variability of quasars is an under-exploited observable of quasars and AGN that enable in-depth characterization of the physics that govern these systems. Some of the arguments laid out in this chapter are preliminary, and hence would benefit from further investigation and in-depth studies. We will describe a couple of these in this short outlook.

3.8.1 MODELING OF SPECTRAL QUASAR VARIABILITY

In Section 3.5.2 we created a simple variability model (outlined in detail in Appendix C on page 149) to reproduce the observed color variability of the SDSS quasars (see Figure 3.6). The model assumed that the quasar spectrum can be model by a simple power-law in the rest-frame optical, and that the emission lines all vary by the same amount with zero time-lag, τ, with respect to the continuum variations. Even though this simple model successfully reproduced the general trend of the observed color variability, the assumptions make it only a toy model pointing towards future, more realistic modeling.

There are three obvious improvements that needs to be included in any future attempts to model the color variability. First of all the continuum emission should be modeled as a broken power-law instead of a single power-law, as this is generally a better approximation of the non-thermal AGN continuum as illustrated in the composite quasar spectrum in Figure 3.1 and described in Vanden Berk et al. (2001); Krolik (1999) and Section 3.2. Secondly, the variability reverberation, or echo, of the emission lines to the continuum variability is non-zero and more complex, than simply just scaling the total emission line flux all at once, as was used for the model presented here. The individual emission lines have been shown to have very different time-lags to the observed continuum variation (e.g., Peterson et al. 2004; Zu et al. 2011), and the response of the different lines is very different. While some lines respond by a large fractional increase in their flux, others respond more moderately or not at all to the continuum variations (Wilhite et al. 2005). As described in Section 3.2.5, the observed time-lags can be seen as a direct consequence of the size of the structure, hence, both of these observational facts might simply be a consequence of the different emission being produced in different regions of the AGN. Nevertheless, this needs to be included in any modeling. Last but not least, it has been shown that the variability at redder rest wavelengths tends to disappear (Wilhite et al. 2005; Vanden Berk et al. 2004). With a broken power-law prescription one could imagine to limit the allowed variability on the red part of the power-law, and adjust the redder part of the emission line flux, such that it reacts to the more modest continuum variation.

The exact details of these improvements to the toy model presented in Section 3.5.2 still needs to be worked out, but with a more physical model that takes these crucial steps into account, we will be able to get a clear view of the interplay between the individual emission lines and the continuum variation, as described by the observed color variability of the ~9000 quasars in SDSS Stripe 82.

3.8.2 PHOTOMETRIC REVERBERATION MAPPING

In this chapter we also indicated that the observed color variability on time-scales of months to years is not caused by simple changes in the mean steady state accretion rate. We showed that the emission line fluxes (their delayed/sub-linear responses) leave a signature in the photometric data. Correcting for these emission line effects leaves a strong and nearly constant color variability both seen as a function of $L_{\rm bol}/L_{\rm Edd}$ and

M_{BH}. Through statistical photometric precision broad band reverberation mapping these emission line effects can possible be disentangled from the continuum variation, and determine the black hole masses and sizes of the BLR of thousands of quasars. The improved modeling of how the actual quasar spectra vary mentioned above, held up against the observational results will help in this disentanglement. Developing the field of photometric reverberation mapping (already initiated in, e.g., Chelouche & Daniel 2012; Haas et al. 2011) could first be done as a feasibility study on the SDSS Stripe 82 data used here, and later be applied to the Pan-STARRS 1 and LSST data sets, enabling black hole mass estimates of *classes* of AGN as opposed to individual objects. Obtaining *directly* measured black hole masses of thousands of objects across a wide range of redshifts, will help better understand the proposed scaling relations between M_{BH} and, e.g., luminosity, velocity dispersion and bulge masses (see Section 1.5). Especially at high redshift these relations are essentially extrapolations, and detecting any evolution in them will be a key diagnostic of the growth and fueling of AGN and their host galaxies.

A statistical (and photometric) approach to reverberation mapping of AGN will serve as a compliment to traditional monitoring campaigns, and direct comparison can be performed by considering predefined regions in the space spanned by redshift, luminosity and black hole mass.

3.9 – EPILOGUE –

In this chapter (Schmidt et al. 2012a) we have shown how the *color* variability can be used as a probe of the physics governing the central engine of quasars (and AGN). We confirm that the optical quasar spectrum is hardened, i.e., becomes bluer, when the quasar increases its brightness. This is in agreement with several previous studies Giveon et al. (1999); Trèvese et al. (2001); Trèvese & Vagnetti (2002); Geha et al. (2003); Vanden Berk et al. (2004); Wilhite et al. (2005); Sakata et al. (2011) though here we eliminate any doubt about the strength of the results, caused by co-variances between the magnitudes and colors, as was not done in the majority of these studies.

We also show that the Davis et al. (2007) steady-state accretion disk models have problems reproducing the relatively strong variations in the quasar colors observed in the SDSS Stripe 82 data. This agrees well with our main result of this chapter, namely that the strength of the observed color variability in the ~9000 quasars in Stripe 82 does not agree with being solely caused by changes in the mean steady state accretion rate of the assumed central engine accretion; such a scenario seems on average to produce too small color changes. Hence, by fairly simple means, we have used the color variability of the quasars to give concise predictions of what physics can and cannot be responsible for the observed signal.

If the change in steady state accretion rate is only responsible for part of the observed variability, where does the rest of the observed signal then stem from? What we are looking for is a process which disappears, or weakens, when time-averaging the sample and remains strong when looking at individual objects. We speculate that a stochastic process in the accretion disk could be responsible for the observed color variability. For instance a scenario where flares or ephemeral hot spots were to appear in the accretion disk could create the observed strong trends. However, this needs to be investigated further.

Another very interesting direction to take the results presented in this chapter, is photometric reverberation mapping of tens of thousands of quasars. As described in Section 3.8.2, if the contribution from the emission lines and the continuum to the observed color variability can be disentangled, the outstanding data of Stripe 82 (and in the future Pan-STARRS 1 and LSST) offers the possibility of obtaining *direct* measurements of the black hole masses in thousands of systems all at once. This opens a new window to the understanding of the size and mechanisms that govern the central engine of *ensembles* of quasars and AGN.

CHAPTER 4

STAR FORMATION IN HIGH REDSHIFT GALAXY MERGERS

In this chapter we will present and analyze the spatial distribution of star formation in an extensive sample of galaxy mergers at $z > 1$. Our sample, drawn from the 3D-HST survey, is the first merger sample to be systematically selected at these redshifts and consists of 61 objects, including pre-mergers (close pairs) and post-mergers (single objects with multiple components and/or tidal features). Total stellar masses and star formation rates are derived from multi-wavelength photometry. The slitless spectra from 3D-HST provide emission line maps, which are projected back onto the undispersed continuum images, providing a comprehensive, high-resolution, empirical picture of where star formation happens in galaxy mergers at the epoch where the cosmic star formation rate density peaked. We find that detectable star formation can occur in one (or both) galaxy centers, or in tidal tails. The most common case (69%) is that star formation is largely concentrated in a single, compact region, coincident with a peak in rest-frame optical continuum emission. This suggests that $z \sim 1.5$ mergers typically occur between galaxies with different gas fractions. No correlations between star formation morphology and redshift, total stellar mass, or star formation rate is found. An initial comparison with a restricted set of cosmological hydrodynamical simulations of mergers between similarly massive and gas-rich objects, supports this by predicting that star formation usually occurs in both components of the merger, when the gas fraction of the individual components is the same.

This chapter is based on Schmidt et al. (2012b).

4.1 – PROLOGUE –

In the previous two chapters we have been focusing on AGN and quasars, the extraordinary active phase in the life and evolution of galaxies. As mentioned both in Section 1.5 and Section 2.2 one of the causes for AGN and quasars is thought to be the merging of galaxies. However, the exact mechanisms that trigger AGN have been quite extensively debated in the literature (e.g., Sanders et al. 1988; Canalizo & Stockton 2001; Dunlop et al. 2003; Georgakakis et al. 2009), and recently several authors have been arguing against clear evidence that merging is a *necessity* for igniting AGN (Cisternas et al. 2011; Schawinski et al. 2011; Kocevski et al. 2012). Irrespective of this debate it is clear that many AGN and quasars reside in interacting or dynamically perturbed systems. Hence, understanding how mergers behave and evolve might shed light on the processes that can ignite AGN.

Furthermore, as described in Chapter 1, both merging of galaxies and the formation of stars are among the most important mechanisms in galaxy formation and evolution in general. The results of billions of years of star formation and galaxy evolution, i.e., the spatial distribution of present-day stars in galaxies, is well known. How it looked at redshift 1–3, the main epoch of galaxy formation where most quasars are observed and the galaxy merger rate peaked, is still fairly unknown. Probing this epoch is crucial to get a better understanding of how the galaxies evolved from the initial density fluctuations presented in Chapter 1 into the galaxies we observe today. One way to do this is by mapping the *spatial* extent and distribution of star formation in high redshift ($z > 1$) galaxy mergers.

Gaining knowledge about the spatial distribution of star formation in mergers would help answering where gravitational interaction triggers the formation of stars. Is it mostly triggered in the galaxies' outer regions, in pronounced tidal tails, or are the stars centrally concentrated in the merging components with no clear burst from the interaction? Closely related to this question is the uncertainty about in what phase of the merging star formation happens. Does the majority of the star formation triggering happen at first passage, where the two galaxies are almost undisturbed, or at final coalescence where the two galaxies merger to one heavily disturbed system? It is not hard to imagine that both of these issues are governed by the mass ratio of the merger, the gas fraction of the individual components, and at what angle we are observing the galaxies at, i.e., the relative inclination of the merging components and the line of sight to the merger. Simulations have done a great job in predicting observables of exactly these quantities (e.g., Mihos & Hernquist 1996; Springel 2000; Cox et al. 2008; Hopkins et al. 2010), but models of star formation in mergers have not been tested against observations at high redshifts yet; another reason why such studies are needed.

However, there are several complicating factors making high redshift star formation studies very challenging in practice. First of all, the emission line features of galaxy spectra probing star formation, i.e., Hα (Kennicutt 1983; Gallagher et al. 1984; Kennicutt et al. 1994; Kennicutt 1998a,b), [OII] (Gallagher et al. 1989; Kennicutt 1998b; Kewley et al. 2004), and [OIII] (Kennicutt 1992; Teplitz et al. 2000; Hippelein et al. 2003), move into the near-infrared for redshifts above ~1, making collecting these data observationally much more challenging than in the UV and optical, where star formation in the low-redshift Universe has been extensively studied. Secondly, simple photometric observations are not (always) enough to get a clear view of spatial extent of star formation in galaxies. People have been using photometry in various pass-bands to probe the extent of the light from star forming regions (Pogge & Eskridge 1987; Mirabel et al. 1998; Gavazzi et al. 2002), but a more direct way of probing star formation is via the spectral signatures mentioned above. To probe the spatial extent of star formation three-dimensional spectroscopy is needed via for example slitless or integral field unit spectroscopy. The SINS sample (Förster Schreiber et al. 2009) described in Section 4.2.1 is the largest existing sample of galaxies at the main epoch of galaxy formation with the spatially resolved near-infrared three-dimensional spectral information needed to probe the extent of star formation at these redshifts.

But this is unfortunately not enough if the goal is to study the interplay between star formation and merging of galaxies. Yet another complicating factor is to obtain and define a sample of mergers at high redshift. Mergers were not a systematically selected part of the SINS sample, so only ~10 of the SINS galaxies have perturbed morphologies and are characterized as merging systems, making the SINS survey unsuitable for such a study. Hence, no proper statistical samples of galaxy mergers with near-infrared three-dimensional spectroscopy exist today.

However, these challenges have recently been overcome with the 3D-HST survey and the slitless spectroscopy capabilities of the *Wide Field Camera 3* on the *Hubble Space Telescope*. In this chapter we will use the 3D-HST survey (described in Section 4.3) to define the largest sample of morphologically disturbed galaxies at the main epoch of galaxy formation with near-infrared three-dimensional spectral information, enabling us to start answering some of the crucial questions outlined above, regarding the build-up of galaxies just 3–4 Gyr after the Big Bang.

> *Galaxy mergers and star formation are the two main reasons galaxies look the way they do – and now we can study their interplay at the main epoch of galaxy formation.*

4.2 STAR FORMATION & GALAXY MERGERS

As described in Chapter 1 two of the key elements of galaxy evolution are the linear and non-linaer merging of dark matter halos and their accompanying baryons, and the monolithic collapse and condensation of baryonic gas into stars in these dark matter potential wells. Hence, getting a handle on both mergers and star formation (regions) at different redshifts can help clarify and describe the various evolutionary stages of a galaxy's life. In this section we will describe the processes leading to the formation of stars in the bottom of dark matter halo potential wells, and how the interaction and merging between galaxies can further induce and enhance this star formation. We do this by partially following Mo et al. (2010), hence, for a detailed and extensive in-depth treatment of the mechanisms involved in star formation mentioned here, please refer to their book and the references therein.

Having isolated (for now) a dark matter halo, the gas eventually making up the observable galaxy, and hence our main source of information, will start condensing while sinking towards the bottom of the dark matter potential well as first described in (White & Rees 1978; Efstathiou & Silk 1983). In it simplest form a condensation is the process of becoming more dense and to cool. From a chemical point of view a condensation reaction is the formation of molecules from smaller reactants. This is exactly what is thought to happen when gas condensate to form galaxies as well as stars.

The internal energy of the gas reservoirs in galaxies decrease mainly by radiative cooling. The gas is under gravitational influence of the dark matter halo and to some extent its own self-gravitation and therefore contracts and increases its density as it cools. The gas mainly looses its energy via bremsstrahlung, where electrons are slowed down by interaction with atomic nuclei in the gas and hence looses energy and cools. When the gas has cooled and contracted significantly for actual interactions to happen cooling via collisional ionization (atoms ionized by colliding electrons), recombination (an electron is captured by an ionized atom), and collisional excitation (and atom is excited from a non-ground state and afterwards decays to the ground state) also sets in. Complimentary to the radiative cooling, Compton scattering, where electrons and photons exchange energy, also contribute.

Allowing the gas to cool radiatively of course assumes that the gas can stream freely from the cloud, i.e., it is assumed that the galaxy is not opaque at the emitted wavelengths. The cooling, and thereby the contraction of the gas, is counteracted by the produced radiation that, instead of streaming freely, re-reacts with the gas resulting in photoionization heating. Hence, the cooling of the gas clouds in the galaxy is a balance between self gravity of the gas particles, the cooling radiation and the radiation pressure. These are the same mechanism that were important in the structure formation model described in

Section 1.3, and the Jeans formalism therefore also applies in this case. Thus, what determines whether the gas cloud continues to collapse or comes to a hold is the size of the initial perturbation of the particular region in the gas cloud, expressed via the Jeans length, λ_J, from Equation (1.3) on page 6. The mass within a sphere of diameter λ_J is called the Jeans mass, M_J. If the amount of gas within the Jeans sphere is larger than the Jeans mass, i.e., if $M_{\text{gas}} > M_J$, the considered gas parcel is not stopped by the counteracting pressure and will continue to collapse.

Such a basically free-falling contraction of the gas is referred to as the 'cold mode' of gas accretion. This is in contrast to the 'hot mode' of gas accretion, where it is not only the radiation pressure of the gas that counteracts the collapse. In the hot mode, the gas is heated to the virial temperature by shocks happening during the condensation. Cold mode accretion usually happens in less massive dark matter halos whereas hot mode accretion predominantly happens in massive halos.

Eventually, whether the gas is condensing via cold or hot mode accretion, the gas will reach an equilibrium where the temperature, i.e., the kinetic energy, balances the cooling mechanisms. Thus, a thermal equilibrium between the density of the gas, ρ_{gas}, and the gas temperature, T_{gas}, is reached. At this stage instabilities can be obtained if the temperature or the density of the equilibrated system is perturbed, either by thermal instabilities, i.e., by further internal cooling or heating, or by hydrodynamical instabilities as for instance turbulence where parcels of gas starts moving around within the condensed gas cloud.

All of the above applies under the assumption that only the energy of the radiating gas contributes to the balances and equilibria. However, if stars have already formed, stellar winds or stellar explosions add new terms to the equation. Furthermore, if an AGN has been ignited at the center of the galaxy (Section 2.2) this radiation will also counteract further collapse and needs to be included, which complicates the matter even further.

If all the radiative processes are nevertheless overcome by the gravitational contraction and the radiative and collisional cooling, the most dense regions of the gas cloud will eventually become so dense that they can start transforming[1] into so-called giant molecular clouds. Giant molecular clouds mostly consist of H_2, but depending on the enrichment of the gas, i.e., how many times the gas has been recycled via stellar winds and explosions, it might also contain, e.g., CO, H_2O, and HCN. If the recombination of hydrogen to form H_2 is more efficient than the photodissociation of H_2, these clouds become stable. It is generally accepted that the overall star formation rate of a galaxy is determined by its ability to form these dense molecular clouds.

---

[1] The main cause for this transformation, whether it being thermal instability, gravitational instability or over-densities induced by galaxy interactions and/or mergers, is still unclear.

Assuming that the galaxy efficiently forms giant molecular clouds, the next challenge to overcome, is to create an over-density of gas and molecules within this cloud, so dense that it starts burning hydrogen in its core, i.e., so dense that a star is born. The efficiency of forming stars from a giant molecular cloud has been estimated to be $\epsilon_{\rm SFR} \sim 0.002$, i.e., only of the order 0.2% of the available molecular gas is turned into stars (see Mo et al. 2010). This last step of turning the densest regions of the molecular cloud into stars is a complicated matter and has not yet been fully understood. The reason for the very low star formation efficiency in molecular clouds is believed to be due to a combination of magnetic fields, winds from already formed stars dispersing the surrounding molecular material to prevent further star formation, and/or (maybe most importantly) turbulence in the molecular cloud itself. Assuming that all of the above conditions are in place and satisfied new stars are born.

Thus, it is clear, that without large enough gas reservoirs to form giant molecular clouds, stars will not ignite. Hence, it is not surprising that one of the more well-established relations, describing the star formation in galaxies, show a tight correlation between the gas surface density, $\Sigma_{\rm gas}$, and the star formation rate per surface area, $\Sigma_{\rm SFR}$. This relation is called the Schmidt-Kennicut law (Schmidt 1959; Kennicutt 1998a) and can be expressed as

$$\Sigma_{\rm SFR} = (2.5 \pm 0.7) \times 10^{-4} \left(\frac{\Sigma_{\rm gas}}{M_\odot \ {\rm pc}^{-2}} \right)^{1.4 \pm 0.15} \frac{M_\odot}{\rm yr \ kpc^2} \quad . \tag{4.1}$$

This empirical relation was obtained assuming that the Hα emission traces star formation, which we will also assume later in this chapter. As the Schmidt-Kennicut law relates the global star formation rate to gas surface density of galaxies it cannot tell us much about the actual star formation in individual clouds within galaxies, but only provide information about the formation and evolution of galaxies as a whole. It has actually been shown that within galaxies a clear relation between $\Sigma_{\rm gas}$ and $\Sigma_{\rm SFR}$ is lacking. More localized observational empirical relations between star formation rate and atomic as well as molecular gas have been established (see, e.g., Bigiel et al. (2008), but also Calzetti et al. (2012) and references herein), enabling a more detailed study of the star formation itself within single galaxies, rather than the global star formation in populations of galaxies Equation (4.1) provides.

Apart from the self-regulated star formation described above, another way to obtain the dense gas and giant molecular clouds needed to ignite stars is by compressing the material via external forces. An example of such an event is external interaction from another galaxy and its halo, or in the more extreme case galaxy merging. The gravitational interaction of close encounters can produce tidal tails as it was first shown in Toomre & Toomre (1972) and induce density waves in the interacting systems. Such an enhancement in the density and kinetic energy of the gas and molecular clouds can trigger bursts of star formation by kick-starting the chain of processes outlined above. Not only do the galaxy interactions and mergers produce density waves and tidal tails, they also enhance the angular momentum loss of the gas. This is an important step in funneling gas into the central regions of galaxies as described in Section 2.2. If the loss is big enough, this can feed a possible AGN engine. More modest angular momentum loss can lead to enhanced density and temperature in the inner regions and trigger nuclear star bursts. Hence, the interaction and merging of galaxies can enhance and induce star formation and is therefore an important player in the evolution of galaxies. This has been confirmed both observationally and via simulations as described in the following.

4.2.1 Observing the Star Formation in Mergers

The results of ~10 billion years of galaxy-scale star formation and galaxy evolution are well known thanks to countless observational efforts, i.e., we know the spatial distribution of stars in the population of present-day galaxies. This has for example led to the powerful correlations between star formation and the gas in galaxies mentioned above. Also the particular cases of star formation in interacting galaxy systems has been studied in the low-redshift ($z < 1$) Universe (e.g., Joseph & Wright 1985; Lambas et al. 2003; Hammer et al. 2005; Barton et al. 2000, 2007; Lin et al. 2007; Bridge et al. 2007; Jogee et al. 2009; Robaina et al. 2009, 2010). From these studies it has become clear that both self-regulated star formation and merging must be key ingredients in galaxy formation and evolution.

Even though galaxy mergers are observed to enhance star formation in galaxies and trigger some of the most violent starbursts known (e.g., Rieke et al. 1985; Joseph & Wright 1985; Melnick & Mirabel 1990; Klaas & Elsaesser 1991; Wang et al. 2004), it appears that the net effect of major mergers in the global star formation history of the galaxy population has been relatively modest since at least a redshift of 1 (e.g., Robaina et al. 2009; Jogee et al. 2009). Whether or not it played a more important role at larger redshift is a present very uncertain. Hence, a crucial step towards fully understanding how galaxies have evolved is to study the star formation properties in merging systems at high redshifts, i.e., at $z > 1$. Some of the questions that needs to be observationally addressed are: Where did stars form in merging galaxies at higher redshift? Which phase(s) of the merging process seems to trigger most star formation?

The fact that the interplay between merging and star formation in the general evolution of galaxies has been investigated in much less detail at higher redshift is mainly due to the observational challenges. For example, tracing the star formation via Hα at $z > 1$, as Kennicutt (1998a) did it when establishing Equation (4.1) in the local Universe, requires observations in the near-infrared, and these have been far less practical than observations in optical wavebands. Nevertheless, the epoch from $1 < z < 3$ is immensely important for understanding the formation and evolution of galaxies, since this is the cosmic time when the majority of the stars we see in galaxies today were formed (e.g., Hopkins & Beacom 2006; Karim et al. 2011). Only a few studies have tried to address the impact of (major) galaxy mergers on the amount and geometry of star formation at $z \sim 1.5$ (e.g., Swinbank et al. 2004; Law et al. 2007; Förster Schreiber et al. 2009; Wright et al. 2009).

As mentioned above, to get an understanding of the star formation processes it is crucial to be able to investigate the resolved star formation within individual galaxies, instead of the global star formation. Hence, it is essential to obtain knowledge about the *spatial* extent of star formation. This is most efficiently done via three-dimensional (3D) spectroscopy like integral field unit (IFU) data or slitless spectroscopy, where both the spectral (to identify emission lines) and the spatial information is available. Until recently, large samples of galaxies, and in particular galaxy mergers, with spatial 3D spectroscopic information at high redshift did not exist. The most prominent current sample of galaxies with 3D spectroscopy at $1 < z < 3$ is as mentioned the SINS sample of Förster Schreiber et al. (2009). They have by now collected near-infrared SINFONI IFU spectroscopy and imaging of ~100 star forming galaxies on ESO's *Very Large Telescope*. Using Hα as a star formation tracer, this data enabled them to spatially analyze the gas kinematics (Shapiro et al. 2008), morphologies and physical properties of these systems. Mergers were not a systematically selected part of their sample, so only a small fraction of the SINS galaxies have significantly perturbed morphologies and are characterized as merging systems. Thus, studying the spatial extent of star formation in mergers at high redshift in a systematic manner with existing data has so far been extremely difficult, not to say impossible. The study presented in this chapter will take the first steps in changing that.

4.2.2 MODELING THE STAR FORMATION IN MERGERS

Along with the extensive observational programs, that have addressed the interplay between star formation and galaxy mergers below redshift 1, theoretical simulations of mergers with dust prescriptions enabling the study of star formation have been carried out (e.g., Mihos & Hernquist 1996; Barnes & Hernquist 1996; Springel 2000; Cox 2004; Cox et al. 2006, 2008; Matteo et al. 2007). The strength of such models is that they can be followed in 'real-time' as galaxies merge and evolve over cosmic time, as opposed to observations which resemble snapshots of billions of years of evolution.

These simulations have predicted and confirmed several of the 'modes' and 'morphologies' of star formation observed, such as nuclear star bursts, enhanced star formation in tidal features resulting from the gravitational interaction of the merger components, and regions of star formation triggered by the direct interaction as the galaxies collide.

Hence, simulations are now advanced enough that they can efficiently reproduce the star formation observed at $z < 1$. However, due to the lack of observations at higher redshifts, it has not been possible to address whether the simulations in the present form are able to reproduce the observed star formation in mergers here. However, as illustrated below and in particular in Section 4.7, such comparisons have recently become possible too.

In Section 4.7 we will use the N-body and smoothed-particle hydrodynamics (SPH) cosmological merger simulations of T. J. Cox and Patrik Jonsson (Cox et al. 2006, 2008; Jonsson 2006; Jonsson et al. 2010) as representatives for the status quo of merger simulations with proper dust prescriptions and hence star formation. These simulations are based on the GADGET code (Springel et al. 2001) with the addition of a proper dust prescription from the radiative transfer code SUNRISE (Jonsson 2006).

The simulations calculate the radiative cooling of the gas outlined above, via a simple function of the local gas density, ρ_{gas}, and the internal energy per unit mass, and ensures that the empirical Schmidt-Kennicut law from Equation (4.1) is satisfied. One of the largest uncertainties and challenges in modern simulations of star formation is the treatment of supernova feedback, i.e., how much radiation is fed back into the system as stars explode. In the simulations used here, the supernovae feedback is fed back on a 8 Myr timescale from an available energy reservoir, resulting in an effective change in temperature and pressure (density) of the surrounding gas. Before this can take effect the actual stars need to be formed. As the resolution of the simulations only goes down to particles of $\geq 10^5 M_\odot$, a stochastic prescription is used to fragment these particles into collisionless stars, i.e., if the over-dense regions satisfy the criteria for star formation, they are turned into stars if a randomly drawn number satisfy some preset star formation efficiency ϵ_{SFR}.

The 'raw' GADGET mergers only trace the massive particles and show how the actual merging happens, but do not resemble actual observations which traces light and not mass. Realistic observations are obtained via SUNRISE. SUNRISE traces photons as they interact with the gas particles and molecules, either via scattering, collisions, absorption and/or re-emission, until they eventually leave the enshrouding cloud of gas and dust. Generating enough such photons, matched to the spectral energy distribution of the source calculated from stellar population synthesis models, and tracing them through the medium, eventually paints a picture of the system in the direction of the artificial observer (see Jonsson 2006 for details). Hence, by collecting these photons an observed spectrum of the object is obtained. In practice this can be done in each individual pixel of the collecting area, and therefore provides a three-dimensional data-cube of monochromatic images of the GADGET merger as observed from the ray-traced photons escaping the system. It is three such data-cubes, with two spatial and one spectroscopic dimension, we will use in Section 4.7.

For further information on these simulations and their dust prescription please refer to Cox (2004); Jonsson (2006); Cox et al. (2006, 2008) and Jonsson et al. (2010).

4.2.3 THIS STUDY: EXPLOITING NEW DATA

The goal of this chapter is to use the new unprecedented data from the 3D-HST survey, to address the spatial extent and triggering of star formation for a statistical sample of galaxy mergers at the peak of cosmic star formation. The 3D-HST survey, which we will present in more detail in Section 4.3, provides large samples of objects with near-infrared slitless grism spectroscopy making such a study possible for the first time.

The power of the 3D-HST data, i.e., the ability to obtain resolved line emission, enabling studies of the spatial extent of star formation for large samples of galaxies at $z > 1$ in different cosmological fields, has already been illustrated in a series of papers. van Dokkum et al. (2011) construct and study a spectroscopic stellar mass-limited sample of 34 galaxies at $1 < z < 1.5$ to quantify the properties of massive galaxies via Hα equivalent widths. Nelson et al. (2012) use the 3D-HST data to determine the spatial extent of star formation in high-z galaxies with Hα equivalent widths > 100 Å and compare this with the near-infrared continuum light distribution, and Fumagalli et al. (2012) estimate the Hα equivalent width of an extensive sample of high redshift systems.

Using the same 3D-HST data we will explore the spatial extent and distribution of star formation in a 'population snapshot' of *merging* systems at $z \sim 1.5$. It is the first time an extensive and more representative statistical sample of morphologically selected mergers at $z \sim 1.5$ with spatially resolved line emission, i.e., spatially resolved star formation, is available and can be exploited to understand the star formation properties of mergers at these redshifts in detail. For this sample we make $\sim 0.2''$ resolution maps of emission lines (Hα and [OIII]), which delineate the spatial extent of the (unobscured) star formation in these mergers and allow us to study their star formation properties in a statistical and unbiased way. This is done under the assumption that the Hα (for $z \sim 0.7 - 1.5$) and [OIII] (for $z \sim 1.2 - 2.3$) emission of the systems trace the star formation. This has been shown to be a fair assumption indeed, for both Hα (Kennicutt 1983; Gallagher et al. 1984; Kennicutt et al. 1994; Kennicutt 1998a,b) and [OIII] (Kennicutt 1992; Teplitz et al. 2000; Hippelein et al. 2003), even though using [OIII] as a quantitative indicator of star formation rate (which is not what we will do here) includes several complicating factors.

Furthermore, we will create an initial sample of pseudo-observations from the simulated data-cubes described in Section 4.2.2, which we can compare to the 3D-HST data. As mentioned, simulations predict a merger sequence and star formation picture with both centrally concentrated triggered star-bursts at final coalescence, enhanced star formation in tidal features, and black hole growth and accretion. It is predictions like these we hope to confirm or reject at $z > 1$ by making direct comparisons with the observational results from the 3D-HST survey. Furthermore, such comparisons will help exploring which parameters, e.g., viewing angle, merger phase, gas fraction, mass ratio, etc. play a crucial role in determining, for example, the distribution of star formation from observations

In Section 4.3 we describe the 3D-HST survey from which the merger sample was selected. We then describe the actual selection of our sample in Section 4.4, before the procedure used to map the spatial extent of the star formation is described in Section 4.5. In Section 4.6 we present and discuss the results of the 3D-HST data. In Section 4.7 we describe how the simulated grism spectra were created and compare them to the 3D-HST results, before we summarize and conclude in Section 4.8. As in the previous science chapters we end this chapter in Section 4.9 with a short outlook presenting some of the improvements and possibilities this study leads to.

4.3 THE 3D-HST SPECTROSCOPIC SURVEY DATA

To construct the 0.2″ resolution emission line maps which we will use as a proxy for star formation maps, we take advantage of the near-infrared 3D spectroscopy survey possibilities that the *Wide Field Camera 3* (WFC3) on *The Hubble Space Telescope* (HST) has brought. The 3D-HST survey[2] (Brammer et al. 2012) is a 248 orbit near-infrared spectroscopic Hubble treasury program (Cycles 18 and 19, PI van Dokkum). It provides near-infrared imaging with the F140W filter and grism spectroscopy with the G141 grism over well-studied extragalactic survey fields (AEGIS, COSMOS, GOODS-S, GOODS-N, and UKIDSS/UDS). The grism spectroscopy is slitless so both spatial and spectroscopic information is available for every single object in the survey fields. The 3D-HST survey provides rest-frame optical spectra for a sample of ~7000 galaxies at $1 < z < 3.5$ (van Dokkum et al. 2011; Brammer et al. 2012).

As of August 2011 the survey had gathered 68 pointings over the GOODS-S, GOODS-N, COSMOS and AEGIS fields (van Dokkum et al. 2011). As the work presented here is limited to pointings with extensive ancillary data, the present study is based on 30 of these 68 pointings. The 3D-HST data was reduced and analyzed using the slitless spectroscopy data extraction software aXe (Kümmel et al. 2009). The reduction of the grism spectroscopy include a detection of sources in the F140W image using the Source-Extractor (Bertin & Arnouts 1996), for which the respective grism spectra can be identified and extracted.

---

[2] `http://3dhst.research.yale.edu/`

4.3.1 THE 3D-HST GRISM SPECTROSCOPY

The key element of the 3D-HST survey for constructing the emission line maps is the slitless grism spectroscopy component, where light from the full field-of-view is dispersed onto the detector via a grism. The WFC3 G141 grism used in 3D-HST disperses the light over the wavelength range from 1.05 μm to 1.7 μm with a relatively low spectral resolution of $R \sim 130$. In parallel 3D-HST has also obtained optical grism spectroscopy with the *Advanced Camera for Surveys* (ACS), which we will not use for the current study.

Since the grism spectroscopy is slitless, the WFC3 G141 grism basically produces an emission line image that is superimposed onto a sequence of dispersed monochromatic continuum images, and some of the key features of slitless spectroscopy therefore need to be considered and taken into account. First of all, the width of emission/absorption lines in the dispersion direction in slitless spectroscopy is not only caused by velocity broadening (which is negligible for the low resolution 3D-HST spectra) and the intrinsic broadening of the wavelength dispersion: as slitless spectroscopy simply produces shifted monochromatic images, the spatial extent of the dispersed emission line image reflects the spatial distribution of the line emission both along and perpendicular to the dispersion direction. It is exactly this feature we will take advantage of when mapping the spatial extent of star formation as described in Section 4.5. Figure 5 in van Dokkum et al. (2011) illustrates this 'morphology-broadening' of the emission lines.

Another issue with slitless spectroscopy is the wavelength coverage of the spectra. Since the detector onto which the field-of-view is dispersed has a finite size, ~10% of the spectra will fall on the edge of the detector and is therefore cut off. This is quantified by the value `FCOVER`, which simply gives the fraction of the G141 wavelength range that actually falls on the detector when dispersing each object.

Lastly, 'contamination' is an important property of slitless spectroscopy. Since the focal plane is not blocked out with a slit or a mask as is usually done in standard spectroscopy, all the light of a given object, and all other objects in the observed field, are dispersed onto the detector. Hence, spectra will often overlap and therefore 'contaminate' each other as explained in Brammer et al. (2012). In 3D-HST catalogs this contamination is quantified by the value `FCONTAM`, which is defined as the ratio of the contamination from the aXe contamination model and the flux measured in the 1D spectra in the G141 range. The aXe software estimates this contamination via modeling of the expected grism spectra from the F140W source extraction. The contamination model of the grism spectra is not perfect. For example light from objects outside the field-of-view will be dispersed into the detector but will not appear in the sextractor catalog of the field which the contamination model is based on, hence `FCONTAM` in principal runs from 0 to ∞, but for approximately correct models give the fraction of contaminating flux (from 0 to 1) in the 1D spectrum.

For more information on the data reduction methods, the data products of the 3D-HST survey and the survey itself please refer to Brammer et al. (2012).

4.4 SELECTING MERGERS

We select our sample of mergers from the pointings in the first (internal) data release of the 3D-HST survey based on three different inputs: (i) The 3D-HST survey catalog, (ii) spectral energy distribution (SED) modeling, and most importantly (iii) visual inspection of near-infrared (F140W) morphologies. The two former selection steps are performed to minimize the number of objects to visually inspect. We will describe each of these three steps in the following.

To obtain robust star formation rate estimates via the SED fitting described in Section 4.4.2, we require extensive ancillary photometric catalogs. We therefore focus on the 30 pointings of data available in GOODS-S (6) and COSMOS (24), where the photometric data in the FIREWORKS (Wuyts et al. 2008) and the NEWFIRM medium band survey (NMBS; Whitaker et al. 2011) catalogs are available, respectively. Hence, this work is performed on approximately 1/4 of the final 120 pointings 3D-HST data product.

4.4.1 GRISM CATALOG CUTS

The first step in defining our merger sample is to minimize the number of objects to visually inspect by selecting a well-defined sample of objects based on the data products of the 3D-HST survey. We ensure that the spectra of the sample cover the majority of the G141 wavelength range by demanding `FCOVER` > 0.75. Since we are looking for merging objects we do not put any constraints on `FCONTAM`, as spectra of close pairs will always have a high level of contamination. We will rely on the visual inspection (Section 4.4.3) to remove objects with heavy contamination from objects, which are not part of the potentially merging system.

Each individual object in the 3D-HST catalog has been matched to the available ancillary photometric catalogs. Each 3D-HST detection is assigned a match radius given by

$$R_{\mathrm{match}} = \sqrt{\Delta \mathrm{R.A.}^2 + \Delta \mathrm{Dec}^2}\ , \qquad (4.2)$$

TABLE 4.1: *Grism (Top) and SED (Bottom) Catalog Cuts*

0.75	$\leq$	`FCOVER`		
0.0	$\leq$	`FCONTAM`		
		$R_{\rm match}$	$\leq$	$0.3''$
		$m_{\rm F140W}$	$\leq$	23.5
0.7	$\leq$	$z_{\rm grism}$	$\leq$	2.3

9.0	$\leq$	$\log\left(\frac{M_*}{[M_\odot]}\right)$	$\leq$	12.0
-9.5	$\leq$	$\log\left(\frac{sSFR}{[\mathrm{yr}^{-1}]}\right)$		
1.0	$\leq$	$\log\left(\frac{SFR}{[M_\odot/\mathrm{yr}]}\right)$		

where

$$\Delta\mathrm{R.A.} = \mathrm{R.A.}_{\rm grism} - \mathrm{R.A.}_{\rm phot} \tag{4.3}$$

$$\Delta\mathrm{Dec} = \mathrm{Dec}_{\rm grism} - \mathrm{Dec}_{\rm phot} \quad . \tag{4.4}$$

Hence, $R_{\rm match}$, corresponds to the distance to the nearest photometric match. Since the 3D-HST catalog is selected on the deep ($H_{\rm F140W} \approx 26.1$; Brammer et al. 2012) high-resolution near-infrared HST F140W images, and the photometric catalogs are ground based and shallower, not all 3D-HST objects show up in the photometric catalogs. Some objects are therefore matched to another object's photometry. We adopt $R_{\rm match} \leq 0.3''$ to avoid contamination by other objects in the photometric catalog. Also, faint objects have the risk of being assigned to a bright(er) counterpart's photometry, so we restrict ourselves to objects with $m_{\rm F140W} \leq 23.5$.

Last but not least, the 3D-HST catalogs provide a redshift estimate of the objects based on the extracted grism spectra. The catalog grism redshifts, $z_{\rm grism}$, are obtained by collapsing the 2D grism spectrum into a 1D spectrum, combining it with available photometry, and then estimating the redshift with an updated version of the EAZY code (Brammer et al. 2008). As illustrated in Figure 1 of Brammer et al. (2012), the redshift range where Hα and/or [OIII] emission fall in the G141 grism wavelength range is $0.7 < z < 2.3$. We are interested in tracing the star formation in the merging systems via either Hα or [OIII] emission and therefore use $z_{\rm grism}$ to select objects in this particular redshift range.

Applying these five initial cuts (listed in the top half of Table 4.1) reduced the full sample of 21460 detections in the 30 GOODS-S and COSMOS pointings to 1542 objects.

4.4.2 FITTING SEDS TO PHOTOMETRY

Besides putting constraints on the 3D-HST catalogs, we also want to make sure that our sample has a high chance of showing pronounced emission line features in the spectra, i.e., pronounced star formation, to be mapped. To ensure this, we select objects with large star formation rates (SFR) and specific SFR (sSFR) as well as stellar masses (M_*) in a fixed range. We obtain the SFR, sSFR and M_* of each individual object from modeling the SED based on the ancillary photometric catalogs with the code MagPhys presented in da Cunha et al. (2008). We use the 37 NMBS bands (Whitaker et al. 2011) for the COSMOS objects and the 17 FIREWORKS bands (Wuyts et al. 2008) for the GOODS-S objects. Both catalogs span from the far UV to MIPS 24 μm. The 3D-HST catalog redshift $z_{\rm grism}$ is used as a prior when fitting the ancillary photometric data for each objet.

Despite the fact that the photometry measurements in most cases belong to two or more blended components, selecting the high-SFR objects based on SED fits to the photometry is still very effective in selecting objects with strong emission line features.

As shown in the bottom part of Table 4.1 we select objects with SFR $> 10 M_\odot$/yr, sSFR $> 10^{-9.5}$yr^{-1} and $10^9 M_\odot < M_* < 10^{12} M_\odot$. A SFR of $10 M_\odot$/yr roughly corresponds to $F_{\rm H\alpha} \sim 10^{-16}$ erg/s/cm^2 at $z = 1.5$, which corresponds to a (collapsed) emission line signal-to-noise (S/N) of ~8 at 1×10^{-16} erg/s/cm^2 (Brammer et al. 2012). Hence, concentrated star formation, i.e., emitted from a modest amount of pixels of this order, should be well detected in the 3D-HST spectra. On the other hand, if the total emission line flux $F_{\rm H\alpha}$ is spread over a larger area, the S/N per pixel might become too low for clear detection (see Section 4.6).

There are 353 of the initial 1542 objects in the $0.7 < z < 2.3$ range that satisfy these SED criteria.

4.4.3 VISUAL INSPECTION OF NEAR-INFRARED MORPHOLOGY

After reducing the number of objects to 353, the crucial step in the merger sample selection process, the visual inspection, can be performed. The visual inspection is based on the near-infrared F140W morphology of the objects from the 3D-HST direct imaging. The near-infrared images show the rest-frame optical emission at the redshifts of our galaxies. Hence, the near-infrared morphology traces the distribution of the (intermediate-age) stellar component of the galaxies, and it is therefore different from the stars being formed now (current star formation), which are traced by the emission lines. The criteria used to select the mergers from the 353 objects are that (i) they should show a morphology that differs from the bulk of the ‘normal’ isolated galaxies, and (ii) they have to show several distinct components in the continuum image, either multiple objects within the on average 2×2 arcsec F140W thumbnails or pronounced tidal features extending from the main continuum emission component. It should be noted, that because our merger selection is based only on this visual classification, and the low resolution of the G141 grism ($R \sim 130$) does not offer any kinematic information of the individual companions, making it impossible to address whether or not the systems are gravitationally bound, in principle our merger sample consists of *potentially* merging systems. We will therefore assume that the majority of objects are indeed bound and are at the same redshift, and are therefore gravitationally interacting. We do that primarily based on the relatively small distances involved at $z \sim 1.5$ where $0.3''$ corresponds to a proper distance of approximately 3 kpc.

The visual inspection is not only used to select morphologically disturbed systems. By inspecting the grism spectrum of each object we can also remove objects that do not show any obvious emission line features, despite the SFR, mass, and redshift cuts described above. Without such emission line features, creating a star formation map is impossible and these objects are therefore discarded.[3] Inspecting the grism spectra also ensures that any strong contamination is due to the different components of the merging system and not due to interlopers.

From the parent sample of 353 catalog-selected objects, the visual inspection discards 292 objects as either being isolated 'normal' galaxies (252/292), having high contamination not stemming from the merging components (13/292) or having no obvious emission line features in their spectra (24/292). Hence, our final sample of (potential) mergers from 6 GOODS-S and 24 COSMOS 3D-HST pointings consists of 61 systems corresponding to ~17% of the 353 catalog-selected objects.

In Figure 4.1 we have plotted the 61 mergers as large solid points together with the selection regions from Table 4.1 (gray shaded regions). The small gray dots represent the 292 objects discarded by the visual classification, i.e., the general galaxy populations satisfying the selection cuts in Table 4.1. The distributions of the 61 merger's SFR, sSFR, $z_{\rm grism}$ and M_* are shown as histograms on the axes of the scatter plots. In Figure 4.2 we show the $m_{\rm F140W}$ magnitude distribution of these objects. In both the histograms in Figure 4.1 and in Figure 4.2 the dotted lines correspond to the 16th and 84th percentiles of the distributions, whereas the dashed lines show the median values. In Section 4.7 we will use these values to determine the parameter space to sample when simulating 3D-HST grism spectra.

---

[3] Assuming that the estimated SFR is correct, these objects are in them selves very interesting, as they might be highly dust obscured systems suffocating the star formation emission. 7% of the 353 catalog-selected objects, corresponding to 24 of the 292 visually discarded objects, fell in this category but are ignored for the present study.

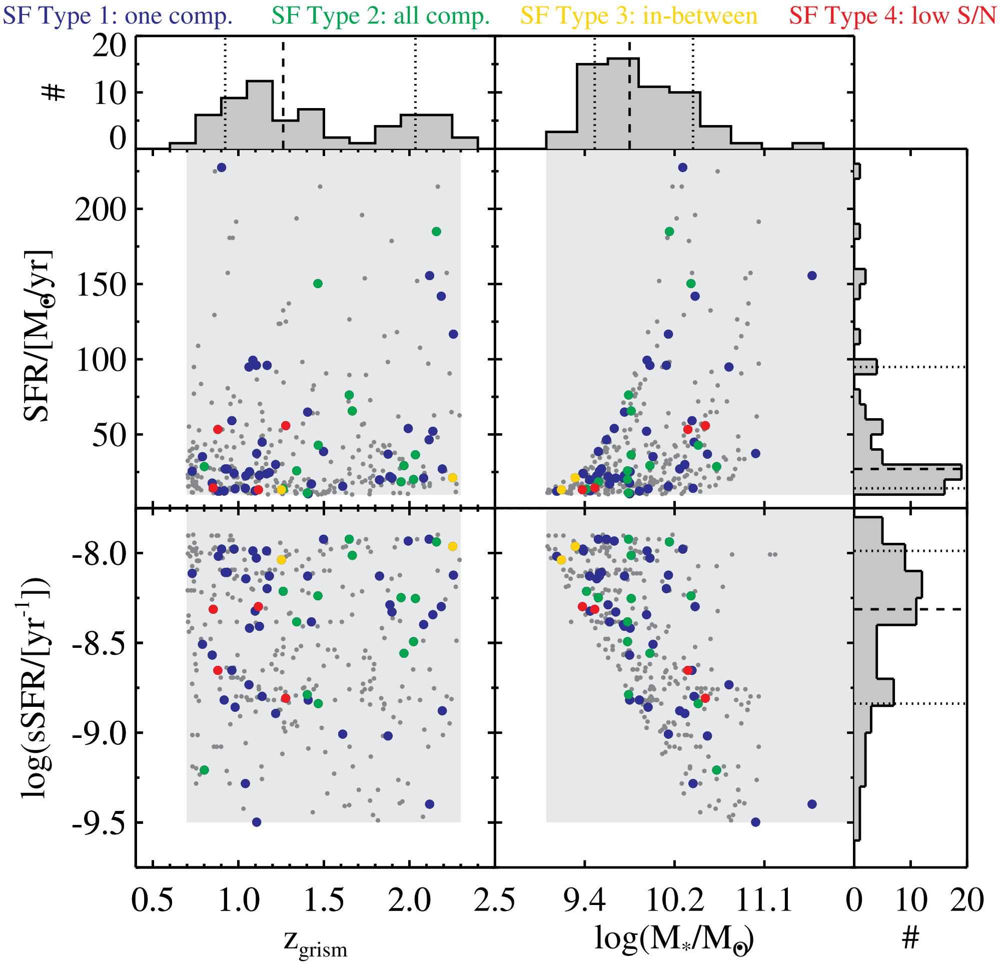

FIGURE 4.1: *The 61 3D-HST mergers (large points) plotted in the main region of the catalog selection space given in Table 4.1 (gray shaded regions). The small gray points represent the 292 objects discarded by the visual inspection. The histograms attached to the scatter plots show the distribution of z_{grism}, M_*, SFR and sSFR for the 61 mergers. The dotted and dashed lines in the histograms indicate the 16th and 84th percentiles and the median of the distributions, respectively. These values are used to define the parameter space of the simulated grism spectra described in Section 4.7. The points are color coded according to the morphology of their emission line maps (SF type) described in Section 4.6 as indicated on the top. No obvious trends between SF type and SFR, sSFR, redshift or stellar mass are found.*

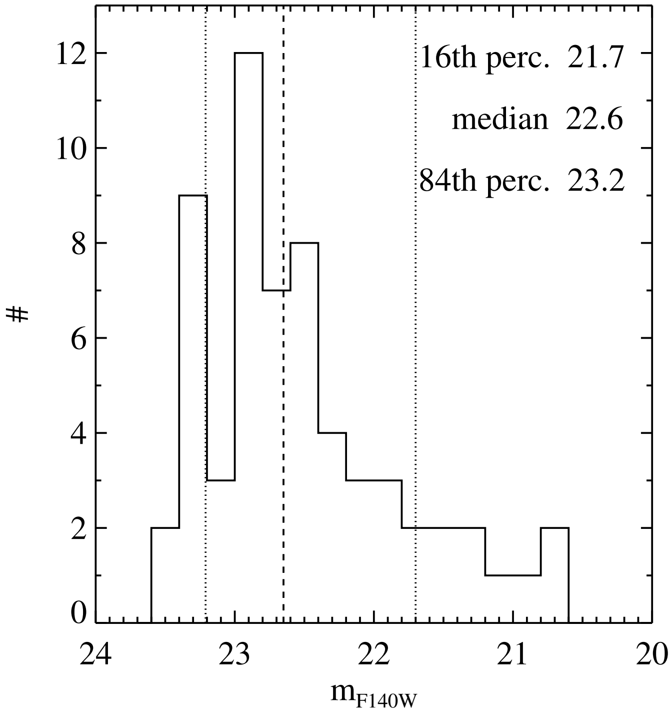

FIGURE 4.2: *The distribution of* m_{F140W} *magnitudes for the 61 3D-HST mergers. The dotted lines indicate the 16th and 84th percentiles of the distribution, whereas the dashed line shows the distribution median. The 16th and 84th percentile values stated in the upper right corner are used when defining the parameter space for the simulated grism spectra in Section 4.7.*

In the left panel of Figure 4.3 we compare the 3D-HST catalog redshifts to the photometric FIREWORKS and NMBS redshifts of the merger sample. It is clear that the overall agreement is fairly good with a few outliers resulting mostly from poor photometric redshifts.

4.5 EMISSION LINE MAPPING

To quantify the extent of (unobscured) star formation in the 61 3D-HST mergers described in the previous section, we rely on the spatial information of the Hα and [OIII] emission lines the slitless grism spectroscopy described in Section 4.3.1 provides. From the grism spectra we create emission line maps by subtracting a model of the continuum light in the grism spectra, such that only the probed emission line feature is remaining. This can then be mapped back onto the near-infrared continuum light-distribution of the object. In practice we:

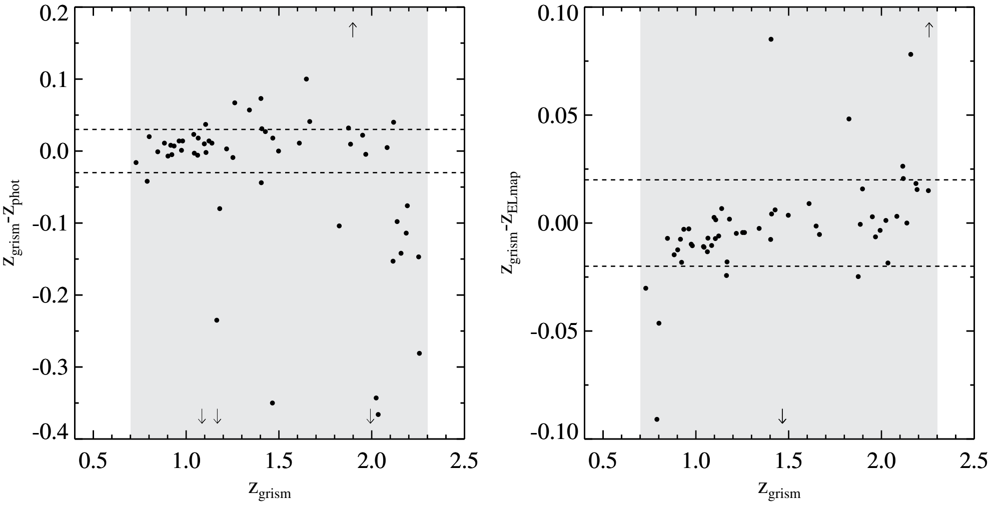

Figure 4.3: *The left panel shows the comparison of the photometric redshift from the FIREWORKS and NMBS catalogs (z_{phot}) with the grism redshifts from fitting the collapsed 3D-HST spectra (z_{grism}). The horizontal dashed lines indicate agreement on the 3% level between z_{grism} and z_{phot}. The arrows indicate objects falling outside the shown region. The right panel compares the catalog grism redshift (z_{grism}) with the redshift determined via the emission line mapping method described in Section 4.5 (z_{ELmap}). These redshifts are obtained from the same data with the same emission lines, hence, the (minor) disagreement mainly reflects the uncertainty in estimating λ_{CCmax} as described in Section 4.5.2.*

1. create a 2D continuum model for the grism spectrum.
2. subtract this continuum model from the 2D grism spectrum itself.
3. cross-correlate the F140W thumbnail with the continuum subtracted spectrum to determine the emission line map redshift, z_{ELmap}.
4. map the (cut-out) continuum subtracted thumbnail corresponding to z_{ELmap} back onto the F140W thumbnail.

Each of these steps are described in details in the following subsections and are illustrated in Figure 4.4.

4.5.1 Continuum Modeling and Subtraction

The 2D continuum models we subtract from the 3D-HST spectra of the mergers are based on a third order polynomial fit to a one-dimensional spectrum obtained from a weighted collapse of the individual 2D spectra. The polynomial fit to the 1D spectrum is turned back into a 2D continuum model by concatenating rows with the 1D polynomial form

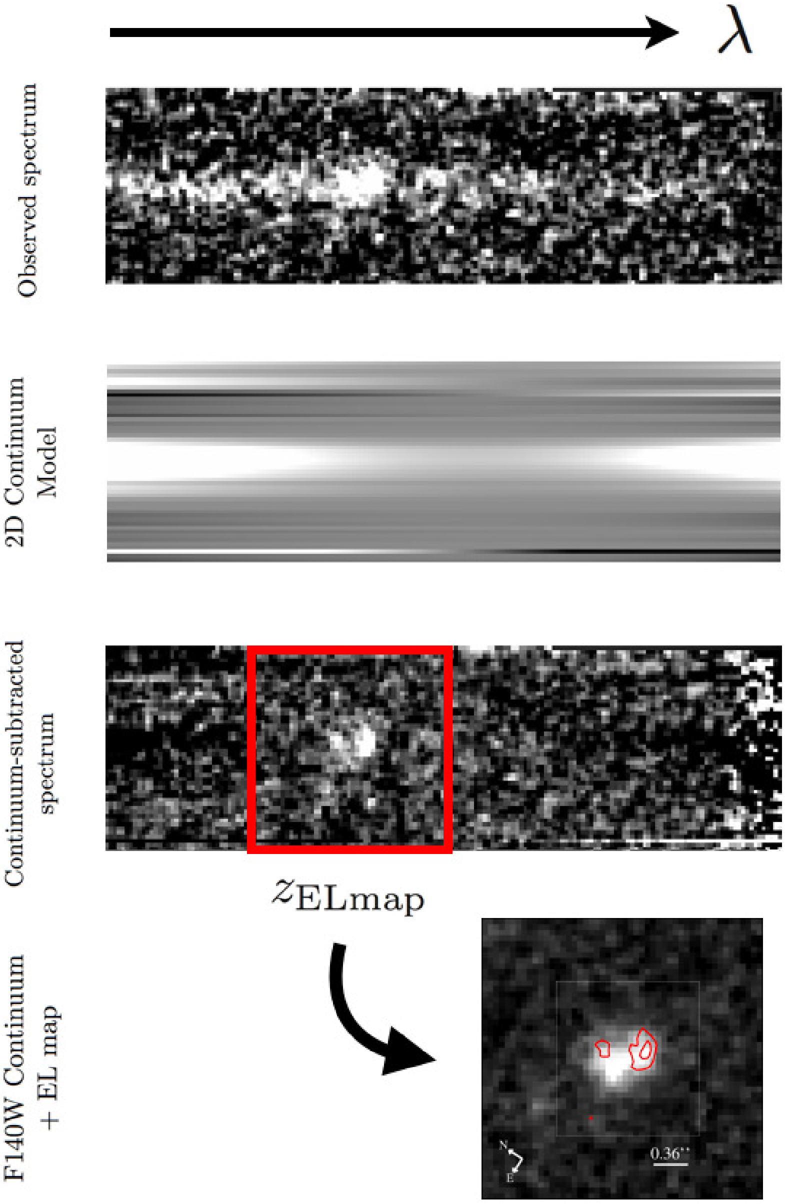
λ
Observed spectrum
2D Continuum
Model
Continuum-subtracted
spectrum
z_{ELmap}
F140W Continuum
+ EL map
N
E
0.36''

FIGURE 4.4: *Illustration of the steps performed to obtain the emission line (star formation) maps as described in Section 4.5. The top panel shows a standard 3D-HST spectrum with a prominent emission line feature. The 2D continuum model is shown just below that. The third panel shows the spectrum from the top panel after subtraction of the 2D continuum model. The red square marks the* z_{ELmap} *emission line map cut-out. The bottom panel illustrates how the emission line map* z_{ELmap} *cut-out is turned into (red) contours on the near-infrared F140W thumbnail of the object.*

weighted by a 'slit-profile', which is obtained from a weighted average of the columns blueward and redward of the probed emission line feature in the full 2D spectrum. Thus, we have a continuum model with the shape of a polynomial in the dispersion direction scaled in each row to a 'slit-profile' of the original spectrum. Subtracting this model from the 3D-HST spectrum returns a two-dimensional emission line map as illustrated in Figure 4.4, where all that is left is the emission line feature. This approach is the same as the one used in Nelson et al. (2012). Often when dealing with slitless spectroscopy and the 3D-HST grism spectra in particular, the goal is to remove contamination in a systematic manner. However, we relied on the visual inspection to remove badly contaminated objects, as mergers per definition are contaminated. This allowed contamination sometimes complicates the continuum subtraction, but in virtually all cases the probed emission line region is continuum-subtracted to a satisfactory level.

4.5.2 THE EMISSION LINE MAP REDSHIFT, z_{ELMAP}

We used a simple cross-correlation between the near-infrared F140W thumbnail continuum image of each object and the corresponding full 2D emission line map in order to map the emission line map back onto the near-infrared image. In practice we calculate

$$\mathrm{CC}(\lambda) = \sum_{i}^{N_{\mathrm{width}}} \sum_{j}^{N_{\mathrm{width}}} f_{i,j,\mathrm{F140W}} \times f_{i,j,\mathrm{ELmap}} \tag{4.5}$$

for each individual of the first $k = N_{\mathrm{2D}} - N_{\mathrm{width}}$ columns in the full 2D emission line map, where N_{2D} is the number of columns in the 2D emission line map and N_{width} is the width of the F140W thumbnail. The $f_{i,j,\mathrm{F140W}}$ and $f_{i,j,\mathrm{ELmap}}$ is the flux in the pixel (i, j) for the F140W thumbnail and 2D emission line map cut-out (indicated by the red box in Figure 4.4), respectively.

The maximum of the cross-correlation function CC indicates the wavelength, λ_{CCmax}, where there is the largest overlap between the near-infrared light distribution of the object and the kth cut-out of the full emission line map. The λ_{CCmax} corresponds to the kth column of the 2D emission line map plus $N_{\mathrm{width}}/2$.

In order to transform this wavelength into a redshift, prior knowledge about the probed emission line feature is needed. This can either be obtained via wavelength ratios if multiple emission line features are available for the same object, for instance at redshifts between 1.1 and 1.5 where both the Hα and [OIII] line fall within the G141 range, or by invoking the emission line rest wavelength (λ_{EL}) using prior knowledge about the redshift. Since only a small fraction of our objects show multiple significant emission line features, we use the latter approach to estimate the emission line map redshifts, z_{ELmap}. Using z_{grism} to identify the emission line feature, the emission line map redshifts can be determined as

$$z_{\mathrm{ELmap}} = \frac{\lambda_{\mathrm{CCmax}} - \lambda_{\mathrm{EL}}}{\lambda_{\mathrm{EL}}} \quad . \tag{4.6}$$

In the right panel of Figure 4.3 we compare the redshifts determined from the emission line maps with the grism redshifts. A few outliers (marked by the arrows) are observed but overall we find an excellent agreement between the two redshifts: the horizontal dashed lines show redshift agreement on the 3% level. This is expected as z_{ELmap} and z_{grism} are obtained from the same data with the same emission lines. The fairly modest scatter reflects the uncertainty in estimating λ_{CCmax}.

In Figure 4.5 we show a collection of emission line maps (red contours) from our 3D-HST merger sample. The individual maps correspond to the region at λ_{CCmax} that has been mapped back onto the near-infrared continuum image as illustrated in Figure 4.4. The different 'morphologies' of these maps will be addressed in Section 4.6. In Appendix D on page 151 we show the full sample of 3D-HST merger emission line maps.

4.6 RESULTS: THE SPATIAL EXTENT OF STAR FORMATION IN HIGH-z MERGERS

As noted in the introduction, star formation and mergers are important parts for understanding how high-z galaxies evolved into the galaxies we observe in the low-z Universe. In the previous sections we have described how we select the mergers from the 3D-HST data, and subtract the continuum light in the spectra to create the emission line (star formation) maps. In this section we will characterize the morphological *type* of the spatial extent of the star formation in the 61 3D-HST mergers. However, we first divide our sample into two different kinds of mergers: the 'pre-mergers' and the 'post-mergers'. By pre-mergers we mean objects that show multiple clearly-distinct and pronounced continuum peaks in the near-infrared images, i.e., the optical continuum emission comes from multiple objects in the process of merging or about to merge. Examples of those are shown to the left in the two top panels of Figure 4.5 (maps framed in yellow). The post-mergers, on the other hand, are systems that have undergone merging and now appear to be dominated by a nuclear feature in the continuum with pronounced tidal features surrounding it. Examples of some of these systems are shown to the right in the two top panles in Figure 4.5 (maps framed in red). Dividing the 61 mergers into these two sub-samples return 33 pre-mergers and 28 post-mergers.

To characterize the location and the spatial extent of the star formation in the 3D-HST mergers we categorize the star formation maps into the following four morphological types (SF type):

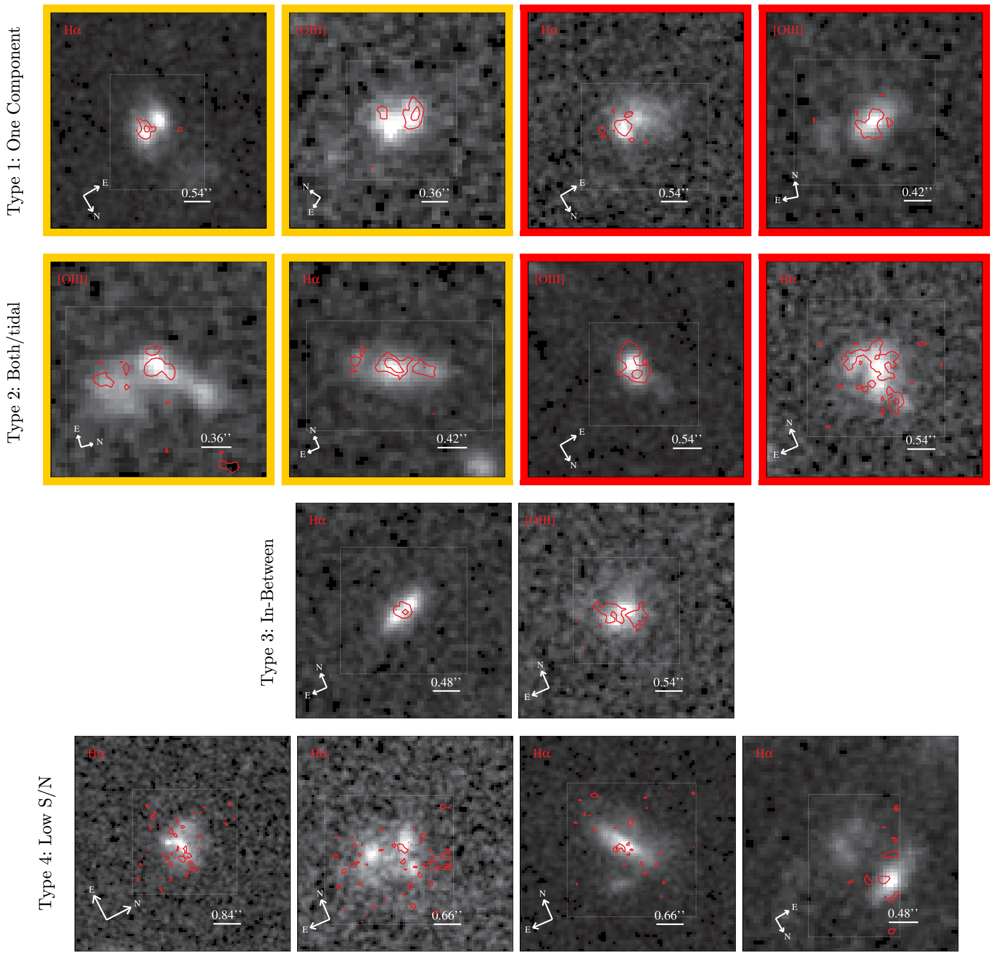
Type 1: One Component
Hα
0.54''
[OIII]
0.36''
Hα
0.54''
[OIII]
0.42''
Type 2: Both/tidal
[OIII]
0.36''
Hα
0.42''
[OIII]
0.54''
Hα
0.54''
Type 3: In-Between
Hα
0.48''
[OIII]
0.54''
Type 4: Low S/N
Hα
0.84''
Hα
0.66''
Hα
0.66''
Hα
0.48''
N
E

Figure 4.5: *Examples of Hα and [O*III*] emission line maps (red contours) plotted on the WFC3 F140W thumbnails (gray scale) for a sub-subsample of the 61 3D-HST mergers. The dashed boxes indicate the mapped region. Each row represents one morphological type of star formation distribution as described in Section 4.7.2. From top to bottom each row show maps of SF type 1 ('one component' maps), SF type 2 ('both (all) components' maps), SF type 3 ('in-between' maps), and SF type 4 ('low S/N' maps). The left-hand emission line maps of the top two rows (framed in yellow) are pre-mergers, whereas the right-hand maps (framed in red) are post-mergers. The full sample of emission line maps is shown in Appendix D on page 151.*

1. **One Component:** The star formation is significantly more pronounced in one of the merger components. For the pre-mergers this means one of the multiple objects and for the post-mergers this refers to the nuclear region or the tidal feature(s).
2. **Both (All) Components:** The star formation is pronounced/detected in all (or the majority if more than two) components of the system, it being individual objects or the nuclear and tidal region.
3. **In-between:** The mapped star formation appears to be emerging from in-between the merging components. None of the post-mergers show this feature, so indeed this means in-between clearly distinguishable objects.
4. **Low S/N:** The S/N per pixel of the emission line features in the 2D grism spectrum was too low to produce a convincing emission line map as might be the case for extended star formation as pointed out in Section 4.4.2.

Each of the four rows of star formation maps in Figure 4.5 show examples of these four SF types. The results from characterizing the two classes of mergers with these four SF types are shown in Figure 4.6. The error-bars are obtained by bootstrapping the results, i.e., by randomly drawing 61 SF types from the results a thousand times and then using the 2σ width of the resulting SF type distributions as error-bars (hence, no error-bar on the post-merger SF type 3 in Figure 4.6; none were found). For both the pre- and post-mergers, the star formation is most prominent in just one of the components (SF type 1) for roughly 3/4 of the objects. In roughly 1/5 of the objects, star formation was detected in all components (SF type 2). Hence, the distribution of the spatial extent of star formation among the objects in the two subsamples is consistent with being the same. In Table 4.2 we have listed the exact fractions for the total sample of 3D-HST mergers resulting from the classification of the emission line maps.

That the fraction of objects with more prominent star formation in just one component (SF type 1) is three times larger than the cases with star formation of type 2 might be a consequence of dust obscuration. As mentioned, we are only able to probe the unobscured star formation, so in cases where one component (or the tidal feature) is much more dust-obscured than the other, we would end up with star formation maps of type 1. It could also simply be due to different SFRs in the different components. Since the mergers are selected on morphology and we, as mentioned, do not have any kinematic information, the fraction of SF type 1 objects might be biased by chance superpositions of objects on the sky at different redshifts, such that we only see line emission from one object in the near-infrared. However, the fraction of such contaminants is not expected to be significant compared to the difference observed in the two SF types due to the small distances involved at these redshifts.

Hence, assuming that the fraction of systems that look like mergers due to a projection

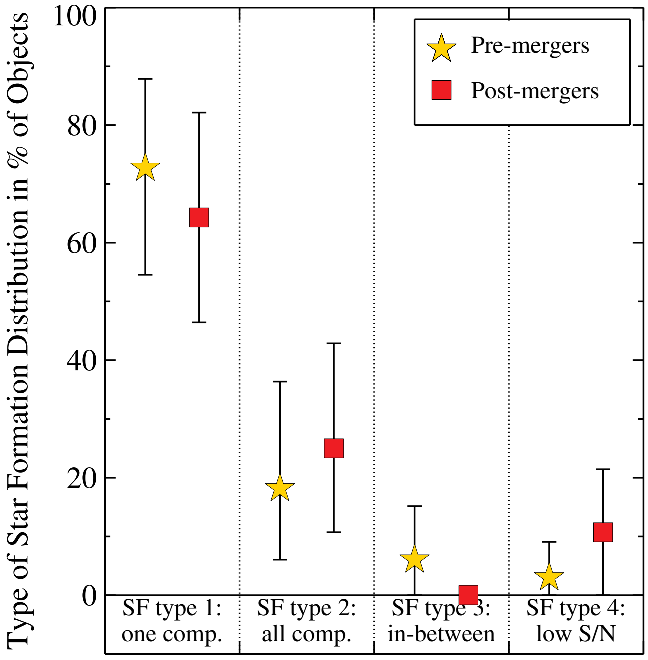

Figure 4.6: *The morphological types of the star formation distribution (Section 4.6) observed in the 61 3D-HST mergers split into 'pre-mergers' of multiple individual objects (33 objects, yellow star) and 'post-mergers' of systems with a nucleus and tidal features (28 objects, red squares). The error-bars are obtained via bootstrapping. The two samples have similar star formation distributions. Only* $\sim 20\%$ *of the objects show star formation in all merger components (individual objects or nucleus and tidal feature) whereas more than 60% of the systems only show star formation in one component.*

effect on the sky is insignificant, the results (maybe as expected) indicate that most mergers happen between objects of different gas fractions and/or different SFR. We will show below that this is backed up by an initial comparison with simulated mergers.

In Figure 4.1 we have color coded the points of the 61 mergers according to the SF types to look for dependencies between the morphology of the star formation maps and SFR, sSFR, $z_{\rm grism}$, and M_*. As we do not find any significant correlations this suggests that all star formation morphologies occur at all redshifts irrespective of SFR and mass.

4.7 Simulating 3D-HST Spectra

With the exceptional data of the 3D-HST merger sample presented above, we can for the first time perform meaningful comparisons with the star formation produced in simulated

TABLE 4.2: *The Spatial Extent of Star Formation*

SF type	3D-HST		Simulations	
1: One Component	69%	(42/61)	28%	(83/296)
2: Both (All) Components	21%	(13/61)	59%	(175/296)
3: In-Between Components	3%	(2/61)	0%	(0/296)
4: Low S/N per pixel	7%	(4/61)	13%	(38/296)

See Figure 4.8 for a plot of these values.

mergers at high redshift. Modern high-resolution merger simulations are able to predict the spatial distribution of star formation in mergers and produce simulated images by including the emission by young, newly-formed stars and the transfer of starlight through gas and dust (Cox et al. 2006, 2008; Jonsson et al. 2010). These simulations have shown various examples of star formation triggered by direct galaxy interaction, star formation originating in the central cores of the individual merging components, and star formation appearing in tidal features. In the nearby Universe several examples of such features have been found, e.g., in the antennae galaxy (Wang et al. 2004). Whether the predictions of simulations are also representative of the star formation in mergers at $z \sim 1.5$, i.e., when the Universe was only 3–4 Gyr old, has not yet been tested. The simple reason for this is as mentioned that the data has not been available before now. With 3D-HST, the sample size is becoming large enough that we can start looking at a *population* (snapshot) of mergers instead of individual case studies, and hence comparisons with predictions from simulations become feasible. In this section we will make an initial illustrative attempt at comparing state of the art merger simulations (described in Section 4.2.2) with the observational results from 3D-HST. In the following we describe how we create simulated WFC3 G141 grism spectra from a small initial set of these high resolution simulated mergers.

4.7.1 SPH MERGER SIMULATIONS

To simulate grism spectra, having knowledge about both the spatial and spectral extent of the object is crucial. As input for our grism simulations we therefore used the three-dimensional data-cubes from the N-body/SPH mergers from GADGET described in Section 4.2.2. For our initial comparison of simulations with actual data, we used the three models listed in Table 4.3. The three simulations are all of mass ratio 1:2 and the two merging objects in each simulation both have a gas fraction (f_{gas}) of 40%. The three mergers happen at three different relative orientations (Θ) of the two merging components. Hence, what we are testing here is the effect of orientation on the detected star formation distribution. We further have a series of time steps or snapshots of each galaxy

TABLE 4.3: *The Simulated 3D Input Data-Cubes*

Name	N_t	N_{view}	Mass ratio*	f_{gas}	Θ
Sim. 1	6	3	(Sb) 1:2 (Sc)	0.4:0.4	30°
Sim. 2	5	3	(Sb) 1:2 (Sc)	0.4:0.4	90°
Sim. 3	3	3	(Sb) 1:2 (Sc)	0.4:0.4	150°

* *Parenthesis give approximate galaxy type.*

(N_t) and a set of viewing angles (N_{view}) that we can simulate spectra for. Thus, the simulation parameter space (sparsely) sampled here is spanned by time, orientation of the two mergers and viewing angle. As described below we combine this with a set of parameters determined by the 3D-HST sample to define the total parameter space to be covered by the simulated spectra. For further information on the input simulations we refer to Cox et al. (2006, 2008); Jonsson (2006) and Jonsson et al. (2010).

4.7.2 SIMULATING 3D-HST GRISM SPECTRA

Simulating the grism spectra from the merger simulation output data-cubes presented above and in Section 4.2.2 is fairly straightforward. The grism spectrum is created by dividing the data-cube into wavelength slices, i.e., images corresponding to a 'filter' of width $\Delta\lambda$. Offsetting or dispersing this sequence of images and co-adding the fluxes results in a grism spectrum. Using this approach we have turned the input data-cubes into a sequence of simulated WFC3 G141 grism spectra. We use a pixel scale of 0.06 arcsec and a spectral resolution of $\Delta\lambda = 22\text{Å}$ according to the 3D-HST grism spectra characteristics described in Brammer et al. (2012). For the HST PSF we use a Tiny Tim PSF[4] and for the system throughput we used the G141 sensitivity curve made available on the WFC3 website.[5]

We let the observed sample of mergers from 3D-HST determine the dimensions of the simulation parameter space. By simply re-scaling each data-cube we simulate grism spectra corresponding to the 16th and 84th percentiles of the magnitude (m_{F140W}), redshift (z_{grism}) and SFR distributions of the 3D-HST data as indicated by the dashed lines in the corresponding histograms in Figure 4.1 and 4.2.

To adjust the SFR of the data-cubes before turning them into grism spectra, we assume that the SFR to Hα-flux conversion follows the empirical relation of Kennicutt et al.

---

[4] `http://www.stsci.edu/hst/observatory/focus/TinyTim`

[5] `http://www.stsci.edu/hst/wfc3/analysis/grism_obs/wfc3-grism-resources.html`

(1994):

$$\frac{\mathrm{SFR}}{[M_\odot/\mathrm{yr}]} = \frac{1}{1.26\times10^{41}}\,\frac{L_{H\alpha}}{[\mathrm{erg/s}]}\ . \tag{4.7}$$

This relation assumes a Salpeter initial mass function. The m_{F140W} is obtained by scaling the flux integrated over the F140W passband independently. Combining the values from Table 4.3 with these data-determined parameters, we end up with a sample of 336 simulated spectra spanning the parameter space $(N_t, N_{\mathrm{view}}, \Theta, m_{\mathrm{F140W}}, z, \mathrm{SFR})$.

Added to the 'raw' grism spectra are Poisson noise and noise terms corresponding to the read noise and dark current quoted on the WFC3 website, as well as the background sky levels presented in Brammer et al. (2012).

In Figure 4.7 we present a sequence of simulated G141 grism spectra. The spectra shown have various combinations of N_t, N_{view}, Θ, m_{F140W}, z, and SFR. The thumbnails on the left are the noise-free F140W images of the objects.

4.7.3 CREATING STAR FORMATION MAPS OF SIMULATED SPECTRA

To create the star formation maps for the simulated spectra, we treat them in exactly the same way as the actual 3D-HST spectra as outlined in Section 4.5. First we visually inspected the spectra to make sure that an emission line feature was available; for some of the high redshift, low SFR combinations the emission lines did not show up in the noise-added spectra. Having eliminated spectra without emission line features, as well as cases of viewing angle and time step were the two merging objects could not be distinguished, the sample of 336 simulated spectra was reduced to 296. For these 296 spectra we created emission line maps by the method described in Section 4.5.

As for the actual 3D-HST data, we characterized the 296 emission line maps of the simulated spectra according to the four SF types. The results from this classification is shown in Table 4.2 together with the results for the 3D-HST mergers presented in Section 4.6. The simulated spectra in Figure 4.7 show examples of both spectra resulting in star formation maps of type 1 (top three spectra) and type 2 (bottom three spectra).

4.7.4 COMPARING THE 3D-HST DATA WITH SIMULATIONS

We compare the 3D-HST SF types we inferred in Section 4.6 with the simulated SF types, in order to investigate what effects we are dealing with when trying to characterize the extent of star formation in the selected sample of 3D-HST mergers. In Figure 4.8 we compare the SF types of the 61 3D-HST mergers (blue circles) with the SF types

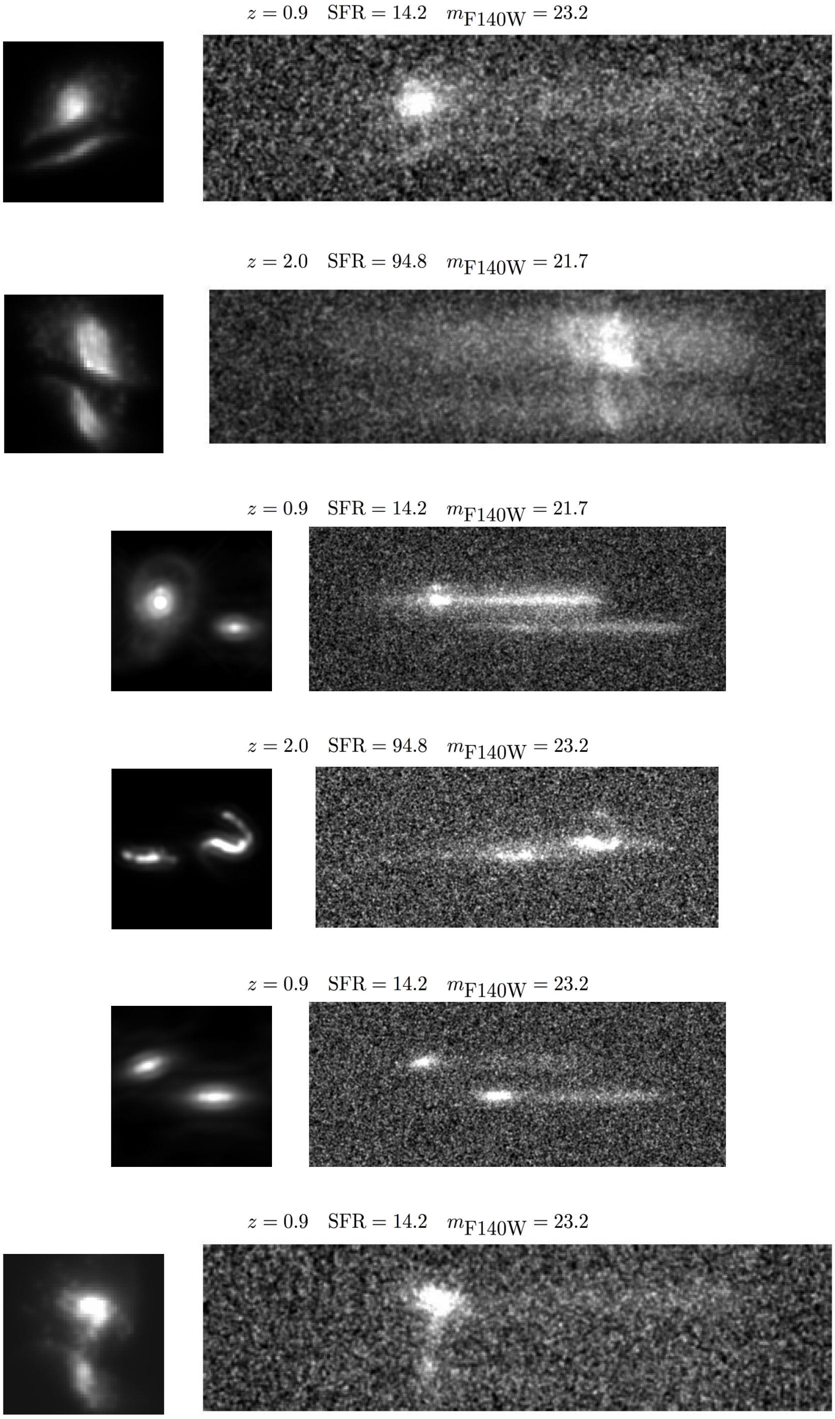

FIGURE 4.7: *Examples of simulated 3D-HST grism spectra. Various combinations of SFR, m_{F140W}, z, time-step in the merging of the components (N_t), viewing angles (N_{view}), and initial relative orientation of the two merging components (Θ) are shown. The first three spectra result in star formation maps of SF type 1, while the bottom three spectra result in SF type 2 star formation maps.*

of the 296 simulated mergers (green triangles). The plotted percentages are given in Table 4.2. The error-bars are again obtained via bootstrapping. The comparison needs to be done with care, as the sample of simulated mergers from Table 4.3 does not span the gas fraction and mass ratio dimensions; they are as mentioned fixed at 40% and 1:2, respectively. Nevertheless, the 3D-HST survey enables comparison of predictions from simulations with a statistical sample of high-z mergers for the first time. Even though the spanned simulation space is fairly limited, it seems that the fraction of spectra where the S/N per pixel is too low to create an actual emission line map is fairly consistent. Likewise, the fraction of cases where the star formation seems to emerge from in-between the merging components (SF type 3) is comparable: in fact we did not find any objects in this category for the simulations. Looking at the two SF type 3 cases we found for the data (third row of Figure 4.5) could indicate, that the reason we are not seeing star formation from the components themselves is due to dust-obscuration rather than more pronounced star formation in-between the components. To find similar trends in the simulated emission line maps we would need to investigate a range of gas fractions rather than just the $f_{\rm gas} = 40\%$ cases presented here.

The comparison also shows that the simulated spectra on average have almost three times as many star formation maps where both components have pronounced star formation (SF type 2), i.e., nearly 2/3 of the simulated maps as compared to ~1/5 for the 3D-HST star formation maps. Correspondingly the fraction of single component emission line maps is significantly lower for the simulations as compared to the observations. This suggests that the majority of mergers does not happen between galaxies of comparable gas fraction as the parameter space spanned by our simulations implicitly assume. This is in good agreement with the results presented in Section 4.6.

4.8 CONCLUSION

In this chapter (Schmidt et al. 2012b) we have presented the first statistical sample of mergers with 3D spectroscopy at redshift $z \sim 1.5$. The sample consists of 61 morphologically selected mergers from the Hubble treasury program 3D-HST with total masses and star formation rates for the systems derived from multi-wavelength photometry. From the slitless grism spectroscopy we created emission line maps of the rest-frame optical emission lines Hα and [OIII] as a proxy for the *spatial* extent of star formation in the full sample. This provides a comprehensive empirical picture of where star formation happens in galaxy mergers at $z \sim 1.5$ where the cosmic star formation and merger rate peaked.

We have also carried out an initial comparison of the 3D-HST mergers with recent SPH simulations of galaxy mergers that include star formation and dust extinction. These

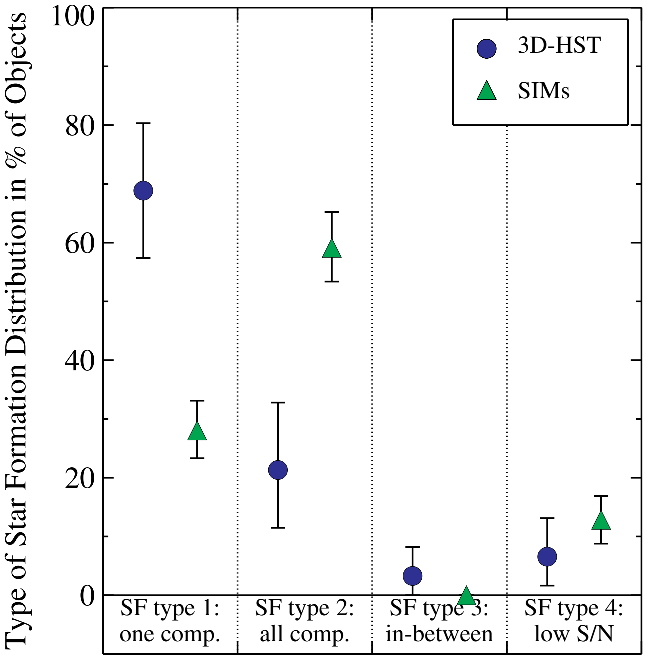

FIGURE 4.8: *Comparing the spatial extent of star formation (SF type; Section 4.6) in the 61 3D-HST mergers (blue circles) with the 296 simulated spectra (green triangles). Each point has been assigned an error-bar obtained via bootstrapping. Direct comparisons should be done with care due to the limited size of the parameter space the simulations span. Nevertheless, it is evident that the simulations show many more cases where star formation is seen in both the merging components than the 3D-HST data do, indicating that the majority of mergers have different gas fractions prior to merging.*

simulations point towards larger and more extensive simulation programs aimed at spanning the full parameter space of the observations, to address whether these support the predictions of modern merger simulations at high redshift or not (see Section 4.9.4).

The main conclusions of the study are:

- The spatial distribution of star formation in $z \sim 1.5$ mergers shows a broad range of morphologies. It is often concentrated in a single, compact region, but can also be located in tidal tails or in between the main stellar bodies of the progenitors.
- In the majority (69%) of the early-stage, pre-coalesence mergers the star formation is significantly more pronounced in just one of the merging components.This is likely due to different gas fractions of the progenitors. Alternatively, dust content may differ in quantity or configuration.
- The range of star formation morphologies show no clear correlations with the es-

timated SFR, sSFR, $z_{\rm grism}$, or M_*, suggesting that all star formation morphologies are presents at all epochs irrespective of star formation rate and mass.

- Simulated mergers among galaxies with similar masses and similar gas fractions typically predict similarly intense star formation in both merger components, at odds with the observations: as opposed to 21% of the observed mergers, as many as 59% of the simulated mergers show detectable star formation in both components.This discrepancy lends further support to the notion that $z \sim 1.5$ mergers typically occur between galaxies with different gas fractions.

4.9 OUTLOOK

As mentioned, little is know about the star formation in galaxy mergers at high redshift. Until recently large 3D spectroscopic samples of galaxy mergers at high redshift ($z > 1$) did not exist. The study presented in this chapter opens up for such investigations and illustrates the possibilities and prospects of the 3D-HST survey presented in Section 4.3. Using just 1/4 of the full 3D-HST survey, we were for the first time able to quantify the spatial extent of star formation in the largest sample of mergers at $z \sim 1.5$. Furthermore, we, also for the first time at these redshifts, performed an initial comparison of the observational results with simulated star forming mergers. Here we will describe some of the obvious, and maybe not so obvious, next steps the work presented in this chapter opens up for.

4.9.1 EXPANDING THE ANALYSIS TO THE FULL 3D-HST SURVEY

The first obvious extension of the work presented in this chapter is expanding the analysis to all 120 pointings of the 3D-HST survey. The data of the last pointings are being taken and reduced at the moment of writing, meaning that the final data product of the survey will be available by autumn 2012. Even though the present study has presented the largest sample of morphologically selected disturbed pre- and post-merging systems at the main epoch of galaxy formation to date, the sample size is still fairly modest. Hence, expanding the sample by a factor of four will without a doubt strengthen the arguments presented here and provide stronger statistical evidence for the conclusions made. However, a detailed assessment of the quality of the results, i.e., whether the assumptions made are fully justified, also needs to be carried out as described in Section 4.9.2 below.

Using the full sample of ~7000 galaxies with convincing grism redshifts in the final product of 3D-HST to define a larger merger sample will not only support the work presented here, but will also allow a full scale investigation of the merger rate at the probed redshifts of $1 < z < 3.5$. Today, there are no indications of the amount of mergers at this epoch based on such a large spectroscopic sample. Knowing how the rate of merges has evolved the last several Gyr is, as mentioned in Section 1.4, a crucial key for understanding the build-up of the galaxies we observe in the nearby Universe.

4.9.2 IFU FOLLOW-UP OF OBJECTS

An important step in expanding the analysis to the full 3D-HST survey, is to assess the quality of the results presented here. As mentioned the merger selection is based on visual inspection of the near-infrared morphology of the 3D-HST objects (after reducing sample sizes by down-selection on redshift, star formation rate, quality of the spectra, etc.) and hence are in principle only merger *candidates*. The low resolution of 3D-HST unfortunately does not provide any kinematic information, so to assess whether the selected (potential) mergers are indeed gravitationally bound systems, higher resolution spectroscopic follow-up is needed. Also confirming that the fraction of interlopers (i.e., chance superpositions of objects on the sky) is indeed insignificant, as was assumed here, will be an important step in expanding to the full 3D-HST survey.

Obtaining kinematics for the 3D-HST mergers is not only interesting for confirming that the merger selection function is indeed doing what we designed it to, namely selecting highly star forming interacting or highly disturbed galactic systems. It will also provide information about the mass ratios of the selected mergers and provide total masses of the individual systems. Furthermore, spectroscopic follow-up at higher resolution will enable us to asses whether the assumption that the emission line features are probing star formation is indeed true. As described in Section 3.2, AGN also have pronounced emission lines, and there is therefore the risk that some of the more centrally concentrated emission line maps presented in this chapter are probing AGN emission instead of star formation. However, with the higher resolution data we would be able to obtain emission line ratios like [OIII]/Hβ and/or Hα/[NII], which as described in Section 3.2.3 and illustrated by the BPT diagram in Figure 3.2 on page 66, is capable of distinguishing between the two possible causes for the strong emission lines.

We have recently submitted a proposal with the aim of doing all of the above (ESO P90: "Spatially Resolved Kinematics of Mergers at $z \sim 2$"; PI K. B. Schmidt). The goal is to observe the Hα (and [NII]) emission of 20 out of the 61 3D-HST mergers with the high resolution near-infrared IFU SINFONI. The candidates have been selected based on z_{grism} to make sure that the Hα feature falls in the near-infrared H or K band which are available on SINFONI.

4.9.3 STAR FORMATION AND MOLECULAR GAS AT HIGH REDSHIFT

Another aspect of galaxy formation that the 3D-HST survey enables the exploration of, is the question of whether the location of star formation coincide with molecular gas at high redshifts. With ALMA's call for early science cycle 0 proposals in 2011 and cycle 1 coming up, it has now become possible to follow up 3D-HST systems at redshifts $z \sim 2$ obtaining molecular gas maps of high enough resolution to directly compare them with the sub-arcsec resolution of the 3D-HST grism spectroscopy. This will enable us to answer whether star formation at high redshift is also primarily happening in dense molecular clouds as has been found to be the case in the Milky Way (Williams et al. 2000; Blitz 1993) and the very local Universe (Blitz et al. 2007).

ALMA will also be able to estimate the *total* star formation in the mergers. As mentioned we are only able to probe the unobscured star formation with the Hα and [OIII] emission lines. However, in the radio where ALMA operates, the star formation is no longer obscured by the dust, and hence the light from it can propagate freely. This would help explain or clarify whether the mergers with the somewhat spurious 'in-between' type 3 star formation maps (as well as the rest of the emission line maps. See Figure 4.5) are caused by asymmetric star formation, tidal features with enhanced star formation, or dust obscuration in the main components. Such dust-free star formation maps would also illustrate how well (or poorly) obscured line emission trace the underlying total star formation.

4.9.4 COMPREHENSIVE AND DETAILED SAMPLING OF THE SIMULATION'S PARAMETER SPACE

The last improvement of the present study we will mention here concerns the comparison between the observational results and the simulated grism spectra. The simulation parameter space used in the present study is spanned by $(N_t, N_{\rm view}, \Theta, m_{\rm F140W}, z, {\rm SFR})$. This is a significant fraction of the parameters that are expected to influence the appearance of the created star formation maps. However, as stressed, the mass ratio of the mergers as well as the gas fraction of the individual merging components most probably play a key role in the appearance of the obtained star formation maps as well. Expanding the simulations' parameter space with these two quantities will provide a broader more general picture and enable a true direct comparison between observations and simulations, and a more straight forward assessment of the similarities and/or differences of the two. Furthermore, improving the resolution of the individual parameters will also give a more detailed picture (here we only used 2 redshifts, 2 magnitudes, and 2 star formation rates), but also challenges the bookkeeping and simulation power as the total number of simulations quickly grows. For example, sampling each of the 8 parameters just 4 times each would result in $4^8 = 65,536$ simulated spectra to analyze (and visually inspect!), which is unfeasible with the current approach. For such a parameter space automated classification schemes of the mergers and their star formation maps are needed.

4.10 – EPILOGUE –

In this chapter we have presented an initial study of the interplay between star formation and merging of galaxies at the main epoch of galaxy and star formation, with the goal of quantifying the spatial extent of star formation in morphologically disturbed systems at high redshift for the first time. We did this by using the initial 30 pointings of slitless grism spectroscopy from the 3D-HST survey. We were able to create high (spatial) resolution maps of Hα and [OIII] assumed to trace the star formation in 61 merging systems at $z \sim 1.5$. This showed that in ~70% of the systems pronounced star formation was found in only one of the merger components, most often at its center. In roughly 20% of the systems star formation (line emission) was detected in all components. This result indicates that merging at the main epoch of galaxy formation happens more often between objects with different gas fractions, such that the fuel for forming stars and the dust obscuration of the merging systems are significantly different. We searched for correlations between the morphology of the spatial extent of the star formation and redshift, star formation rate, specific star formation rate, and total stellar mass, but found none. This suggests that all star formation morphologies occur at all epochs irrespective of star formation rate and mass.

With the presented merger sample we were able to perform the first comparison between merger simulations and a statistical sample of mergers with the crucial 3D spectroscopic information at high redshift. This initial comparison was done spanning the parameter space of merger inclination, viewing angle, magnitude, star formation rate, and redshift with three simulations of constant mass ratio and gas fraction. This limits the conclusions that can be drawn from the comparison, but directly points towards larger efforts also spanning ranges in mass ratio and gas fraction, as these are also expected to be important players for the star formation in mergers. We found that for the simulated mergers around 60% of the systems show pronounced star formation in both components, whereas 30% only showed star formation in one. This supports the conclusion from the data, that mergers at high redshift does not primarily happen between mergers of the same gas fractions (and mass ratio) as the simulations implicitly assume. This supports the claimed importance of mass ratio and gas fraction on these parameters.

These studies were performed using only 1/4 of the full data product of 3D-HST. The full data product of 3D-HST will be available by the end of 2012 and as described in Section 4.9 stronger conclusions with a larger sample of mergers is straight forwardly obtained by extending the present initial study to the full 3D-HST survey.

CHAPTER 5

SUMMARY & CONCLUSION

In the three main chapters of this thesis we have studied the intrinsic variability of quasars, the most powerful active phase in the evolution of galaxies, and the *spatial* extent of star formation in mergers at $z > 1$, the possible precursors of AGN and quasars. In Chapter 2 we showed that large complete and pure samples of quasars are easily selected via their intrinsic variability by modeling the structure function of each individual quasar as a power-law. Such a selection will become very valuable in the near future, where wide time-domain surveys will provide quasar candidate samples of millions of objects making spectroscopic follow-up unfeasible. Obtaining large samples of quasars is desirable due to their many applications in observational cosmology. In particular we showed in Chapter 3 how samples of quasars with variability information can be used to probe the physics of the quasar central engine. We argued that the observed *color* variability of quasars is too strong to stem solely from changes in the steady state accretion rate of the central engine. We suggested that a stochastic process, like for instance ephemeral hot spots in the accretion disk, could solve this discrepancy. In Chapter 4 we presented the first statistical sample of mergers with three-dimensional spectroscopic information at $z \sim 1.5$. Making maps of the *spatial* extent of star formation (traced by emission lines), we showed that more than 2/3 of these systems have pronounced star formation in only one of the merger components. By comparing with an initial set of simulated spectra, this illustrates that most mergers at $z \sim 1.5$ happen between objects with different gas content.

5.1 CONTRIBUTING TO THE GALAXY EVOLUTION PUZZLE

In this thesis we have exploited two of the most powerful and promising tools at the frontier of galaxy evolution and formation, namely photometric time-domain observations and three-dimensional spectroscopy in the form of slitless grism spectroscopy. The formation and evolution of galaxies is a complicated and very diverse matter, as we believe this thesis, and in particular Chapter 1 and the introductory sections of each of the three main science chapters, Chapters 2–4, illustrates. And that is even considering the fact that these have only mildly scratched the surface of the topic. The areas needed to be covered and dealt with to get a coherent picture of how galaxies formed and evolved are countless. From the description of the cosmological framework of the Universe describing how the first structures formed, over the condensation of gas in dark matter potential wells and the physical mechanisms leading to the formation of stars, to the description and investigation of populations of galaxies at various redshifts to analyze snapshots of the results of billions of years of evolution and star formation, just to mention a few. In order to address these very different and versatile aspects of astronomy and astrophysics the 'galaxy & cosmology community' needs to come together as a whole, and everyone must provide their pieces to the puzzle. With the present study of the intrinsic variability of quasars as a selection tool for finding new quasars and as a probe of the underlying physics mechanisms, as well as the first study of a statistical sample of galaxy mergers with three-dimensional spectroscopy at the peak of cosmic star formation rate density just 3–4 Gyr after the Big Bang, we believe, that we have added significant pieces to the galaxy evolution puzzle, and that these will help understanding how galaxies form, grow and evolve.

Below we have distilled and boiled down the conclusions of the three main science chapters. We have summarized the possibilities this work has lead to and recapped what immediate improvements would strengthen the arguments and results presented, as it was outlined in the individual chapter outlooks (Sections 2.9, 3.8 and 4.9). We will end this thesis in Section 5.6 by taking a peek into the immediate future of 'yours truly'.

5.2 SELECTING QUASARS WITH VARIABILITY

In Chapter 2 we presented how the well established intrinsic variability of quasars can be used as an efficient selection tool in up-coming time-domain surveys like Pan-STARRS 1, LSST and Gaia. As these surveys will provide quasar candidate samples of millions of objects, spectroscopic follow-up to confirm quasars will become unfeasible. Hence, it is crucial to develop methods that can provide candidate samples so pure and complete, that follow-up confirmation is no longer needed. Variability is as illustrated a promising candidate for solving this issue. Our focus was mainly on Pan-STARRS 1, as this survey lacks the quasar UV excess information, which is crucial for the standard color selection methods applied in, e.g., SDSS. Using the largest sample of spectroscopically confirmed quasars with extensive time-domain data, consisting of approximately 9000 quasars from SDSS's Stripe 82, as a testbed, we developed a new variability selection method.

Representing the time-domain data of each individual object by a power-law structure function as described in Section 2.3.2, and referred to as the A–γ selection, we obtained impressive results both with the well sampled Stripe 82 data, ~60 epochs over ~8 years, as well as with mock Pan-STARRS 1 data created by down-sampling the Stripe 82 data to the Pan-STARRS 1 cadence of approximately 6 epochs over 3 years. In particular we obtained quasar candidate samples with estimated completeness and purity above 90% in basically all test cases with a Stripe 82-like time sampling as summarized in Table 2.5 on page 43. Even in the case where we eliminated the previously so crucial UV excess information, we achieved completeness and purity above 90%. In the cases of the 10-times sparser Pan-STARRS 1 sampling, the completeness and purity of our samples drops below 90% in most cases, but in specific regions of color space this variability selection is still 5-10 times more complete than existing color-selections. This illustrates another power of variability selection of quasars: its performance is independent of redshift and photometric (optical) band.

Besides providing promisingly complete and pure quasar samples by very simple means, the A–γ selection confirmed that the fraction of overlooked quasars in Stripe 82 is very modest, i.e., less than 10%. Furthermore, the results suggest that also other objects can be selected convincingly via this simple method, as we were able to cleanly separate 97% of a sample of confirmed RR Lyrae from the non-varying stellar contaminants and the quasars.

The work presented in Chapter 2 therefore clearly illustrates the power of variability selection of large samples of quasars (and possibly other variable objects). As described throughout this thesis, the areas of galaxy evolution where large quasar samples significantly contribute are many. Hence, as time-domain data of significant fractions of the full sky become available these and the coming years, the capabilities of quasar variability as a confident quasar selection method is very promising and comforting.

5.3 THE COLOR VARIABILITY OF QUASARS

In Chapter 3 we gave a specific example of how large statistical quasar samples, in this particular case with variability, can provide direct information about the physics governing quasars and AGN. The central engine of quasars, as described in Section 2.2, consists of a supermassive black hole surrounded by an accretion disk. Whether such a model is the sole power of the observed variability in quasars is still unknown. By exploiting the unprecedented sample of confirmed quasars with variability information from SDSS Stripe 82 also used in Chapter 2, we used the *color* variability to address the question of whether or not the standard steady state accretion disk model is the underlying mechanism giving rise to the observed color variability in quasars. We parametrized the color variability by the quantity $s_{\lambda_1\lambda_2} \equiv \frac{\partial m_{\lambda_2}}{\partial m_{\lambda_1}} - 1$, which we obtained via Metropolis-Hastings Markov chain Monte Carlo fitting in the magnitude-magnitude space of each of the ~9000 quasars in the sample.

We confirm the proposed relation between the magnitude and color of quasars, namely, that quasars become bluer when they get brighter. We observed a distinct redshift dependence in this relation, which we showed was linked to the spectral emission lines exiting/entering the photometric SDSS bands as different redshifts. From simple modeling it turned out that a continuum and emission line variability in phase, with an emission line variability amplitude of 10% or less than the continuum amplitude, was able to reproduce the general form of this redshift dependence. As mentioned in Section 3.8 and below, this offers and illustrates the possibility of performing photometric reverberation mapping.

Testing for any dependencies between the quasar color variability and the mass of the central black hole, the Eddington ratio, the light curve variability amplitude A' (A at quasar rest frame), and the structure function power-law exponent γ, only showed a $s_{\lambda_1\lambda_2}$–A correlation, indicating that quasars with large variability amplitudes have less color variability than quasars with modest variability amplitudes. Hence, the observed color variability is independent of M_{BH} and luminosity of the quasar.

Lastly, we found that the color variability of the individual quasars is on average much stronger than the time averaged ensemble color variability at fixed mass, which resembles the color variability arising from changes in the accretion rate of the central engine. Hence, the study implies that changes in the steady state accretion of the standard quasar engine model cannot explain the observed color variability of quasars. This was confirmed by a comparison of the observations to a set of steady-state accretion disk models.

We speculate that ephemeral hot spots in the accretion disk could produce variability with the observed characteristics, and hence, the work presented in Chapter 3 exemplifies how quasars, and quasar variability in particular, can be used to modify AGN models and add pieces to the general galaxy evolution puzzle.

5.4 STAR FORMATION IN MERGERS AT $z \sim 1.5$

The final science chapter, Chapter 4, was concerned with the spatial extent of star formation in galaxy mergers, the proposed precursor of AGN and quasars, at the main epoch of galaxy and star formation at $z \sim 1.5$. Using the outstanding data of three-dimensional slitless grism spectroscopy and near-infrared imaging from the 3D-HST Hubble treasury program (Section 4.3), we were able to create a statistical sample of galaxy mergers at the peak of cosmic star formation rate density ($z \sim 1.5$) with spatial spectral information for the first time. Making $0.2''$ spatially resolved maps of the (unobscured) star formation, assuming that the star formation is well traced by Hα and [OIII] emission, we categorized the spatial extent and distribution of star formation in this sample of mergers. We found that a diverse morphology of the spatial extent of star formation in mergers at $z > 1$ is present. Nevertheless, in more than 2/3 of the merging systems star formation in just a single merger component was significantly more pronounced than the star formation in the other(s), indicating that merging at these redshifts happen between objects of distinctly different gas fractions and/or star formation rates.

Besides star formation rate, gas fraction and mass ratio, the viewing angle and inclination of the individual merging components are expected to influence the morphology of unobscured star formation. To test this, we used recent N-body SPH simulations with proper radiative transfer and dust prescriptions to create a sample of simulated mock 3D-HST grism spectra resembling the properties of the 61 3D-HST mergers. The estimated star formation morphologies of the simulations were significantly different from the observations: 2/3 of the simulations show star formation in both merger components as opposed to only 21% of the 3D-HST mergers. These initial and illustrative simulations are done on mergers with fixed gas fraction and mass ratio, i.e., the significant difference between them and the 3D-HST mergers supports the conclusion that mergers happen between objects of distinctly different gas fractions and/or star formation rates.

Even though this study was only done on 1/4 of the full 3D-HST data product and the performed simulations are illustrative and point towards more comprehensive studies as pointed out in Section 4.9.4 and noted below, the results of Chapter 4 nicely illustrates the prospects of high resolution spatial studies of star formation in statistical samples of mergers at high redshift.

5.5 OUTLOOK

As is the case for most, if not all science, there is always room for improvement, and results very seldom answer questions without posing and revealing several new interesting perspectives on the subject to be investigated. The work presented here is no exception and also leads to new interesting studies and leaves room for improvement, confirmation and strengthening of the key results and arguments presented.

The variability selection of quasars is an area of research which is heavily investigated in these years, and therefore evolves from one day to the other. Following the results presented in Chapter 2 (Schmidt et al. 2010) a series of papers (MacLeod et al. 2010, 2011; Butler & Bloom 2011; Palanque-Delabrouille et al. 2011; Kim et al. 2011) also presented quasar selection methods based on variability, some of which pointed out limitations, such as computational speed and limited contaminants in the testing phase, of the A–γ selection. This very well illustrates the nature of research, namely, that we stand on each others shoulders to reach higher, as it was already realized in the early days by Newton through Chartres' quote on page 2. Hence, combining all efforts and quasar selection methods should be a main goal for the quasar selection in upcoming surveys. As mentioned in Section 2.9.1, a probabilistic combination might be a way to include all the knowledge gained over the years on how to select quasars in the most efficient and complete way, and thereby exploit all the available photometric as well as time-domain observations.

Besides improving the important selection of quasar samples, the intrinsic variability also offers the possibility of finding even rarer and more glorified objects such as gravitationally lensed quasars. Since quasars are in most cases point sources (especially at higher redshift) using their variability signature can be used to search for unresolved small-separation quasar lenses as described in Section 2.9.2.

Exploiting the variability of quasars in more direct ways to gain further knowledge about AGN and quasars is well illustrated by the results presented in Chapter 3. But also here there is room for improvement. Especially the modeling of the spectral quasar variability's redshift dependence, described in Section 3.5.2, is not giving full credit to the many previous studies of the interplay between the variability of emission lines and the spectral continuum's variations. Despite the presented model's illustrative purpose and capability to actually reproduce the observed redshift trend by very simple means, the model needs to be improved upon. One of the things the model does not take properly into account is the reverberation, i.e., the delay or echo, of the emission line variations to the continuum variations. For example we assumed an instant response, which is know to be pretty far from the truth. Nevertheless, the fact that we actually saw the clear imprint of emission lines on the observed color variability from the broad band photometry of SDSS, illustrates the prospects of performing *photometric* reverberation mapping of large samples of quasars from the very local Universe all the way out to redshift 7. A couple of recent studies have already engaged in developing such mechanisms (Haas et al. 2011; Chelouche & Daniel 2012), and more will most certainly come, as direct estimates of the size and mass of the components of the AGN engine in tens of thousands of objects over a wide redshift range will be a huge improvement to the present knowledge gained from just a few tens of AGN at very low redshift.

In a similar way, though not by as many factors, the sample size of galaxy mergers at $z \sim 1.5$ with three-dimensional spectroscopic information was improved by the study presented in Chapter 4. As mentioned the sample was established using only 1/4 of the 3D-HST survey, which offers the possibility of expanding our study to the full scale 3D-HST data product, which will become available by the end of 2012.

Sample size is however not always the key to more insight. Also thorough and detailed case studies are needed. The 3D-HST mergers serve as obvious follow-up candidates for high resolution IFU spectroscopy. First of all it will provide information about the merging nature of the system, that the relatively low resolution grism spectroscopy of 3D-HST cannot provide, but it will also enable an assessment of the possible AGN contribution to the emission lines via emission line ratios (see Section 3.2.3). Follow-up in radio of these systems would also be extremely valuable. It would be very interesting to see whether the unobscured star formation actually represents and traces the total star formation, and whether the observed star formation morphologies are a consequence of dust obscuration or different star formation rates. Radio follow-up would also be able to quantify, whether or not star formation is coincident with dense molecular clouds at high redshift, as has been found to be the case in the nearby Universe.

Last but not least, the comparison between the 3D-HST mergers and the simulated mock 3D-HST grism spectra is by no means spanning the full parameter space expected to influence the observed star formation maps. As mentioned, the SPH merger simulations used in Section 4.7 all have a mass ratio of 1:2 and are between galaxies all having a gas fraction of 40%. As both gas fraction and mass ratio are expected to influence the appearance of star formation, future comparisons with simulations should sample these parameters as well.

Most of what is mentioned above describes studies that can be initiated immediately, assuming that proposals get excepted, access to the different data repositories can be established, and the time investment needed is available. But also the (near) future is very bright for both galaxy evolution studies in general, as well as quasar variability and high-redshift mergers in particular. With Pan-STARRS 1 (Kaiser et al. 2002) fully operational as of mid 2010 continuously surveying 3/4 of the sky, LSST (Ivezic et al. 2008; LSST Collaboration 2009) steadily moving along in the planning process, Gaia (Perryman et al. 2001) hopefully being launched next year, WFC3 and its slitless grism spectroscopy capabilities fully operational on HST (e.g., Brammer et al. 2012), proposal submission for ALMA[1] cycle 1 coming up, multi-object near-infrared IFUs and spectrographs like KMOS (commissioning in August 2012; Sharples et al. 2004) and MOSFIRE (first light in April 2012; McLean et al. 2010), soon available, and of course JWST (e.g., Gardner et al. 2006) in the more distant future, galaxy formation and evolution studies are entering a very bright era. Many of the possibilities these new instruments, missions and surveys offer will greatly benefit the areas of research presented in this thesis, but also serve the greater good of the community. Mentioning all the possibilities these observational capabilities offer would be far too lengthy for this chapter, so instead we refer to the individual papers and websites, and just mention a couple of very interesting projects (from the point of view of this thesis).

---

[1] http://www.eso.org/sci/facilities/alma/index.html

First and foremost, and almost needless to say, the 24 high resolution near-infrared IFUs and the 7 arcmin diameter field-of-view on KMOS make large follow-up programs of interesting 3D-HST objects, mergers as well as non-mergers, much more efficient than observing one object at the time with existing instruments. This enables statistical studies of star formation histories, merger rates, environmental effects on galaxy evolution, etc. MOSFIRE will aid similar studies with the only difference that spatial spectroscopic information of individual objects will not be available. In the following section we will describe a specific example of the capabilities of MOSFIRE combined with WFC3 selected samples of *very* high redshift galaxies.

We have mentioned that the quasar redshift record at the moment is at 7 (Mortlock et al. 2011b), but as explained in, e.g., Schleicher et al. (2008), ALMA (as well as JWST) will help pushing the detection of quasars towards even higher redshift. This is particularly interesting, since it helps determining the ionization-state of the very early Universe and can put even tighter constraints on the models of AGN engines and the formation of the first galaxies and supermassive black holes and their seeds. For instance if we find a quasar at $z = 10$ we need to figure out how to create $\sim 10^8 M_\odot$ supermassive black holes residing in gas-rich galaxies and ignite the quasar in less than 0.5 Gyr after Big Bang!

These are obviously just a few of the advances in studies of galaxies and their evolution the future will bring. In the following we will end this thesis by looking into the very near future...

5.6 THE IMMEDIATE FUTURE OF 'YOURS TRULY'

In the very near future 'yours truly' will move to the University of California Santa Barbara (UCSB) to be part of Prof. Dr. Tommaso Treu's group. The planned work for this employment is two-fold, but will, as the majority of this thesis, concern the formation and evolution of galaxies at high redshift.

The primary science goal will be to spectroscopically follow-up redshift 8 galaxy candidates selected from the "Brightest of Reionizing Galaxies" (BoRG) survey (Trenti et al. 2011). When the BoRG survey is complete, it will consist of ~310 arcmin of HST WFC3 multi-band (F606W, F098M, F125W and F160W corresponding to V, Y, J, and H) deep imaging with a F125W 5σ limiting AB magnitude of ~26.8 (Bradley et al. 2012). The main purpose of the survey is to find the brightest galaxies at the epoch of reionization in the early Universe, i.e., at $z \gtrsim 7.5$, to gain insight in the reionization epoch and explore the properties of galaxy and star formation at these early epochs. With the four visual/near-infrared bands and the dropout technique, where objects due to their spectral shape fall below the detection threshold in a certain band at a particular redshift and therefore 'disappear' in the observations, a list of potential bright galaxies at $z \gtrsim 7.5$ has been compiled. To confirm that these dropouts are indeed some of the first starbursting galaxies in the Universe, spectroscopic follow-up is needed. The BoRG $z \sim 8$ candidates seem to cluster and form over-densities (Trenti et al. 2012) and are therefore perfect targets for the new near-infrared multi-object spectrograph MOSFIRE on Keck, mentioned above. We have recently submitted the Keck proposal "Probing the physics of reionization with MOSFIRE spectroscopy of $z \sim 8$ galaxy over-densities" (PI T. Treu) to follow-up 15 of the most promising BoRG candidates.

The secondary science goal will be to exploit the scientific environment of UCSB in a continued exploration of some of the results presented in this thesis. Prof. Treu's group has extended experience in gravitational lens searches, hence, the already initiated search for the first variability selected quasar lenses described in Section 2.9.2, will without doubt benefit greatly from this new scientific environment. Investigating the possibility of performing photometric reverberation mapping (Section 3.8.2) will also thrive in Prof. Treu's group, where new interesting studies of spectroscopic reverberation mapping in AGN have been carried out (Pancoast et al. 2011; Brewer et al. 2011). Last but not least, the possible extension of the 3D-HST work described in Chapter 4 will unquestionably take advantage of the near-infrared-spectroscopy expertise and the knowledge about star formation in the early Universe gathered at UCSB.

> *Hence, it is not only the near future of galaxy formation and evolution studies that looks really bright!*

APPENDIX A

INDIVIDUAL STRUCTURE FUNCTION PARAMETER INFERENCE BY MCMC

We use a simple Markov chain Monte Carlo (MCMC) approach (e.g., Metropolis et al. 1953; Hastings 1970; Press et al. 1992; Hansen 2004) to infer the parameters A and γ (Equation (2.8)) and their confidence regions. Our MCMC procedure takes as input a catalog of magnitudes, photometric errors and observed frame MJDs, converted to $N(N-1)/2$ pairs of observations (Equation (2.6)). We then initialize it as follows:

- Pick a starting point for A and γ: we chose 0.1 for both
- Define an initial Gaussian proposal distribution (PD)
- Set an initial temperature β for the chain

The PD width sets the mobility of the Markov chain. Based on several tests we set the initial PD width to 0.05 in both the $\log A$ and γ directions. During a 'burn-in' period at the start of sampling, we draw samples from a modified posterior probability distribution function (PDF) for the parameters, given by the product of the prior PDF, and the likelihood raised to the power of λ. This parameter is an inverse temperature; we start from $\lambda_0 = \frac{1}{\beta_0} = 10^{-3}$. The λ parameter is then increased geometrically to unity as 500 samples are drawn, at which point burn-in is declared over, the inverse temperature is fixed at $\lambda = 1$, and the subsequent samples are stored and used to compute various statistics. For the post burn-in sampling, we use an updated Gaussian proposal distribution, whose widths are set to 10% of the standard deviations of the parameter ($\log A$ and γ) values sampled during burn-in.

At each point in parameter space proposed, we calculate the (un-normalized) log posterior probability distribution of the step, which we define as

$$\log P = \log P(A) + \log P(\gamma) - \lambda \frac{\chi^2}{2} \,. \tag{A.1}$$

Here, χ^2 is related to the logarithm of the likelihood defined in Equation (2.9)

$$\chi^2 = -2 \log \mathcal{L} = \left(\sum_{ij} \log(2\pi V_{\mathrm{eff},ij}^2) + \sum_{ij} \frac{\Delta m_{ij}^2}{V_{\mathrm{eff},ij}^2} \right) \,, \tag{A.2}$$

with the effective variability defined as in Equation (2.11):

$$V_{\mathrm{eff},ij}^2 = V_{\mathrm{mod}}(A, \gamma | \Delta t_{ij})^2 + \delta\Delta m_{ij}^2 = \left(A \Delta t_{ij}^{\gamma} \right)^2 + \delta\Delta m_{ij}^2 \,. \tag{A.3}$$

The sums in Equation (A.2) are over the data pairs derived from the light curve of the given object, and $\delta\Delta m_{i,j}$ is the photometric error on the ij^{th} magnitude pair. In Equation (A.1) $\log P(A)$ and $\log P(\gamma)$ represent the (log) prior PDFs for the parameters A and γ, which we chose to be uninformative. We assigned the following functional forms:

$$P(A) \propto \frac{1}{A} \tag{A.4}$$

$$P(\gamma) \propto \frac{1}{1+\gamma^2} \,. \tag{A.5}$$

If γ is negative or A lies outside the range $[0, 1]$, the log prior density for that parameter is set to -10^{32}, lowering the overall posterior probability for that particular iteration step to effectively zero. In this way we enforce our assumption that the power law exponent is positive and that the average variability on a 1 year timescale is less than 1 magnitude (as found by, e.g., Vanden Berk et al. 2004; Bauer et al. 2009b). Samples are accepted or rejected via the Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970). Care is taken to make the comparison of the log posterior values at constant inverse temperature.

For our final A and γ values, we choose to take the position of the global peak of the likelihood, an approximation of the 'best-fit' point. We approximate this by keeping track of the value of the likelihood as we sample, and then using the sample with the highest value as our estimate. In practice the posterior PDF is not dominated by the prior, such that the peaks of the likelihood and the posterior PDF are usually quite close together.

The uncertainties on the parameters are estimated by considering the 16th and 84th percentiles of the one-dimensional marginalized distributions. In the case of a (symmetric) Gaussian distribution, this would correspond to the 1σ error bar.

To confirm our choice of initial conditions, cooling schedule and PD evolution as sensible, we tested our sampler on simulated light curves for both sine wave and Gaussian white noise sources, and recovered the correct parameters (zero γ and analytically calculated A).

APPENDIX B

FITTING IN COL.-MAG. AND MAG.-MAG. SPACE

The photometric errors of the $(g - r)$ color are correlated with the photometric errors of the g and r band. Thus, when estimating the color variability of objects in general, and quasars in particular, as is done in Chapter 3, care has to be taken that the co-variances of the errors are either removed, taken into account, or avoided. In Chapter 3 we avoided the co-variances by estimating the color relation in magnitude-magnitude space and then 'translated' that into a color variability in color-magnitude space as described in the text. In the following we illustrate the problems one can run into, if the color variability is instead estimated directly in color-magnitude space.

The correlated errors between for instance the g band and the $(g - r)$ color are easily illustrated, by simply drawing a set of random 'observations' from a gaussian distribution with a standard deviation corresponding to the approximate photometric error at the given magnitude. In the left panel of Figure B.1 such a sequence of simulated data is shown. The data have been drawn from 2D gaussian distributions in g and r with mean magnitudes of approximately 18, 19, 20, 21, and 22 and estimated errors of 0.02, 0.025, 0.04, 0.06, and 0.15 respectively. The solid line shows the $g = r$ relation for reference. The color trend (deviation of the data from the $g = r$ line) has been put in to mimic the average color trend of quasars at the given magnitudes. Plotting these simulated observations in color-magnitude space, as done in the right panel of Figure B.1, clearly illustrates the error correlations. The stripy pattern of the color-magnitude diagram is not a consequence of a color change in the object, but a consequence of the fact that the errors in $(g - r)$ are correlated with the errors in g; this is the reason that the 'length' (or artificial color change) of each set of points grows for fainter magnitudes.

In Figure B.2 we show how this effect looks when dealing with real data. Figure B.2 shows the SDSS Stripe 82 photometric data of the quasars SDSS J0320-0051 (left) and

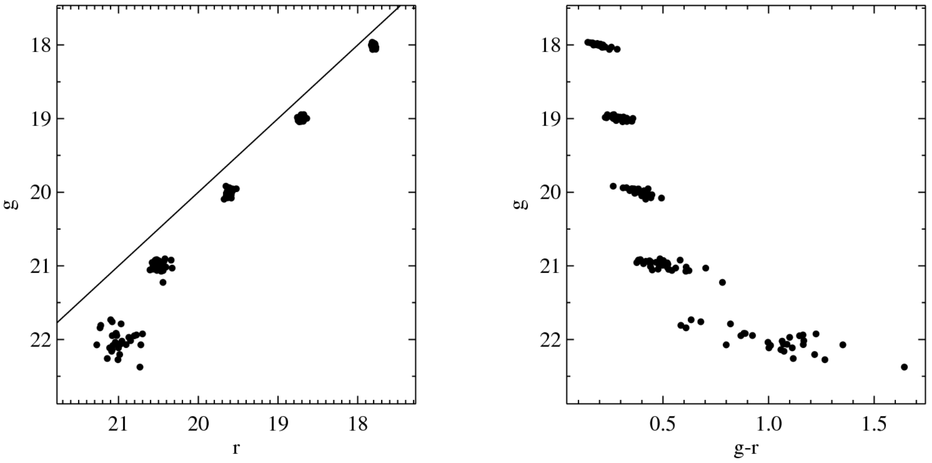

FIGURE B.1: *Simulated data illustrating the effect of error co-variances between magnitude and color, which has not been accounted for in a number of previous analyses. The left panel shows 5 'clouds' of simulated observations in r-g space drawn from 2D gaussian distributions at g = 18, 19, 20, 21 and 22 with 'photometric' errors of 0.02, 0.025, 0.04, 0.06 and 0.15 respectively. In the right panel these data are plotted in the g-$(g-r)$ color-magnitude space. The stripy pattern seen here is due to the correlation between the errors on the g magnitude and the $(g-r)$ color and not a change in color. Our test showed that this effect need to be accounted for when analyzing data of SDSS Stripe 82 quality, as is also illustrated in Figure B.2.*

SDSS J2141-0050 (right) in magnitude-magnitude, magnitude-color and color-magnitude space. The top panel corresponds to the left panel of Figure B.1. The stripy nature of the data when turned into colors is clearly visible in the center and bottom panels of Figure B.2. The data have been color coded according to the observation time. In each panel one solid and two dashed lines are shown. The solid line has been fitted to the shown data, whereas the dashed lines are 'translations' of the fits from the other two spaces. It is clear that in the case of the superb data set of SDSS Stripe 82 the difference between fitting in r-g (top), g-$(g-r)$ (center) and $(g-r)$-g (bottom) space is negligible. However, if one imagines that only data from year 6 and 7 were available for SDSS J2141-0050 (right panel) and the color variability was estimated based on either g-$(g-r)$ or $(g-r)$-g, it is clear that the fit would deviate significantly from the 'real' color relation because of the co-variant errors.

If this effect is not taken into account or avoided when estimating color variability, and the error co-variance 'color change' is interpreted as an actual color change of the quasar, there is a high probability that the results and conclusions will be erroneous. As mentioned in the text color variability has been estimated in color-magnitude space in the past with only a few exceptions. Hence, there might be cases in the literature where this effect has not been taken properly into account and therefore might affect the validity of

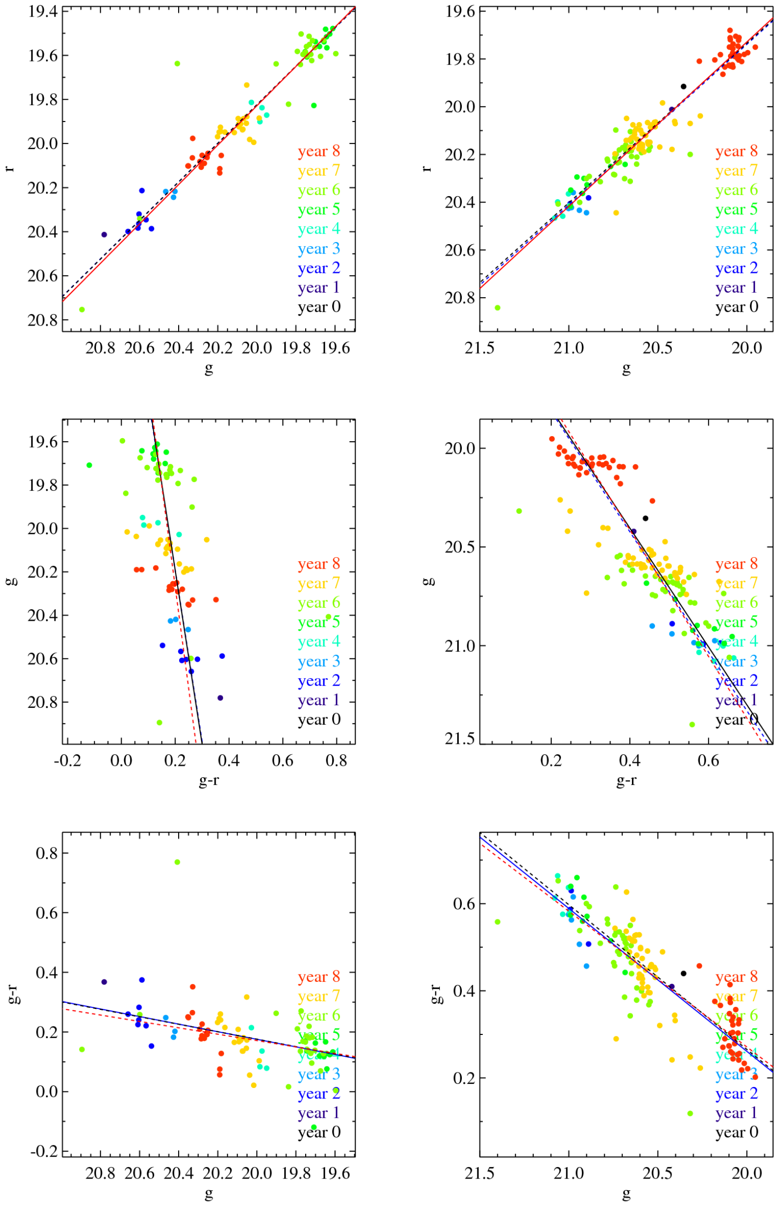

FIGURE B.2: *Comparison between fitting in magnitude-magnitude, magnitude-color and color-magnitude space. The objects SDSS J0320-0051 (left) and SDSS J2141-0050 (right) from the SDSS Stripe 82 sample are shown in the r-g (top), g-$(g-r)$ (center) and $(g-r)$-g (bottom) spaces. The observations are color coded according to the observation year. Each plot has one solid and two dashed lines. The solid line corresponds to the fit (performed as described in Section 3.4) done in the given space, whereas the dashed lines are 'translations' from the two other spaces into the present one. The red line has been fit in r-g space (top), the black line in g-$(g-r)$ (center) and the blue line in $(g-r)$-g (bottom). The stripy-ness in the bottom two panels comes mostly from the correlated errors between g and $(g-r)$ as illustrated in Figure B.1.*

the results. It is hard to quantify how much this effect will affect the results and conclusions made so far in the color variability literature, and we have therefore not made any attempts at quantifying it, but will just note that one needs to take this error correlation into account or, as it is done here, estimate the color variability in magnitude-magnitude or flux-flux space to avoid it.

APPENDIX C

SIMPLISTIC SPECTRAL VARIABILITY MODEL

In the following we describe the simplistic model for the spectral (color) variability of the quasars used in Section 3.5.2. The simple spectral variability model is an attempt to reproduce the observed redshift trends in the mean color variability, $\langle s_{gr/ui} \rangle(z)$. The model is based on the composite SDSS spectrum from Vanden Berk et al. (2001), F_{VdB}. We decomposed the composite spectrum in a line component, F_{line}, and a continuum component, F_{cont}, such that

$$F_{\mathrm{VdB}} = F_{\mathrm{line}} + F_{\mathrm{cont}} \quad . \tag{C.1}$$

The underlying continuum of the composite spectrum is well modeled (for $1150\text{Å} \lesssim \lambda_{\mathrm{rest}} \lesssim 4500\text{Å}$) by a simple power-law with a power-law index of $\beta_\lambda = -1.528$. By fixing the power-law continuum model with a pivot point in the IR (ensuring that $s_{gr/ui} < 0$) we simulate the variable continuum by a (time)sequence of power-laws with different β_λ. Adding fractions of F_{line} to each power-law simulates the (potentially) variable emission lines. The amount of variability in the emission lines can be fixed to the variability of the continuum power-law via a constant α in Equation (3.8), which as a reminder reads

$$\alpha = \frac{\delta F_{\mathrm{line}}}{\delta F_{\mathrm{cont}}} \quad . \tag{3.8}$$

We define

$$\begin{aligned}
\delta F_{\mathrm{line}} &\equiv \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} F_{\mathrm{line}}(\lambda, t_j)\, d\lambda - \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} F_{\mathrm{line}}(\lambda, t_{j-1})\, d\lambda && \text{(C.2)} \\
\delta F_{\mathrm{cont}} &\equiv \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} F_{\mathrm{cont}}(\lambda, t_j)\, d\lambda - \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} F_{\mathrm{cont}}(\lambda, t_{j-1})\, d\lambda \ , && \text{(C.3)}
\end{aligned}$$

where $\lambda_{\mathrm{min}} \sim 1150\text{Å}$ and $\lambda_{\mathrm{max}} \sim 4500\text{Å}$. The t_j and t_{j-1} refers to the 'epochs' of the variability model. This implies from Equation (3.8), that for a fixed α the emission line response for each variability model ‘epoch’ is given by

$$\int_{\lambda_{\min}}^{\lambda_{\max}} F_{\mathrm{line}}(\lambda, t_j)\, d\lambda = \int_{\lambda_{\min}}^{\lambda_{\max}} F_{\mathrm{line}}(\lambda, t_{j-1})\, d\lambda + \alpha \int_{\lambda_{\min}}^{\lambda_{\max}} F_{\mathrm{cont}}(\lambda, t_j) - F_{\mathrm{cont}}(\lambda, t_{j-1})\, d\lambda \quad . \tag{C.4}$$

Here the emission lines are assumed to respond instantly to the continuum variation. By integrating the obtained spectra at the different epochs, t_j, over the SDSS bands the model color variability can be estimated. The results shown in Figure 3.6 are for a variability model with fixed α, since such a variability seems to resemble the observed redshift dependence of the color variability the closest.

The pivot point for the results shown in Figure 3.6 was put at $4\lambda_{\max} \sim 1.8\mu$m. Changing the pivot point slightly changes the amplitude of the obtained model predictions (dashed and colored curves in Figure 3.6) and the mean color variability in such a way that pivot points in the far-IR result in smaller amplitudes and larger $\langle s_{gr/ui}\rangle(z)$. The actual curvature of the curves in Figure 3.6 does not change with the pivot point, i.e., the redshift dependence is independent of the continuum power-law model pivot point.

This spectral variability model predicts the redshift dependence in s_{gr} and s_{ui} equally well.

A model similar to the one presented here, was used in Richards et al. (2001) to explain the redshift dependences of the SDSS quasar colors.

APPENDIX D

THE 3D-HST EMISSION LINE MAPS

In Figures D.1 and D.2 we show the full sample of the 61 3D-HST emission line (star formation) maps. The objects have been sorted according to the estimated morphology of the star formation (SF type) described in Section 4.6, such that Figure D.1 contains all 42 SF type 1 objects, and Figure D.2 contains the 13, 2, and 4 systems with SF type 2, 3, and 4, respectively.

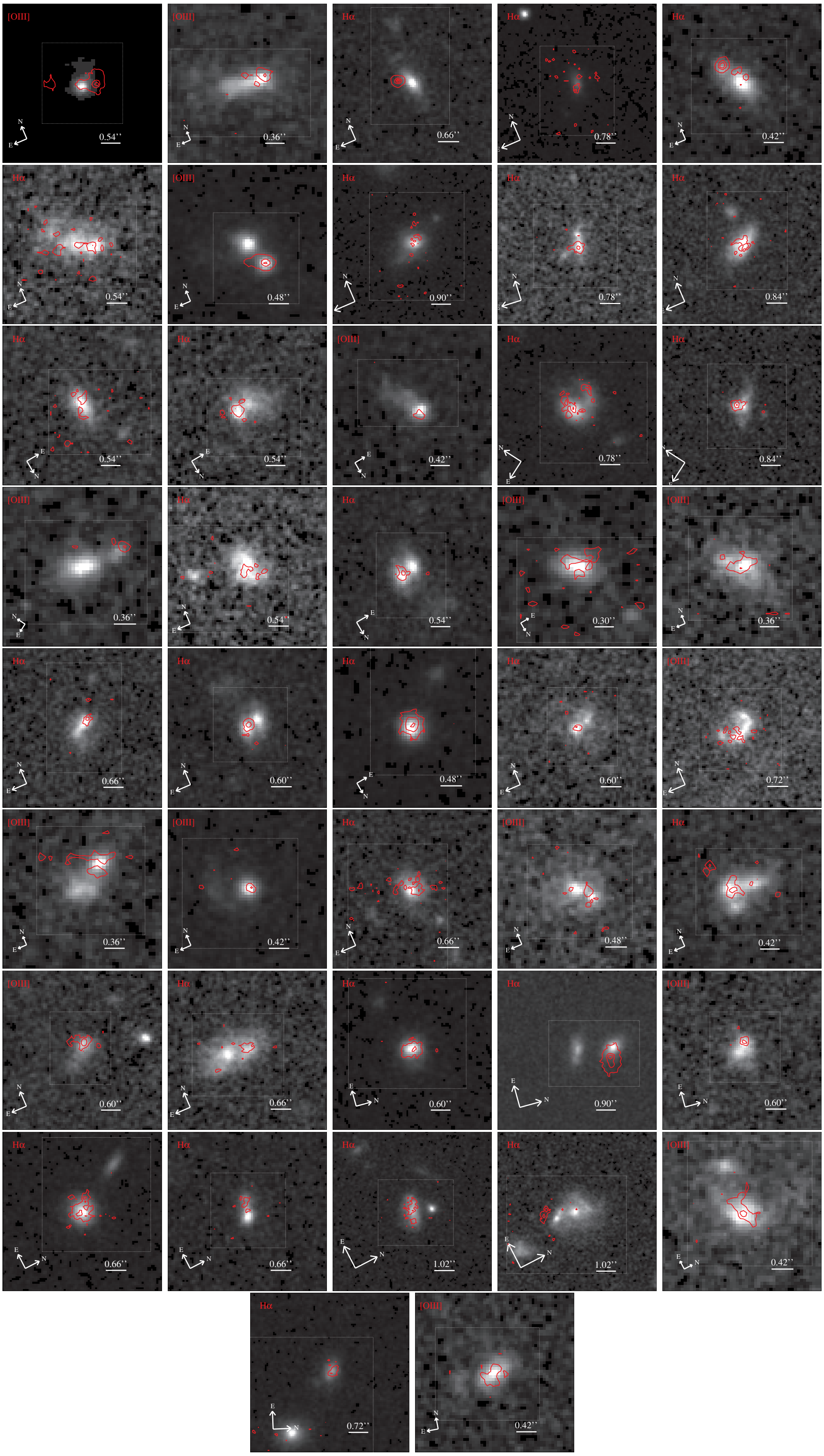
[OIII]
0.54’’
[OIII]
0.36’’
Hα
0.66’’
Hα
0.78’’
Hα
0.42’’
Hα
0.54’’
[OIII]
0.48’’
Hα
0.90’’
Hα
0.78’’
Hα
0.84’’
Hα
0.54’’
Hα
0.54’’
[OIII]
0.42’’
Hα
0.78’’
Hα
0.84’’
[OIII]
0.36’’
Hα
0.54’’
Hα
0.54’’
[OIII]
0.30’’
[OIII]
0.36’’
Hα
0.66’’
Hα
0.60’’
Hα
0.48’’
Hα
0.60’’
[OIII]
0.72’’
[OIII]
0.36’’
[OIII]
0.42’’
Hα
0.66’’
[OIII]
0.48’’
Hα
0.42’’
[OIII]
0.60’’
Hα
0.66’’
Hα
0.60’’
Hα
0.90’’
[OIII]
0.60’’
Hα
0.66’’
Hα
0.66’’
Hα
1.02’’
Hα
1.02’’
[OIII]
0.42’’
Hα
0.72’’
[OIII]
0.42’’
N
E

FIGURE D.1: *The emission line (star formation) maps obtained as described in Section 4.5 of the 42 3D-HST mergers with SF type 1 (‘one component’) star formation morphologies.*

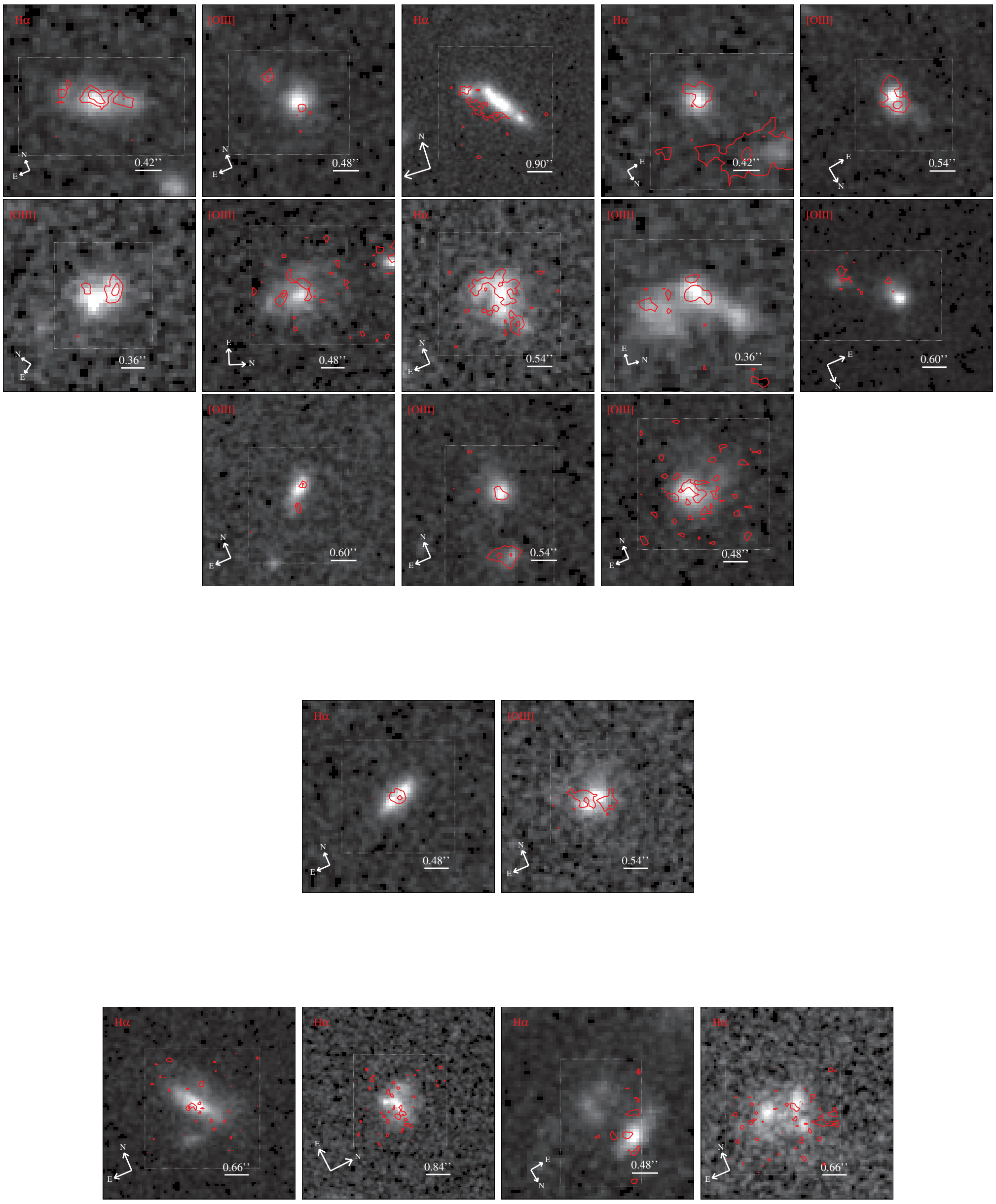
Hα
0.42''
[OIII]
0.48''
Hα
0.90''
Hα
0.42''
[OIII]
0.54''
[OIII]
0.36''
[OIII]
0.48''
Hα
0.54''
[OIII]
0.36''
[OIII]
0.60''
[OIII]
0.60''
[OIII]
0.54''
[OIII]
0.48''
Hα
0.48''
[OIII]
0.54''
Hα
0.66''
Hα
0.84''
Hα
0.48''
Hα
0.66''
N
E

Figure D.2: *The emission line (star formation) maps obtained as described in Section 4.5 of the 13 (top), 2 (center), and 4 (bottom) 3D-HST mergers with SF type 2 ('both (all) components'), SF type 3 ('in-between'), and SF type 4 ('low S/N') star formation morphologies, respectively. The SF type 3 and 4 are mergers also shown in Figure 4.5.*

LIST OF FIGURES

1.1 The ptolemaic (geocentric) and copernican (heliocentric) systems . 3

1.2 The CMB as seen by COBE and WMAP 5

1.3 The Hubble galaxy classes 11

1.4 The unified AGN model 14

2.1 Schematic overview of AGN fueling 23

2.2 SDSS Stripe 82 color space 29

2.3 *gri* light curves of SDSS J203817.37+003029.8 36

2.4 Variability structure functions for SDSS J203817.37+003029.8 . . 38

2.5 SDSS Stripe 82 sample-median structure functions 40

2.6 A–γ space: Stripe 82 test data 41

2.7 A–γ space: mock Pan-STARRS 1 test data 42

2.8 A–γ space: UVX and nUVX selection boxes 45

2.9 A–γ space: Stripe 82 *griz* selection box 48

2.10 A–γ space: mock Pan-STARRS 1 *griz* selection box 49

2.11 Quasar lens candidates selected via variability 56

3.1 Quasar composite spectrum 63

3.2 BPT (emission line ratio) diagram 66

3.3 Example of color variability obtained via MCMC 72

3.4 The color variability s_{gr} and s_{ui} for the 9093 Stripe 82 quasars . . 74

3.5 The role of emission lines in the color variability 75

3.6 Comparison of simple spectral color variability model with data . . 77

3.7 Color variability in gr vs. ui space; s_{gr} vs. s_{ui} 78

3.8 Color variability as a function of Eddington ratio and $M_{\rm BH}$ 80

3.9 2D histogram of Eddington ratio and $M_{\rm BH}$ 81

3.10 Color variability as a function of A' and γ 83

3.11 Color variability as a function of A' and γ – cleaned version . . . 84

3.12 The ensemble vs. individual object quasar color variability . . . 87

3.13 Color variability compared with accretion disk models – SDSS J2037-0112 88

3.14 Color variability compared with accretion disk models – ensemble . 90

4.1 The 61 3D-HST mergers in the selection space 111

4.2 The $m_{\rm F140W}$ magnitude distribution of the 61 3D-HST mergers . . 112

4.3 Redshift comparison plots. $z_{\rm grism}$ vs. $z_{\rm phot}$ and $z_{\rm grism}$ vs. $z_{\rm ELmap}$. . . 113

4.4 Illustration of emission line map creation 114

4.5 Examples of emission line (star formation) maps 117

4.6 Comparison of 3D-HST pre- and post-merger star formation types . 119

4.7 Examples of simulated 3D-HST grism spectra 123

4.8 Comparison of 3D-HST and simulated star formation types . . . 125

B.1 Mock data illustrating co-variances in magnitude and color . . . 146

B.2 Color variability determined in mag.-mag. and mag.-color space . . 147

D.1 The 42 SF type 1 3D-HST Merger's Emission Line Maps 152
D.2 The SF type 2, 3, and 4 3D-HST Merger's Emission Line Maps . . . 153

LIST OF TABLES

1.1 The main AGN types and their characteristics 13

2.1 The UV excess color box 34

2.2 The non-UV excess color box 34

2.3 The *griz* color box 35

2.4 Pan-STARRS 1 survey schedule 37

2.5 Completeness and purity of variability slections 43

4.1 The merger selection cuts 108

4.2 The spatial extent of star formation – data and simulations 120

4.3 The 3D input data-cubes for 3D-HST grism simulations 121

ACKNOWLEDGEMENTS

Thesis acknowledgements are a funny concept. Is it really necessary to put ones gratitude down on paper? I hope that people around me generally know when I appreciate their company, their help and guidance in scientific as well as personal matters, their support in my doings, and their ability to tell me when I'm off track. But that's probably just me having a crazy idea (again), and it does seem to be good practice to write them, so here we go:

First and foremost it is a pleasure to acknowledge everything my thesis advisor Hans-Walter Rix has done for me through the last 3.5 years. I'm truly grateful to him for letting me exploit and be part of his extensive scientific network without having to worry (too much) about the politics involved. His insights and support in my scientific, as well as personal life, have greatly improved my science and the quality of life. Thanks!

Secondly, I am indebted to Phil Marshall who showed me how science is suppose to be: a part of, but not your entire life – "*Earn your beers*". I deeply appreciate the times we have spent together chatting about life and science and am looking forward to continued collaboration.

I would like to thank the many collaborators that helped getting our papers into shape and encouraged, helped, and inspired me to become a better scientist on a daily basis. Thank you Bovy, Brammer, Cox, da Cunha, Dobler, Fan, Hennawi, Hogg, Jester, Jonsson, Knecht, Maoz, Richards, Shields, and van Dokkum.

Also the 3D-HST team needs to be mentioned. I acknowledge them for letting me be part of this truly magnificent survey, even though I haven't been able to put all the time I wish I had into it. I hope, I can stay in the perimeter of the collaboration, despite the fact that I'm moving on to new adventures on the west coast.

Without the support of the Marie Curie Initial Training Network, ELIXIR, none of this would have happened, so needless to say I feel privileged to have been part of this network. A special thank to my fellow ELIXIR comrades, Bellocchi, Cano, Caruana, Caz-

zoli, Chang, Chevallard, Dorner, Lorenzoni, Mosleh, Pacifici, Rosdahl, and Troncoso. I have always looked forward to the meetings and am looking forward to run into you at meetings and conferences in the future – see you around.

Thanks to the IMPRS crew for letting me ‘quereinsteig’ and become part of what was truly an *extra*ordinary generation – we set the bar high, mates! I am thankful, that I managed to find such good friends down here both within and outside IMPRS. A special thank goes to Gs, MG, MC, and BV. It has been a great pleasure sharing these years with you. I appreciate that you never managed to make me crawl (intentionally – I think not). On the contrary, I’m sad to admit that E-town probably won. But I do believe, however, that Rohrbach gave you a good fight. Thanks for making me a little more Italian, a tad more Chilean and a pinch more Indian – I needed that.

I’m indebted to my parents for always supporting me in whatever I do. You made me realize that I could take this path, even though it hasn’t always been easy. Thanks for always being there when needed. The same applies to you, Martin. I *always* appreciate and worship the time we spend together. I’m really grateful you visited as often as you did – epic days to remember for sure. I owe you (at least) one. Thanks Bro – at ease.

Last, but certainly not least: Mette, you make me a better person. Thanks for sticking around and supporting me from afar. I can’t wait for our next adventure – together!

I hope UCSB turns out to be as big a lottery win as MPIA was.

This work was funded primarily by the Marie Curie Initial Training Network ELIXIR of the European Commission under contract PITN-GA-2008-214227, and was carried out at the Max Planck Institute for Astronomy, Heidelberg, as a member of the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD), Germany.

This work is partially based on observations taken by the 3D-HST Treasury Program with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

BIBLIOGRAPHY

Abazajian, K. N., Adelman-Mccarthy, J. K., Agüeros, M. A., et al. 2009, The Astrophysical Journal Supplement, 182, 543

Almeida, C. R., Levenson, N. A., Alonso-Herrero, A., et al. 2011, The Astrophysical Journal, 731, 92

Antonucci, R. 1993, In: Annual review of astronomy and astrophysics. Vol. 31 (A94-12726 02-90), 31, 473

Aretxaga, I., Fernandes, R. C., & Terlevich, R. J. 1997, Monthly Notices RAS, 286, 271, (c) 1997 The Royal Astronomical Society

Atlee, D. W., & Gould, A. 2007, The Astrophysical Journal, 664, 53

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, Astronomical Society of the Pacific, 93, 5, a&AA ID. AAA029.158.269

Bandara, K., Crampton, D., & Simard, L. 2009, The Astrophysical Journal, 704, 1135

Barnes, J. E., & Hernquist, L. 1996, Astrophysical Journal v.471, 471, 115

Barton, E. J., Arnold, J. A., Zentner, A. R., Bullock, J. S., & Wechsler, R. H. 2007, The Astrophysical Journal, 671, 1538

Barton, E. J., Geller, M. J., & Kenyon, S. J. 2000, The Astrophysical Journal, 530, 660

Bauer, A., Baltay, C., Coppi, P., et al. 2009a, The Astrophysical Journal, 699, 1732

—. 2009b, The Astrophysical Journal, 696, 1241

Bennett, C. L., Banday, A. J., Gorski, K. M., et al. 1996, Astrophysical Journal Letters v.464, 464, L1

Bennett, C. L., Halpern, M., Hinshaw, G., et al. 2003, The Astrophysical Journal Supplement Series, 148, 1

Bertin, E., & Arnouts, S. 1996, Astronomy and Astrophysics Supplement, 117, 393

Bhatti, W. A., Richmond, M. W., Ford, H. C., & Petro, L. D. 2010, The Astrophysical Journal Supplement, 186, 233

Bigiel, F., Leroy, A., Walter, F., et al. 2008, The Astronomical Journal, 136, 2846

Binney, J., & Tremaine, S. 2008, Galactic Dynamics: Second Edition

Blandford, R. D., & McKee, C. F. 1982, Astrophysical Journal, 255, 419, a&AA ID. AAA031.158.074

Blitz, L. 1993, In: Protostars and planets III (A93-42937 17-90), 125

Blitz, L., Fukui, Y., Kawamura, A., et al. 2007, Protostars and Planets V, 81

Bloom, J. S., Perley, D. A., Li, W., et al. 2009, The Astrophysical Journal, 691, 723

Bondi, H. 1947, Monthly Notices RAS, 107, 410

Bonning, E. W., Cheng, L., Shields, G. A., Salviander, S., & Gebhardt, K. 2007, The Astrophysical Journal, 659, 211

Bovy, J., Hennawi, J. F., Hogg, D. W., et al. 2011, The Astrophysical Journal, 729, 141

Boyle, B. J., Shanks, T., Croom, S. M., et al. 2000, Monthly Notices RAS, 317, 1014, (c) 2000 The Royal Astronomical Society

Bradač, M., Schneider, P., Steinmetz, M., et al. 2002, 6th European VLBI Network Symposium on New Developments in VLBI Science and Technology, 203

Bradley, L. D., Trenti, M., Oesch, P. A., et al. 2012, eprint arXiv, 1204, 3641

Bramich, D. M., Vidrih, S., Wyrzykowski, L., et al. 2008, Monthly Notices RAS, 386, 887

Brammer, G., van Dokkum, P., Franx, M., et al. 2012, eprint arXiv, 1204, 2829

Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, The Astrophysical Journal, 686, 1503

Brewer, B. J., Treu, T., Pancoast, A., et al. 2011, The Astrophysical Journal Letters, 733, L33

Bridge, C. R., Appleton, P. N., Conselice, C. J., et al. 2007, The Astrophysical Journal, 659, 931

Bromm, V., Coppi, P. S., & Larson, R. B. 1999, The Astrophysical Journal, 527, L5

Burbidge, G., & Napier, W. M. 2009, The Astrophysical Journal, 706, 657

Butler, N. R., & Bloom, J. S. 2011, The Astronomical Journal, 141, 93

Calzetti, D., Liu, G., & Koda, J. 2012, eprint arXiv, 1204, 5659

Canalizo, G., & Stockton, A. 2001, The Astrophysical Journal, 555, 719

Chambers, K. C., & Denneau, L. J. 2008

Chelouche, D., & Daniel, E. 2012, The Astrophysical Journal, 747, 62, 14 pages, 12 figures

Chiu, K., Richards, G. T., Hewett, P. C., & Maddox, N. 2007, Monthly Notices RAS, 375, 1180

Choloniewski, J. 1981, ACTA ASTRONOMICA V. 31, 31, 293, a&AA ID. AAA031.158.140

Cisternas, M., Jahnke, K., Inskip, K. J., et al. 2011, The Astrophysical Journal, 726, 57

Colless, M., Dalton, G., Maddox, S., et al. 2001, Monthly Notices RAS, 328, 1039

Collier, S., & Peterson, B. 2001, The Astrophysical Journal, 555, 775

Conselice, C. J., Rajgor, S., & Myers, R. 2008, Monthly Notices RAS, 386, 909

Cowie, L. L., Songaila, A., Kim, T.-S., & Hu, E. M. 1995, Astronomical Journal (ISSN 0004-6256), 109, 1522

Cox, T. J. 2004, Ph.D dissertation, 27

Cox, T. J., Jonsson, P., Primack, J. R., & Somerville, R. S. 2006, Monthly Notices RAS, 373, 1013

Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R., & Dekel, A. 2008, Monthly Notices RAS, 384, 386, (c) Journal compilation © 2008 RAS

Cristiani, S., Trentini, S., Franca, F. L., et al. 1996, Astronomy and Astrophysics, 306, 395

Croom, S. M., Smith, R. J., Boyle, B. J., et al. 2001, Monthly Notices RAS, 322, L29

Croom, S. M., Boyle, B. J., Shanks, T., et al. 2005, Monthly Notices RAS, 356, 415

Croom, S. M., Richards, G. T., Shanks, T., et al. 2009a, Monthly Notices RAS, 399, 1755

—. 2009b, Monthly Notices RAS, 392, 19, (c) Journal compilation © 2009 RAS

da Cunha, E., Charlot, S., & Elbaz, D. 2008, Monthly Notices RAS, 388, 1595

D'abrusco, R., Longo, G., & Walton, N. 2009, Monthly Notices RAS, 396, 223

Dalal, N., & Kochanek, C. S. 2002, The Astrophysical Journal, 572, 25, (c) 2002: The American

Astronomical Society
Davis, S. W., Woo, J.-H., & Blaes, O. M. 2007, The Astrophysical Journal, 668, 682
de Sitter, W. 1917, Koninklijke Nederlandsche Akademie van Wetenschappen Proceedings, 19, 1217
—. 1918, Koninklijke Nederlandsche Akademie van Wetenschappen Proceedings, 20, 229
de Vries, W. H., Becker, R. H., White, R. L., & Loomis, C. 2005, The Astronomical Journal, 129, 615, (c) 2005: The American Astronomical Society
Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T. 1965, Astrophysical Journal, 142, 414
Dobler, G., & Keeton, C. R. 2006, The Astrophysical Journal, 653, 1391
D'Odorico, V., Petitjean, P., & Cristiani, S. 2002, Astronomy and Astrophysics, 390, 13
Doroshkevich, A. G., Sunyaev, R. A., & Zeldovich, I. B. 1974, In: Confrontation of cosmological theories with observational data; Proceedings of the Symposium, 63, 213, a&AA ID. AAA012.162.028
Dreyer, J. L. E. 1888, Mem. R. Astron. Soc., 49, 1, nGC
Dunlop, J. S., McLure, R. J., Kukula, M. J., et al. 2003, Monthly Notice of the Royal Astronomical Society, 340, 1095
Efstathiou, G., & Silk, J. 1983, Fundamentals of Cosmic Physics (ISSN 0094-5846), 9, 1
Einstein, A. 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 844
—. 1916, Annalen der Physik, 354, 769
—. 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 142
Elvis, M. 2000, The Astrophysical Journal, 545, 63
Eyer, L. 2002, Acta Astronomica, 52, 241
Faber, S. M. 1984, Large-Scale Structure of the Universe, 187
Fan, X. 2006, New Astronomy Reviews, 50, 665, elsevier B.V.
Fan, X., Strauss, M. A., Schneider, D. P., et al. 2001a, The Astronomical Journal, 121, 54
Fan, X., Narayanan, V. K., Lupton, R. H., et al. 2001b, The Astronomical Journal, 122, 2833
Ferguson, H. C., Dickinson, M., & Williams, R. 2000, Annu. Rev. Astro. Astrophys., 38, 667
Fernandes, R. C., Terlevich, R., & Aretxaga, I. 1997, Monthly Notices RAS, 289, 318
Ferrarese, L., & Merritt, D. 2000, The Astrophysical Journal, 539, L9
Förster Schreiber, N. M., Genzel, R., Bouché, N., et al. 2009, The Astrophysical Journal, 706, 1364
Freedman, R., & Kaufmann, W. J. 2002, Universe, iSBN: 0-7167-4647-6
Friedmann, A. 1922, Zeitschrift für Physik, 10, 377
—. 1924, Zeitschrift für Physik, 21, 326
Frieman, J. A., Bassett, B., Becker, A., et al. 2008, The Astronomical Journal, 135, 338
Fumagalli, M., Franx, M., Brammer, G. B., et al. 2012, In preparation
Gallagher, J. S., Hunter, D. A., & Bushouse, H. 1989, Astronomical Journal (ISSN 0004-6256), 97, 700
Gallagher, J. S., Hunter, D. A., & Tutukov, A. V. 1984, Astrophysical Journal, 284, 544
Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, Space Science Reviews, 123, 485
Garnavich, P. M., Kirshner, R. P., Challis, P., et al. 1998, Astrophysical Journal Letters v.493, 493, L53

Gavazzi, G., Boselli, A., Pedotti, P., Gallazzi, A., & Carrasco, L. 2002, Astronomy and Astrophysics, 396, 449

Gebhardt, K., Bender, R., Bower, G., et al. 2000, The Astrophysical Journal, 539, L13

Geha, M., Alcock, C., Allsman, R. A., et al. 2003, The Astronomical Journal, 125, 1

Gehren, T., Fried, J., Wehinger, P. A., & Wyckoff, S. 1984, Astrophysical Journal, 278, 11

Georgakakis, A., Coil, A. L., Laird, E. S., et al. 2009, Monthly Notices RAS, 397, 623

Ghez, A. M., Salim, S., Weinberg, N. N., et al. 2008, The Astrophysical Journal, 689, 1044

Giannantonio, T., Scranton, R., Crittenden, R. G., et al. 2008, Physical Review D, 77, 123520

Giveon, U., Maoz, D., Kaspi, S., Netzer, H., & Smith, P. S. 1999, Monthly Notices RAS, 306, 637

Gnedin, N. Y. 1998, Monthly Notices RAS, 299, 392

Gnedin, N. Y., & Hui, L. 1998, Monthly Notices RAS, 296, 44

Graham, A. W., Erwin, P., Caon, N., & Trujillo, I. 2001, The Astrophysical Journal, 563, L11

Gupta, A. C., & Joshi, U. C. 2005, Astronomy and Astrophysics, 440, 855

Guth, A. H. 1981, Physical Review D (Particles and Fields), 23, 347, a&AA ID. AAA029.162.101

Haas, M., Chini, R., Ramolla, M., et al. 2011, Astronomy & Astrophysics, 535, 73

Haiman, Z., & Hui, L. 2001, The Astrophysical Journal, 547, 27

Hammer, F., Flores, H., Elbaz, D., et al. 2005, Astronomy and Astrophysics, 430, 115

Hansen, S. H. 2004, Monthly Notices RAS, 351, L5

Häring, N., & Rix, H.-W. 2004, The Astrophysical Journal, 604, L89

Harrison, E. R. 1970, Physical Review D, 1, 2726

Hastings, W. K. 1970, Biometrika, 57, 97

Hawkins, M. R. S. 1996, Monthly Notices RAS, 278, 787, (c) 1996 The Royal Astronomical Society

Heavens, A., Panter, B., Jimenez, R., & Dunlop, J. 2004, Nature, 428, 625

Heckman, T. M., Bothun, G. D., Balick, B., & Smith, E. P. 1984, Astronomical Journal (ISSN 0004-6256), 89, 958

Hennawi, J. F., & Prochaska, J. X. 2007, The Astrophysical Journal, 655, 735

Hennawi, J. F., Strauss, M. A., Oguri, M., et al. 2006a, The Astronomical Journal, 131, 1

Hennawi, J. F., Prochaska, J. X., Burles, S., et al. 2006b, The Astrophysical Journal, 651, 61

Hennawi, J. F., Myers, A. D., Shen, Y., et al. 2010, The Astrophysical Journal, 719, 1672

Herschel, W. 1786, Philosophical Transactions of the Royal Society of London, 76, 457

Hewett, P. C., Warren, S. J., Leggett, S. K., & Hodgkin, S. T. 2006, Monthly Notices RAS, 367, 454

Hinshaw, G., Weiland, J. L., Hill, R. S., et al. 2009, The Astrophysical Journal Supplement, 180, 225

Hippelein, H., Maier, C., Meisenheimer, K., et al. 2003, Astronomy and Astrophysics, 402, 65

Hogg, D. W., Bovy, J., & Lang, D. 2010, arXiv, 1008, 4686, a chapter from a non-existent book

Hook, I. M., McMahon, R. G., Boyle, B. J., & Irwin, M. J. 1994, R.A.S. MONTHLY NOTICES V.268, 268, 305

Hopkins, A. M., & Beacom, J. F. 2006, The Astrophysical Journal, 651, 142

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2005a, The Astrophysical Journal, 630, 705, (c) 2005: The American Astronomical Society

—. 2006, The Astrophysical Journal Supplement Series, 163, 1, (c) 2006: The American Astronomical Society

Hopkins, P. F., Hernquist, L., Martini, P., et al. 2005b, The Astrophysical Journal, 625, L71, (c) 2005: The American Astronomical Society

Hopkins, P. F., Younger, J. D., Hayward, C. C., Narayanan, D., & Hernquist, L. 2010, Monthly Notices RAS, 402, 1693

Hoyle, F., & Tayler, R. J. 1964, Nature, 203, 1108

Hubble, E. 1929, Proceedings of the National Academy of Sciences of the United States of America, 15, 168

Hubble, E. P. 1925a, The Observatory, 48, 139

—. 1925b, Astrophysical Journal, 62, 409

—. 1926, Astrophysical Journal, 64, 321

Hutchings, J. B., Johnson, I., & Pyke, R. 1988, Astrophysical Journal Supplement Series (ISSN 0067-0049), 66, 361

Hutchings, J. B., & Neff, S. G. 1992, Astronomical Journal (ISSN 0004-6256), 104, 1

Inada, N., Oguri, M., Shin, M.-S., et al. 2010, The Astronomical Journal, 140, 403

Ivezic, Z., Tyson, J. A., Allsman, R., et al. 2008, eprint arXiv, 0805, 2366

Ivezić, Ž., Smith, J. A., Miknaitis, G., et al. 2007, The Astronomical Journal, 134, 973

Jahnke, K., & Maccio, A. 2010, arXiv, astro-ph.CO, submitted to ApJL

Jahnke, K., Bongiorno, A., Brusa, M., et al. 2009, The Astrophysical Journal Letters, 706, L215

Jarosik, N., Barnes, C., Greason, M. R., et al. 2007, The Astrophysical Journal Supplement Series, 170, 263

Jarosik, N., Bennett, C. L., Dunkley, J., et al. 2011, The Astrophysical Journal Supplement, 192, 14

Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London. Series A, 199, 1

Jogee, S. 2006, Physics of Active Galactic Nuclei at all Scales, 693, 143

Jogee, S., Miller, S. H., Penner, K., et al. 2009, The Astrophysical Journal, 697, 1971

Jonsson, P. 2006, Monthly Notices RAS, 372, 2

Jonsson, P., Groves, B. A., & Cox, T. J. 2010, Monthly Notices RAS, 403, 17, 32 pages, 22 figures, submitted to MNRAS

Joseph, R. D., & Wright, G. S. 1985, Monthly Notices of the Royal Astronomical Society (ISSN 0035-8711), 214, 87

Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, Survey and Other Telescope Technologies and Discoveries. Edited by Tyson, 4836, 154

Karim, A., Schinnerer, E., Martínez-Sansigre, A., et al. 2011, The Astrophysical Journal, 730, 61

Kaspi, S., Brandt, W. N., Maoz, D., et al. 2007, The Astrophysical Journal, 659, 997

Kaspi, S., Maoz, D., Netzer, H., et al. 2005, The Astrophysical Journal, 629, 61

Kaspi, S., Smith, P. S., Netzer, H., et al. 2000, The Astrophysical Journal, 533, 631

Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, Monthly Notices RAS, 346, 1055

Kawaguchi, T., Mineshige, S., Umemura, M., & Turner, E. L. 1998, Astrophysical Journal v.504, 504, 671, (c) 1998: The American Astronomical Society

Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, The Astrophysical Journal, 698, 895

Kelly, B. C., Sobolewska, M., & Siemiginowska, A. 2010, eprint arXiv, 1009, 6011

Kennicutt, R. C. 1983, Astrophysical Journal, 272, 54

—. 1992, Astrophysical Journal, 388, 310

—. 1998a, Astrophysical Journal v.498, 498, 541

—. 1998b, Annu. Rev. Astro. Astrophys., 36, 189

Kennicutt, R. C., Tamblyn, P., & Congdon, C. E. 1994, Astrophysical Journal, 435, 22

Kepler, J. 1609, (Pragae) 1609, 32.

Kewley, L. J., Geller, M. J., & Jansen, R. A. 2004, The Astronomical Journal, 127, 2002

Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, Monthly Notices RAS, 372, 961

Kewley, L. J., Heisler, C. A., Dopita, M. A., & Lumsden, S. 2001, The Astrophysical Journal Supplement Series, 132, 37

Kim, D.-W., Protopapas, P., Byun, Y.-I., et al. 2011, The Astrophysical Journal, 735, 68

Kirkpatrick, J. A., Schlegel, D. J., Ross, N. P., et al. 2011, The Astrophysical Journal, 743, 125, 9 Pages, 8 Figures

Klaas, U., & Elsaesser, H. 1991, Astronomy and Astrophysics Supplement Series (ISSN 0365-0138), 90, 33

Kocevski, D. D., Faber, S. M., Mozena, M., et al. 2012, The Astrophysical Journal, 744, 148

Kolb, E. W., & Turner, M. S. 1990, Front. Phys.

Kollmeier, J. A., Onken, C. A., Kochanek, C. S., et al. 2006, The Astrophysical Journal, 648, 128

Kormendy, J., & Kennicutt, R. C. 2004, Annual Review of Astronomy &Astrophysics, 42, 603

Kormendy, J., & Richstone, D. 1995, Annu. Rev. Astro. Astrophys., 33, 581

Kozłowski, S., Kochanek, C. S., Udalski, A., et al. 2010, The Astrophysical Journal, 708, 927, submitted to ApJ, 42 pages, 14 figures, 2 tables

Krolik, J. H. 1999, Active galactic nuclei : from the central black hole to the galactic environment /Julian H. Krolik. Princeton

Kümmel, M., Walsh, J. R., Pirzkal, N., Kuntschner, H., & Pasquali, A. 2009, PUBL ASTRON SOC PAC, 121, 59

Lacey, C., & Cole, S. 1993, Monthly Notices of the Royal Astronomical Society (ISSN 0035-8711), 262, 627

Lambas, D. G., Tissera, P. B., Alonso, M. S., & Coldwell, G. 2003, Monthly Notices RAS, 346, 1189

Law, D. R., Steidel, C. C., Erb, D. K., et al. 2007, The Astrophysical Journal, 669, 929

Lemaître, G. 1931, Monthly Notices RAS, 91, 483

—. 1933, Annales de la Société Scientifique de Bruxelles A53, 53, 51

Lifshitz, E. M. 1946, Journal of Physics USSR, 10, 116

Lin, L., Koo, D. C., Weiner, B. J., et al. 2007, The Astrophysical Journal, 660, L51

Linde, A. D. 1982, Physics Letters B, 108, 389

Longair, M. S. 1998, Galaxy formation / Malcolm S. Longair. New York : Springer

Lopez, S., Barrientos, L. F., Lira, P., et al. 2008, The Astrophysical Journal, 679, 1144

LSST Collaboration. 2009, arXiv, 0912, 0201

Lynden-Bell, D., & Wood, R. 1968, Monthly Notices RAS, 138, 495

Macciò, A. V. 2008, Dark Galaxies and Lost Baryons, 244, 186, (c) 2008: International Astronomical Union

MacLeod, C., Ivezić, Ž., de Vries, W., Sesar, B., & Becker, A. 2008, CLASSIFICATION AND DISCOVERY IN LARGE ASTRONOMICAL SURVEYS: Proceedings of the International Conference: “Classification and Discovery in Large Astronomical Surveys”. AIP Conference Proceedings, 1082, 282

MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, The Astrophysical Journal, 721, 1014

MacLeod, C. L., Brooks, K., Ivezić, Ž., et al. 2011, The Astrophysical Journal, 728, 26, 47 pages, submitted to ApJ

Madau, P., & Rees, M. J. 2001, The Astrophysical Journal, 551, L27

Maddox, N., Hewett, P. C., Warren, S. J., & Croom, S. M. 2008, Monthly Notices RAS, 386, 1605

Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, The Astronomical Journal, 115, 2285

Marconi, A., & Hunt, L. K. 2003, The Astrophysical Journal, 589, L21

Martini, P., & Weinberg, D. H. 2001, The Astrophysical Journal, 547, 12

Mather, J. C., Cheng, E. S., Eplee, R. E., et al. 1990, Astrophysical Journal, 354, L37

Matteo, P. D., Combes, F., Melchior, A.-L., & Semelin, B. 2007, Astronomy and Astrophysics, 468, 61

Matteo, T. D., Springel, V., & Hernquist, L. 2005, Nature, 433, 604

McLean, I. S., Steidel, C. C., Epps, H., et al. 2010, Ground-based and Airborne Instrumentation for Astronomy III. Edited by McLean, 7735, 47, (c) 2010: American Institute of Physics

McLure, R. J., & Dunlop, J. S. 2002, Monthly Notices RAS, 331, 795

Melnick, J., & Mirabel, I. F. 1990, Astronomy and Astrophysics (ISSN 0004-6361), 231, L19

Messier, C., & Niles, P. H. 1981, Clifton Park, 6413, lCCN: 80-70586 (BKS3); CALL NUMBER: QB6 .M6413

Metcalf, R. B. 2005, The Astrophysical Journal, 629, 673, (c) 2005: The American Astronomical Society

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, Journal of Chemical Physics, 21, 1087

Meusinger, H., Hinze, A., & de Hoon, A. 2011, Astronomy and Astrophysics, 525, 37

Mihos, J. C., & Hernquist, L. 1996, Astrophysical Journal v.464, 464, 641

Mirabel, I. F., Vigroux, L., Charmandaris, V., et al. 1998, Astronomy and Astrophysics, 333, L1

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution. Cambridge University Press, cambridge University Press

Morgan, J. S., Burgett, W., & Teran, J. U. 2008, Ground-based and Airborne Telescopes II. Edited by Stepp, 7012, 95

Mortlock, D. J., Patel, M., Warren, S. J., et al. 2011a, eprint arXiv, 1101, 4965, submitted to MNRAS; 20 pages, 13 figures

Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011b, Nature, 474, 616

Myers, A. D., Brunner, R. J., Nichol, R. C., et al. 2007, The Astrophysical Journal, 658, 85

Myers, A. D., Richards, G. T., Brunner, R. J., et al. 2008, The Astrophysical Journal, 678, 635

Myers, A. D., White, M., & Ball, N. M. 2009, Monthly Notices RAS, 399, 2279

Narayan, R., Yi, I., & Mahadevan, R. 1995, Nature, 374, 623

Nelson, E. J., van Dokkum, P. G., Brammer, G., et al. 2012, The Astrophysical Journal Letters, 747, L28

Netzer, H., & Peterson, B. M. 1997, Astronomical Time Series, 218, 85

Newton, I. 1760, Colonia

Oguri, M., & Marshall, P. J. 2010, Monthly Notices RAS, 405, 2579

Oguri, M., Inada, N., Strauss, M. A., et al. 2008, The Astronomical Journal, 135, 512

Oort, J. H. 1932, Bulletin of the Astronomical Institutes of the Netherlands, 6, 249

Padmanabhan, N., Schlegel, D. J., Finkbeiner, D. P., et al. 2008, The Astrophysical Journal, 674, 1217

Palanque-Delabrouille, N., Yeche, C., Myers, A. D., et al. 2011, Astronomy & Astrophysics, 530, 122, 11 pages, 17 figures, submitted to A&A

Pancoast, A., Brewer, B. J., & Treu, T. 2011, The Astrophysical Journal, 730, 139

Peebles, P. J. E. 1968, Astrophysical Journal, 153, 1

—. 1971, Princeton Series in Physics

—. 1982, Astrophysical Journal, 258, 415, a&AA ID. AAA032.162.001

Peng, C. Y. 2007, The Astrophysical Journal, 671, 1098

Peng, C. Y., Impey, C. D., Rix, H.-W., et al. 2006a, New Astronomy Reviews, 50, 689

—. 2006b, The Astrophysical Journal, 649, 616

Penzias, A. A., & Wilson, R. W. 1965, Astrophysical Journal, 142, 419

Pereyra, N. A., Vanden Berk, D. E., Turnshek, D. A., et al. 2006, The Astrophysical Journal, 642, 87, (c) 2006: The American Astronomical Society

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, The Astrophysical Journal, 517, 565

Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, Astronomy and Astrophysics, 369, 339

Peterson, B. M. 1993, Astronomical Society of the Pacific, 105, 247

—. 2010, Co-Evolution of Central Black Holes and Galaxies, 267, 151

Peterson, B. M., Wanders, I., Bertram, R., et al. 1998, Astrophysical Journal v.501, 501, 82

Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, The Astrophysical Journal, 613, 682

Pogge, R. W., & Eskridge, P. B. 1987, Astronomical Journal (ISSN 0004-6256), 93, 291

Poindexter, S., Morgan, N., & Kochanek, C. S. 2008, The Astrophysical Journal, 673, 34

Pollock, J. T., Pica, A. J., Smith, A. G., et al. 1979, Astronomical Journal, 84, 1658, a&AA ID. AAA026.141.113

Press, W. H., & Schechter, P. 1974, Astrophysical Journal, 187, 425, a&AA ID. AAA011.162.012

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Cambridge: University Press

Primack, J. R., & Blumenthal, G. R. 1984, IN: Formation and evolution of galaxies and large structures in the universe; Proceedings of the Third Moriond Astrophysics Meeting, 163

Rasmussen, C. E., & Williams, C. K. I. 2006, MIT Press, 2006

Rauch, M. 1998, Annu. Rev. Astro. Astrophys., 36, 267

Rees, M. J. 1984, IN: Annual review of astronomy and astrophysics. Volume 22. Palo Alto, 22, 471

Rengstorf, A. W., Brunner, R. J., & Wilhite, B. C. 2006, The Astronomical Journal, 131, 1923

Rengstorf, A. W., Mufson, S. L., Andrews, P., et al. 2004, The Astrophysical Journal, 617, 184, (c) 2004: The American Astronomical Society

Richards, G. T., Vanden Berk, D. E., Reichard, T. A., et al. 2002a, The Astronomical Journal, 124, 1

Richards, G. T., Fan, X., Schneider, D. P., et al. 2001, The Astronomical Journal, 121, 2308

Richards, G. T., Fan, X., Newberg, H. J., et al. 2002b, The Astronomical Journal, 123, 2945

Richards, G. T., Nichol, R. C., Gray, A. G., et al. 2004, The Astrophysical Journal Supplement Series, 155, 257

Richards, G. T., Croom, S. M., Anderson, S. F., et al. 2005, Monthly Notices RAS, 360, 839, (c) 2005 RAS

Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, The Astronomical Journal, 131, 2766, (c) 2006: The American Astronomical Society

Richards, G. T., Myers, A. D., Gray, A. G., et al. 2009, The Astrophysical Journal Supplement, 180, 67

Riechers, D. A., Walter, F., Carilli, C. L., & Bertoldi, F. 2007a, The Astrophysical Journal, 671, L13

Riechers, D. A., Walter, F., Cox, P., et al. 2007b, The Astrophysical Journal, 666, 778

Rieke, G. H., Cutri, R. M., Black, J. H., et al. 1985, Astrophysical Journal, 290, 116

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, The Astronomical Journal, 116, 1009

Robaina, A. R., Bell, E. F., Wel, A. V. D., et al. 2010, The Astrophysical Journal, 719, 844

Robaina, A. R., Bell, E. F., Skelton, R. E., et al. 2009, The Astrophysical Journal, 704, 324

Rosa, G. D., Decarli, R., Walter, F., et al. 2011, The Astrophysical Journal, 739, 56

Ross, N. P., Shen, Y., Strauss, M. A., et al. 2009, The Astrophysical Journal, 697, 1634

Ryden, B. 2003, Introduction to cosmology / Barbara Ryden. San Francisco, iSBN: 0-8053-8912-1 Addison Wesley

Sakata, Y., Morokuma, T., Minezaki, T., et al. 2011, The Astrophysical Journal, 731, 50

Sakata, Y., Minezaki, T., Yoshii, Y., et al. 2010, The Astrophysical Journal, 711, 461

Sanders, D. B., Soifer, B. T., Elias, J. H., et al. 1988, Astrophysical Journal, 325, 74

Schawinski, K., Treister, E., Urry, C. M., et al. 2011, The Astrophysical Journal Letters, 727, L31

Schleicher, D. R. G., Spaans, M., & Klessen, R. S. 2008, eprint arXiv, 0812, 3950

Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., et al. 1998, The Astrophysical Journal, 507, 46

Schmidt, K. B., Marshall, P. J., Rix, H.-W., et al. 2010, The Astrophysical Journal, 714, 1194

Schmidt, K. B., Rix, H.-W., Shields, J. C., et al. 2012a, The Astrophysical Journal, 744, 147

Schmidt, K. B., Rix, H.-W., Cunha, E. D., et al. 2012b, In preparation

Schmidt, M. 1959, Astrophysical Journal, 129, 243

Schmidt, M., & Green, R. F. 1983, Astrophysical Journal, 269, 352

Schneider, D. P., Hall, P. B., Richards, G. T., et al. 2007, The Astronomical Journal, 134, 102

Schneider, P., Kochanek, C., & Wambsganss, J. 2006, 33rd Saas-Fee Advanced Course, Gravitational lensing: strong, weak and micro

Schneider, R., Ferrara, A., Natarajan, P., & Omukai, K. 2002, The Astrophysical Journal, 571, 30

Schödel, R., Ott, T., Genzel, R., et al. 2002, Nature, 419, 694

Scholz, R.-D., Meusinger, H., & Irwin, M. 1997, Astronomy and Astrophysics, 325, 457

Schramm, K.-J., Borgeest, U., Kuehl, D., et al. 1994, Astronomy and Astrophysics Suppl. 106, 106, 349

Scranton, R., Ménard, B., Richards, G. T., et al. 2005, The Astrophysical Journal, 633, 589

Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, The Astronomical Journal, 134, 2236

Sesar, B., Ivezić, Ž., Grammer, S. H., et al. 2010, The Astrophysical Journal, 708, 717

Seyfert, C. K. 1943, Astrophysical Journal, 97, 28

Shakura, N. I., & Sunyaev, R. A. 1973, Astron. Astrophys., 24, 337, a&AA ID. AAA009.066.049

Shapiro, K. L., Genzel, R., Förster Schreiber, N. M., et al. 2008, The Astrophysical Journal, 682, 231

Sharples, R. M., Bender, R., Lehnert, M. D., et al. 2004, Ground-based Instrumentation for Astronomy. Edited by Alan F. M. Moorwood and Iye Masanori. Proceedings of the SPIE, 5492, 1179

Shaver, P. A., Wall, J. V., Kellermann, K. I., Jackson, C. A., & Hawkins, M. R. S. 1996, Nature, 384, 439

Shectman, S. A., Landy, S. D., Oemler, A., et al. 1996, Astrophysical Journal v.470, 470, 172

Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, The Astrophysical Journal, 680, 169

Shen, Y., & Liu, X. 2012, eprint arXiv, 1203, 601, 17 emulateapj pages; submitted to ApJ

Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, The Astronomical Journal, 133, 2222, (c) 2007: The American Astronomical Society

Shen, Y., Strauss, M. A., Ross, N. P., et al. 2009, The Astrophysical Journal, 697, 1656

Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, The Astrophysical Journal Supplement, 194, 45

Shlosman, I., Begelman, M. C., & Frank, J. 1990, Nature (ISSN 0028-0836), 345, 679

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1991, Astrophysical Journal, 371, L1

—. 1992, Astrophysical Journal, 396, L1

Soszynski, I., Udalski, A., Szymanski, M., et al. 2003, Acta Astronomica, 53, 93, (c) 2003: Acta Astronomica

Spitzer, L., & Saslaw, W. C. 1966, Astrophysical Journal, 143, 400

Spitzer, L., & Stone, M. E. 1967, Astrophysical Journal, 147, 519

Springel, V. 2000, Monthly Notices RAS, 312, 859

Springel, V., Yoshida, N., & White, S. D. M. 2001, New Astronomy, 6, 79

Stalin, C. S., Gopal-Krishna, Sagar, R., & Wiita, P. J. 2004, Monthly Notices RAS, 350, 175

Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, The Astronomical Journal, 123, 485

Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, The Astronomical Journal, 124, 1810

Suganuma, M., Yoshii, Y., Kobayashi, Y., et al. 2006, The Astrophysical Journal, 639, 46

Sullivan, M., & Collaboration, T. S. L. S. 2005, 1604-2004: Supernovae as Cosmological Lighthouses, 342, 466

Sumi, T., Woźniak, P. R., Eyer, L., et al. 2005, Monthly Notices RAS, 356, 331

Swinbank, A. M., Smail, I., Chapman, S. C., et al. 2004, The Astrophysical Journal, 617, 64

Teplitz, H. I., Malkan, M. A., Steidel, C. C., et al. 2000, The Astrophysical Journal, 542, 18

Terlevich, R., Tenorio-Tagle, G., Franco, J., & Melnick, J. 1992, Royal Astronomical Society, 255, 713

Tolman, R. C. 1934, Proceedings of the National Academy of Sciences of the United States of America, 20, 169

Toomre, A., & Toomre, J. 1972, Astrophysical Journal, 178, 623, a&AA ID. AAA008.151.039

Tran, H. D. 1995, Astrophysical Journal v.440, 440, 578

Tremaine, S., Gebhardt, K., Bender, R., et al. 2002, The Astrophysical Journal, 574, 740

Trenti, M., Bradley, L. D., Stiavelli, M., et al. 2011, The Astrophysical Journal Letters, 727, L39

—. 2012, The Astrophysical Journal, 746, 55

Trèvese, D., Kron, R. G., & Bunone, A. 2001, The Astrophysical Journal, 551, 103

Trèvese, D., & Vagnetti, F. 2002, The Astrophysical Journal, 564, 624

Trouille, L., Barger, A. J., & Tremonti, C. 2011, The Astrophysical Journal, 742, 46

Trump, J. R., Weiner, B. J., Scarlata, C., et al. 2011, The Astrophysical Journal, 743, 144, apJ accepted. 8 pages, 6 figures

Udalski, A., Kubiak, M., & Szymanski, M. 1997, Acta Astronomica, 47, 319, (c) 1997: Acta Astronomica

Ulrich, M.-H., Maraschi, L., & Urry, C. M. 1997, Annu. Rev. Astro. Astrophys., 35, 445

Urry, C. M., & Padovani, P. 1995, PUBL ASTRON SOC PAC, 107, 803

van Dokkum, P. G., Brammer, G., Fumagalli, M., et al. 2011, The Astrophysical Journal Letters, 743, L15, submitted to ApJ Letters

Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, The Astronomical Journal, 122, 549

Vanden Berk, D. E., Wilhite, B. C., Kron, R. G., et al. 2004, The Astrophysical Journal, 601, 692

Vestergaard, M. 2002, The Astrophysical Journal, 571, 733

Vestergaard, M., & Peterson, B. M. 2006, The Astrophysical Journal, 641, 689

Wang, Z., Fazio, G. G., Ashby, M. L. N., et al. 2004, The Astrophysical Journal Supplement Series, 154, 193

Webb, J. R., Smith, A. G., Leacock, R. J., et al. 1988, Astronomical Journal (ISSN 0004-6256), 95, 374

Webb, W., & Malkan, M. 2000, The Astrophysical Journal, 540, 652

Whitaker, K. E., Labbe, I., van Dokkum, P. G., et al. 2011, eprint arXiv, 1105, 4609

White, S. D. M., & Rees, M. J. 1978, Monthly Notices RAS, 183, 341, a&AA ID. AAA021.162.070

Wilhite, B. C., Brunner, R. J., Grier, C. J., Schneider, D. P., & Vanden Berk, D. E. 2008, Monthly Notices RAS, 383, 1232

Wilhite, B. C., Vanden Berk, D. E., Kron, R. G., et al. 2005, The Astrophysical Journal, 633, 638

Williams, J. P., Blitz, L., & McKee, C. F. 2000, Protostars and Planets IV (Book - Tucson: University of Arizona Press; eds Mannings, 97

Willott, C. J., Albert, L., Arzoumanian, D., et al. 2010, The Astronomical Journal, astro-ph.CO, 546, 16 pages, 10 figures, AJ in press

Winkler, H. 1997, Royal Astronomical Society, 292, 273

Woo, J.-H., & Urry, C. M. 2002, The Astrophysical Journal, 579, 530

Wright, S. A., Larkin, J. E., Law, D. R., et al. 2009, The Astrophysical Journal, 699, 421

Wu, X.-B., & Jia, Z. 2010, Monthly Notices RAS, 406, 1583

Wu, X.-B., Wang, R., Schmidt, K. B., et al. 2011, The Astronomical Journal, 142, 78

Wuyts, S., Labbé, I., Förster Schreiber, N. M., et al. 2008, The Astrophysical Journal, 682, 985

Xia, J.-Q., Viel, M., Baccigalupi, C., & Matarrese, S. 2009, Journal of Cosmology and Astroparticle Physics, 09, 003

Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, The Astronomical Journal, 137, 4377

Yun, M. S., Scoville, N. Z., Carrasco, J. J., & Blandford, R. D. 1997, Astrophysical Journal Letters v.479, 479, L9, (c) 1997: The American Astronomical Society

Zackrisson, E., Bergvall, N., Marquart, T., & Helbig, P. 2003, Astronomy and Astrophysics, 408, 17

Zeldovich, Y. B. 1972, Monthly Notices RAS, 160, 1P, a&AA ID. AAA008.162.030

Zel'Dovich, Y. B., & Novikov, I. D. 1967, Soviet Astronomy, 10, 602

Zu, Y., Kochanek, C. S., & Peterson, B. M. 2011, The Astrophysical Journal, 735, 80

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

INDEX

3D-HST, 105

A'–γ, 81–83
A–γ, 30, 40–42, 45, 48, 49, 51, 80, 141
active galactic nucleus, *see* AGN
adiabatic structure formation, 6
AGN, 12
    accretion disk, 15
    central engine, 15, 22
    dust torus, 15
    fueling, 22, 23
    jet, 15

baryon asymmetry problem, 8
Big Bang model, 3
BL Lac galaxy, 12–14
black hole seed, 22
Blazar, 12–14
bolometric luminosity, 65
bottom-up model, 6, 7
BPT diagram, 65, 66, 127
brighter-bluer, 69, 70, 74
broad line radio galaxy, BLRG, 12
broad line region, BLR, 15, 64

central engine, *see* AGN
cold mode gas accretion, 99
color variability, 71
    s_{gr}, 73
    s_{ui}, 77
    model, 75, 149
completeness, c, 43
cosmic microwave background, CMB, 4, 5
cosmological constant, 4, 9
cosmological principle, 5, 8

damped Lyα systems, DLA, 63
dark matter, 7
    cold, CDM, 7
    hot, 7
    warm, 7

Eddington accretion rate, $\dot{M}_{\mathrm{Edd}}$, 24
Eddington luminosity, L_{Edd}, 24
Eddington ratio, $L_{\mathrm{bol}}/L_{\mathrm{Edd}}$, 26, 64
emission line map, 112, 151
emission line ratio, 65

F/G stars, 33
filaments, 6, 7
flatness problem, 8

GADGET, 103
geocentric, 2, 3
giant molecular cloud, 99
grand unified theory, 8
grism spectrum, 106
    contamination `FCONTAM`, 106
    coverage, `FCOVER`, 106
    simulated, 120
griz box, 34, 47

heliocentric, 2, 3

Herschel catalogue, 3
hierarchical merging, 10, 98
hierarchical Universe, 7
horizon problem, 8
hot mode gas accretion, 99
Hubble
  class, 10, 11
  constant, H_0, 3
  expansion, 4, 9
  law, 3
  time, t_0, 4

inflation, 8
isothermal structure formation, 6

Jeans
  length, λ_J, 6
  mass, M_J, 99

large scale structure problem, 8
lensed quasar, 55
LINER, 12
LSST, 26, 51
Lyα forest, 63
Lyman limit systems, LLS, 63

MACHO, 7
Markov chain Monte Carlo, MCMC, 31, 141
match radius, R_{match}, 107
merger rate, 11
Messier catalogue, 3
monolithic collapse, 10, 98
monopole problem, 8

narrow line radio galaxy, NLRG, 12
narrow line region, NLR, 15, 64
near-infrared inflection, 64
new general catalogue, NGC, 3
non-UV excess, nUVX box, 34, 45

optically violently variable galaxy, OVV, 12–14

Pan-STARRS 1, 26, 28
Pan-STARRS 4, 26
Press-Schechter theory, 7
probability distribution function, PDF, 141
purity, p, 43, 44

quasar, 12–14, 16
  color selection, 27
  peak of density, 16
  spectrum, 62, 63
  variability timescales, 67
  volume density, 13

Radio galaxy, 13, 14
  volume density, 13
reverberation mapping, 67
  photometric, 93
RR Lyrae, 33, 82

SDSS Stripe 82, 31
SED model, 109
Seyfert galaxy, 12–14
  volume density, 13
simulation data-cubes, 103, 120
SINS sample, 102
star formation
  efficiency, 100
  peak of, 11
stellar locus, 28, 29
structure function, 28, 30, 38, 40, 82
  ensemble, 30
  model, 30, 80
SUNRISE, 103
super-Eddington accretion, 24
supermassive black hole, SMBH, 12, 15

top-down model, 6
transfer function, 68

unified AGN model, 13, 14
unified quasar selection, 55
UV excess, UVX, 26
  box, 33, 44

virial black hole mass, 65
voids, 6, 7

WIMP, 7