Constraining Cosmic String Parameters Using Large-Scale Structure

Yuki Omori

Master of Science

Department of Physics

McGill University

Montréal, Québec

2013-08-15

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

ACKNOWLEDGEMENTS

First and foremost I wish to acknowledge the the efforts and contributions of both my supervisors Robert Brandenberger and Tracy Webb, whose guidance and discussions have proved invaluable to my graduate experience. I also wish to thank Darren Croton for kindly allowing me to use SAGE, which without, this thesis would be incomplete. In addition, I would like to thank Luke Hodkinson for assisting me in converting ROCKSTAR halo catalogues to SUBFIND format, which SAGE is compatible with. I also acknowledge the use of the publicly available N-body code GADGET-2 written by Volker Springel, halo finding code ROCKSTAR by Peter Behroozi and simulation remapping algorithm by Jordan Carlson and Martin White. Furthermore, I gratefully acknowledge the centre for High Performance Computing (HPC) for providing computing time on supercomputers Guillimin and Colosse. One person who deserves special mention is Jeff Cooke, who provided me great help, both in terms of science and mental support. Also deserving of mention are the graduate students residing in the ERP-225, who provided enlightening scientific discussions (most of the times). Finally, I wish to acknowledge the support of family and friends who provided me support to complete this thesis.

ABSTRACT

Cosmic strings are topological defects that appear in many different early universe theories. By constraining the tension $G\mu$ of cosmic strings, the energy scale of the early universe could be probed and various theoretical models could be ruled out. In this study, we use large-scale structure to test its effectiveness in constraining $G\mu$. Using simulations and observational data, we find that the differences in large-scale structure in a universe with $G\mu = 0$, $G\mu = 10^{-5}$, $G\mu = 10^{-7}$ are small. We further analyze structure formed by galaxies with various luminosities and star formation rates. However, we find that the uncertainties in the measurements dominate the signal, and hence, no strict constraint could be placed on the cosmic string tension.

ABRÉGÉ

Les cordes cosmiques sont des défauts topologiques qui apparaissent dans de nombreuses théories de l'univers primordial. En plaçant des limites sur les valeurs de la tension $G\mu$ des cordes cosmiques, on peut explorer l'échelle d'énergie de l'univers primordial et certains modèles théoriques peuvent ainsi être rejetés. Jusqu'à maintenant, la valeur de $G\mu$ a été mesurée grâce aux observations du fond diffus cosmologique et grâce au chronométrage des pulsars. Dans cette étude, nous utilisons les structures à grande échelle de l'univers pour tester leur capacité à contraindre $G\mu$. À l'aide de simulations et de données observationelles, nous observons que les différences sont faibles dans les structures à grande échelle entre des univers avec $G\mu = 0$, $G\mu = 10^{-5}$ et $G\mu = 10^{-7}$. Nous étudions ensuite la structure formée par des sous-classes de galaxies caractérisées par différentes luminosités et différents taux de formation stellaire. Néanmoins, nous concluons que les incertitudes dominent le signal et qu'aucune contrainte stricte ne peut être placée sur la tension des cordes cosmiques.

TABLE OF CONTENTS

LIST OF TABLES

Table page

LIST OF FIGURES

Figure page

CHAPTER 1
INTRODUCTION

1.1 Modern Cosmology

Our understanding of the early universe has dramatically improved in the past century due to experiments such as Planck, WMAP, SPT and ACT. By observing the cosmic microwave background (CMB), many cosmological parameters have been constrained including the amplitude of the linear power spectrum at 8 Mpc/h σ_8, tilt of the initial power spectrum n_s and the energy densities $\Omega_{m,r,\Lambda}$ [Larson et al., 2011]. These CMB observations combined with Baryonic Acoustic Oscillation (BAO) observations and type Ia supernovae observations allow us to constrain the 6-parameter ΛCDM model [Komatsu et al., 2011]. One of the primary goals of these CMB experiments is to measure the angular power spectrum of the temperature fluctuations on the surface of last scattering. The angular power spectrum describes the fluctuation amplitudes of different modes in the universe at the time of recombination.

Until the theory of inflation was proposed by Alan Guth, the origin of these density fluctuations was not well understood due to three main problems: the flatness problem, horizon problem and the structure formation problem. The flatness problem arises from the fact that the density of our universe today is very close to the critical density. This implies that shortly after the Big Bang, the density must have been coincident with the critical density to extreme precision, which leads to the fine tuning problem of the initial state.

The horizon problem arises from observing the CMB to be isotropic. This would not be the case unless the universe was extremely isotropic to begin with. In the standard Big Bang theory, this is not possible due to the finite speed of light and the distance which thermal equilibrium could extend to at the epoch of decoupling [Ryden, 2003].

The structure formation problem is concerned with the origin of the large-scale structure in the universe. Structures such as galaxy clusters are coherent over a mass of about $10^{15} M_{\odot}$, and are thought to have cumulatively increased in size from small initial perturbations to their current size through gravitational collapse. To explain its coherence within the object, the entire mass must have been inside the particle horizon. This requires the existence of perturbations that seeded the structures we observe today at $z < 10^6$. This poses the question of origin of density perturbations at times before the time of equal matter and radiation. However, at these early times, the scale of these large objects were larger than the particle horizon [Mo et al., 2010].

Inflation provides solutions to the aforementioned problems by postulating a period of rapid, exponential growth driven by a scalar field to occur shortly after the Big Bang. With this expansion, the density of the universe rapidly approaches the critical density and expands a small region of the universe, which achieved thermal equilibrium to a size much greater than the present horizon. This also expands primordial quantum fluctuations to cosmological scales and forms the initial infrastructure for large scale structures [Benson, 2000].

The small fluctuations then gravitationally accrete mass (both dark matter and gas) and virialize to form halos. The properties of these gaseous halos are determined

by gravitational, dynamical and radiative processes. Gas contained in a halo loses its energy via radiative cooling, and the halos that cool effectively form galaxies. The critical mass for effective cooling is very similar to that of massive galaxies observed today, and thus it is thought that there is a tight relation between the effectiveness of cooling and the resulting mass of a galaxy [Mo et al., 2010].

During the epoch when the majority of proto-galaxies formed, the universe mostly only contained the basic elements namely hydrogen, helium and dark matter. The most massive and cool gas clouds within the proto-galaxies then started to collapse to produce the first stars [Bromm and Yoshida, 2011]. At this time, galaxies evolved quickly through accretion of smaller mass galaxies. Galaxy clusters situated at locations with high underlying dark matter density continuously grew and formed nodes on the cosmic web. Other over dense regions also ended up as filamentary structures though gravitational collapse connecting the nodes. These large scale structures are the outcome of the growing initial density perturbation in the universe. Hence, the distribution, or the clustering of galaxies provides hints in understanding the structure of the early universe [Aragón-Calvo et al., 2010].

1.2 Topological defects

In the very early stages, matter in the universe possessed a large degree of gauge symmetry when it was in a plasma state. As the temperature decreased, a succession of phase transitions occurred and these symmetries were broken. Consider the simplest case of having a complex scalar field $\phi(x)$ described by the Lagrangian

density

$$\mathcal{L} = \frac{1}{2}\partial_\mu \phi^* \partial^\mu \phi + \frac{1}{2}m^2|\phi|^2 - \frac{\lambda}{4}|\phi|^4 \tag{1.1}$$

where λ and m are real constants. At finite temperatures the quantity to minimize becomes the free energy of the system. This alters the zero temperature potential to an effective potential:

$$V_{\mathrm{T}} = -\frac{1}{2}m^2(T)\phi^2 + \frac{\lambda}{4}\phi^4 \tag{1.2}$$

where the mass $m^2 = m^2(1 - \frac{T^2}{T_c^2})$ and T_{c} is the critical temperature above which the symmetry is restored. At temperatures lower than T_{c}, the potential will develop local minima and ϕ will fall into these minima. The actual value of ϕ in the final state will depend on the random fluctuation in ϕ, which will be different at different spatial locations. Small regions of the universe will begin their transition and form domains of broken phase. The dimensions of these domains are determined by the temperature dependence of correlation length of the ϕ^4 potential. At temperatures below the Ginsburg temperature, T_{G}, a temperature slightly smaller than the critical temperature, thermal fluctuations are insufficient to change the minima and the defects are effectively "frozen out". At that point, the correlation length is microscopic. On scales larger than the correlation length, there is an $\mathcal{O}(1)$ probability to form a defect. Hence, in a particle physics model which admit defect solution, such defect *will* form in the early universe [Vilenkin and Shellard, 2000]. This is known as the Kibble mechanism.

Domain walls and high energy scale local monopoles are ruled out, since they would over-close the universe. Other topological defects such as cosmic strings, global

monopoles and textures are free of such consequences and are interesting for cosmology. The effects of these topological defects has been extensively studied. From these past studies it can be concluded that topological defects cannot be the dominant source for fluctuations [Vilenkin and Shellard, 2000],
[Planck collaboration XXV et al., 2013]. Since many particle physics models predict defects, it is interesting to search for the cosmological signatures in spite of the small contribution to structure formation.

1.3 Cosmic strings

Cosmic strings arise when we consider a complex scalar field ϕ and a potential of the form:

$$V(\phi) = \frac{\lambda}{4}(\phi^*\phi - \eta^2)^2, \tag{1.3}$$

where η is the minimum of the potential. Such type of potential is rotationally symmetric (i.e. U(1))

$$\phi(x) \to e^{i\alpha}\phi(x). \tag{1.4}$$

A particular ground state is not unique under such a transformation:

$$\langle 0|\phi|0\rangle = \eta e^{i\theta} \to \eta e^{i\theta+\alpha}. \tag{1.5}$$

This lack of symmetry for the ground state is known as symmetry breaking. Since causally disconnected regions of space cannot have the same value of the scalar field on the vacuum manifold, the scalar field along a path in space traces out a path on the manifold in position space. With probability $\mathcal{O}(1)$ on horizon scales, a path will

correspond to a full revolution in the vacuum manifold. Since the potential $V(\phi)$ is higher at $\phi = 0$, there exists a line of trapped energy, which is known as a *cosmic string*.

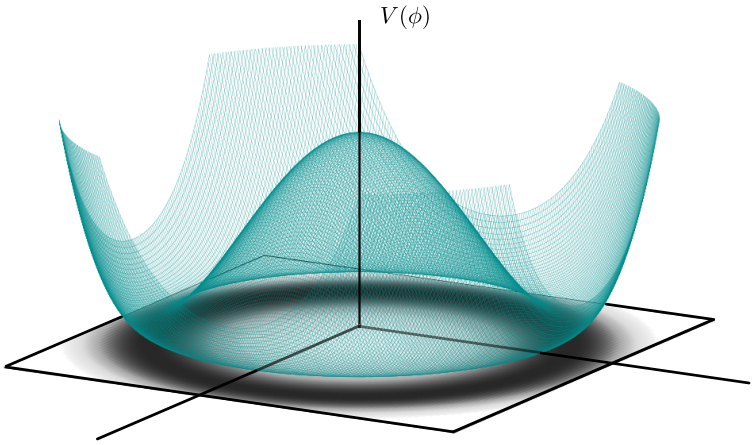

Figure 1–1: Mexican hat shaped potential in the complex plane.

The width of a cosmic string is determined from the Compton wavelength:

$$r \sim \frac{1}{m} \sim \frac{1}{\sqrt{\lambda}\eta} \text{ with } h = c = 1 \tag{1.6}$$

and the energy density is defined to be the difference between the maximum and the minimum of the potential

$$\rho_{\text{cs}} = V(0) - V(\eta) \sim \lambda\eta^4. \tag{1.7}$$

A commonly used parameter to characterize the energy scale is the string tension, which is calculated by integrating the energy density over the cross sectional area of

the string

$$\mu = \int \rho \mathrm{d}A \simeq \eta^2. \tag{1.8}$$

Since the minimum of the scalar field η is related to the energy breaking scale in GUT theory, constraining this value will give us insights on the physics of the very early universe [Danos, 2010].

1.4 Cosmic string wakes

In cylinderical coordinates, the flat spacetime metric can be written as:

$$g^{\mu\nu} = \mathrm{d}t^2 - \mathrm{d}z^2 - \mathrm{d}r^2 r^2 \mathrm{d}\theta^2. \tag{1.9}$$

Unlike regular flat spacetime, the geometry of space perpendicular to a long string has a θ range of $0 < \theta < 2\pi - \Delta\phi$ and the boundary condition $\theta = 0$ is equivalent to $\theta = 2\pi - \Delta\phi$. $\Delta\phi$ is known as the deficit angle:

$$\Delta\phi = 8\pi G \mu \gamma_\mathrm{s} v_\mathrm{s}, \tag{1.10}$$

where v_s is the velocity of the string in the plane perpendicular to it's tangent vector, and γ_s is the Lorentz factor $\gamma_\mathrm{s} = (1 - v_\mathrm{s}/c)^{1/2}$. This metric is locally flat everywhere expect at the tip of the deficit (see Figure 1–2).

As cosmic strings pass through the dark matter fluid, a static observer standing behind the string will observe matter to receive a velocity "kick" (from both sides towards the plane spanned by the tangent vector of the string and its velocity vector)

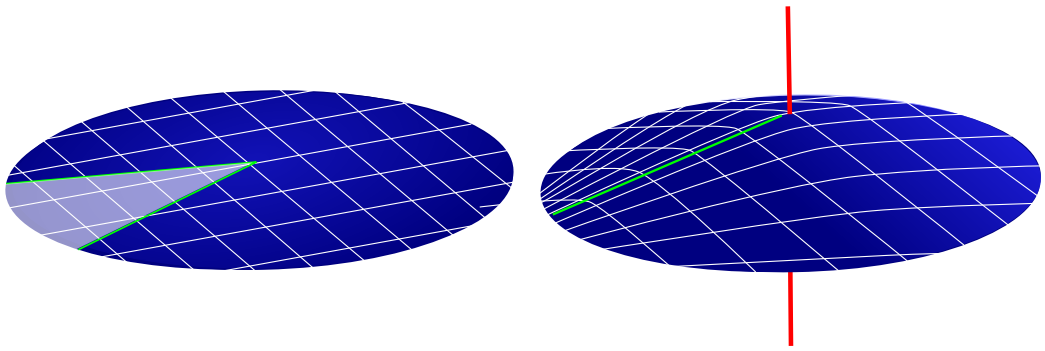

Figure 1–2: Spacetime around a cosmic string. The region shaded in white on the left corresponds to the deficit.

of magnitude

$$v_{\rm kick} = 4\pi G\mu v_{\rm s}\gamma_{\rm s}. \tag{1.11}$$

Matter on each side of the string will accumulate and produce a region of twice the background density [Silk and Vilenkin, 1984]. Overdensity produced via such a mechanism is called a cosmic string wake. A segment of a string laid down at an initial time t_i will produce a wake with dimensions

$$\epsilon t_i \times t_i v_{\rm s}\gamma_{\rm s} \times 4\pi G\mu t_i v_{\rm s}\gamma_{\rm s}, \tag{1.12}$$

where ϵ is a constant of order unity (the front dimension is the correlation length of the string, the second is the depth of the wake (in direction of the string velocity), and the third the initial wake thickness). This initial wake then experiences gravitational accretion growth in thickness and increases in mass. We follow other studies, which have adopted the Zel'dovich approximation (see Chapter 2) to calculate the evolution of these wakes [Danos, 2010], [Duplessis and Brandenberger, 2013],

[Hernández and Brandenberger, 2012]. Particles surrounding the wake are constantly experiencing the Hubble flow and the opposing pull from the wake. Mass shells which are initially moving away from the wake eventually come to rest. The comoving distance at which this occurs we call q and is a function of time. At a given redshift, q can be calculated as:

$$q(z) = \frac{16\pi}{5} \frac{G\mu v_s \gamma_s}{H_0} \frac{(z_i + 1)^{1/2}}{(z + 1)}, \tag{1.13}$$

where z_i is the redshift at which the cosmic string wake was produced and z is the redshift at which the wake is observed. Since the distance of turn-around (distance at which particles experience the same magnitude of acceleration but in opposite directions due to the expansion of the universe and gravitational collapse) is half the distance without gravitational accretion, the density within the turn-arourd radius is doubled. Furthermore, the shell collapses to half of the height, undergoes shocks and virializes, which results in a structure with comoving width

$$w_{\text{comoving}} = \frac{8\pi}{5} \frac{G\mu v_s \gamma_s}{H_0} \frac{(z_i + 1)^{1/2}}{(z + 1)}, \tag{1.14}$$

and density 4 times the background density [Hernández and Brandenberger, 2012].

1.5 Current bounds on the Cosmic string tension

The current bound on the cosmic string tension $G\mu$ is given by CMB observations at $G\mu = 1.7 \times 10^{-7}$ [Dvorkin et al., 2011]. The current non-detection of gravitational waves from cosmic strings loops rule out strings with tension $G\mu = 5.3 \times 10^{-7}$ [Sanidas et al., 2012]. There are also theoretical lower limits including

$G\mu > 10^{-17}$ from femto-lensing of gamma-ray burts [Yoo et al., 2012]. However, since the deficit angle created by strings is small, this effect is unlikely to be seen anytime in the near future. The goal for this research is to use galaxy surveys as a new observational window to constrain the value of $G\mu$.

1.6 Galaxy formation and evolution

Galaxies form in halos with cooled baryonic gas. The lack of pressure support allows the gas to flow towards the center of the gravitational potential to form higher density clumps. In such a process, the initial non-uniform density and cooling efficiency leads to fragmentation of the larger clump, which yields multiple sites for star formation when the temperature becomes sufficiently low in each clump [Binney and Merrifield, 1998]. The relationship between gas density and star formation rates has been confirmed observationally [Heiderman et al., 2010]. In our own galaxy, molecular clouds are distributed in clumps and detailed observations of these gas clumps reveal fine substructures. Not all of these clumps end up forming stars however; only the most massive clouds and cores initiate star formation and the fraction of clouds that do so is low [Mo et al., 2010]. In general, there is a spatial correlation between the gas clouds and young stars, however, the correlation is weaker between gas clouds and older stellar populations. This suggests that the gas gets consumed in the process of star formation and becomes drained quickly [Blitz et al., 2007].

Naively, one would predict that a gas cloud would collapse in its free fall time and form all the stars it could. However, through observations we know that this is not

true and that galaxies are many free fall times older. This could be partially explained by the angular momentum and velocity dispersion a galaxy has, which prevents the gas clouds from collapsing as efficiently as they could [Bournaud, 2011]. Therefore, there is a strong correlation between the internal kinematics and the star formation rate of a galaxy. Other characteristics such as magnetics fields and supersonic turbulence can also play a major role in star formation [Padoan and Nordlund, 2011], [Mac Low and Klessen, 2004].

In general, a giant gaseous cloud produces a wide spectrum of stars; from small mass K-type stars to more massive O-type stars. The spectrum which describes the initial mass spectrum of a stellar population is called the initial mass function (IMF). Treating a star as a black body, the radiative spectrum has Planck's form:

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1}. \tag{1.15}$$

Since massive stars have higher temperatures, it could be deduced that massive stars have a bluer colour and less massive stars are redder. Massive O-type stars turn away from the main sequence first and turn into hyper-giants [Carroll and Ostlie, 2007]. Therefore the existence of blue stars inside a galaxy indicates a relative young age and is associated with recent star formation. In the contrary, the absence of O-type stars indicates that enough time has lapsed for these stars to turn away from the main sequence and therefore signify maturity.

In general, the star formation rate is proportional to the galaxy's age. However, the kinematics and dynamics complicate the scenario, and in particular, merger events drastically alter the internal properties of the colliding galaxies. Merger

events can be classified into four subclasses, wet major/minor mergers and dry major/minor mergers. Wet/dry refer to the abundance of cold gas available in the colliding galaxies and major/minor refer to the relative sizes of the colliding galaxies [Woods et al., 2006]. Wet mergers compress cold gas inside the colliding galaxies, producing over dense regions inducing star formation, whereas dry mergers have no gas and hence mass builds up purely from the increase in number of constituent stars, [Barton et al., 2000], [Barton Gillespie et al., 2003].

Other interaction mechanisms that alter the internal properties of galaxies include:

1. *Cannibalism:* Although galaxies in clusters are unlikely to experience a merger due to the encounter speed being significantly greater than the internal velocity dispersion, energy and momentum in a galaxy could be dissipated due to dynamical friction. This causes the galaxy to fall towards the gravitational potential centre. If the dynamical friction time is sufficiently short, a satellite galaxy is dragged towards the center where the central galaxy is located and merges, a process called galactic cannibalism.
2. *Tidal harassment:* In high density environments, the velocity dispersion is higher than for individual galaxies. When two galaxies encounter each other, the internal energy is increased due to dynamical friction. This results in a less bound galaxy prone to further gravitational interaction.
3. *Ram pressure stripping:* This occurs when a galaxy is in a dense cluster-like environment where the gravitational potential is strong enough to strip the gas

away from the galaxy. This results in a galaxy depleted of cold gas, which shuts down star formation.

4. *Strangulation:* Strangulation occurs for gas that is surrounding a galaxy that has not cooled and or gas that has been expelled from the galaxy itself but is still bound to the halo. Since the halo binding of gas is very weak, the gas is easily stripped away when halos or galaxies merge. Since gas surrounding the galaxy acts as an external reservoir for supplying gas to the galaxy, depletion of such a supply will imply a shorter star forming period.

Recently, there have been enormous efforts in trying to understand the relationship between the environments, and the formation and evolution of galaxies. Specifically, galaxies in the low redshift universe show a strong relationship between environment, morphology, colour and stellar mass [Einasto et al., 1974], [Dressler, 1980], [Tempel, 2011], [Blanton et al., 2005a]. Galaxies which are in a denser environment, have a higher likelihood of interacting with another and show a lower fraction of star formation galaxies [Bassett et al., 2013]. This relation is somewhat visible in the intermediate redshift ranges also [Poggianti et al., 2013]. Such observations motivate us to select galaxy types that are most sensitive to cosmic strings.

Since cosmic string wakes produce regions of dark matter overdensity which galaxies reside in, it could be stipulated that cluster-like environments will be produced. In such environments, cold gas is more likely to be stripped away from the halos and the nested galaxies due to the interaction mechanisms mentioned above. Therefore, it could be naively predicted that the spatial distribution of passive non-star forming galaxies is more sensitive to cosmic strings.

1.7 Galaxy properties

There are three main galaxy properties that are relevant to this research: luminosity, stellar mass and star formation rate (SFR). Bolometric luminosity (total luminosity from all wavelengths) is one of the most fundamental properties of the galaxy since it describes the total energy the galaxy is emitting. Luminosity in different wave bands provides information on the different emission processes. However, luminosity is a derived quantity from observed flux. The measured fluxes must be first corrected for light lost through effects such as atmospheric extinction, quantum efficiency of the detectors, and also the response curve of the filters selected. After such reduction, fluxes are converted to luminosities using the combination of flux-magnitude and magnitude-luminosity relations. Additional, corrections such as dust, K and E corrections (conversion of an object's magnitude to an equivalent measurement in the rest frame of the object) are also applied to consider the radiative processes and physical effects that affect the appearance of galaxies (see Chapter 3).

Stellar mass is also a derived quantity from observed flux and measured redshift. It is estimated by using multi-band photometry and the galaxy's redshift to compare the observed spectral energy distribution (SED) with synthetic templates with various stellar population models (which correspond to different galaxy morphologies) spanning a range of star formation histories, ages, metallicities and dust content. For each synthetic model, the K-band mass-to light ratio ($M_*/L_{\rm K}$), stellar mass and minimum χ^2 is calculated and the probability of fit is determined. This probability

is then integrated over the parameter space to obtain the probability distribution for each galaxy, which is then used to estimate the final stellar mass [Bundy et al., 2005].

Calibration of star formation rates can be done through the electromagnetic spectrum from X-ray to IR. For local resolved galaxies, the SFR is measured by counting individual objects or events that indicate recent star formation [Chomiuk and Povich, 2011]. For galaxies which are further away and are unresolved, the SFR is estimated by targeting the continuum or line emissions in the spectra unique to massive and short lived O/B/A type stars. Some of the commonly used emission lines for this purpose are the Balmer lines Hα, Hβ and [OII]λ3727 [Argence and Lamareille, 2009]. Galaxies with Balmer emission indicate the presence of young massive stars since only photons with sufficiently high energy are able to cause such transitions in the inter-stellar hydrogen. This however assumes that the star formation has been constant over the recent history of the galaxy and that the initial mass function is known so that the total number of stars formed including the brightest and dimmest stars can be extrapolated. For dust enshrouded galaxies, the SFR could be approximated from emission in the mid/far IR range of the galaxy spectrum, which is caused by dust absorbing and re-emitting photons from bright stars [Calzetti, 2012].

1.8 Galaxy surveys

There exists two main types of observations in the optical window. One is photometric observations, which are done by capturing an image of the target object using multiple filters. By using different filters, different segments of the entire electromagnetic spectrum are discretely obtained, which are then used to fit template

galaxy models. The advantage of this type of observation is the cost and speed of reaching high signal to noise ratio (S/N). Therefore, large sky surveys, which ideally spend a minimal amount of time on each object to cover a large sky area in the same given time are often done through this method. Alternatively, photometric observations are also used for pencil beam (deep small area) observations, since photons are collected more efficiently than by the spectroscopic method, which allows us to attain high S/N for high redshift objects, which are often dimmer (see for example [Cuillandre and Bertin, 2006]). The disadvantage for this method is uncertainty in the estimation of galaxy properties, since there could be degeneracies in the models being fitted to the measured photometric data points. One example of this is redshift estimation, which becomes increasingly challenging at $z > 1.5$ [Brammer et al., 2008].

The second type is spectroscopic observations, where the incoming light is dispersed by a slit or a grism. The advantage of this type is the completeness of the spectral information we obtain. This allows us to observe chemical finger prints in the form of emission and absorption lines, which enables us to make accurate and precise estimations of physical properties such as SFR, as discussed previously [Appenzeller, 2012]. However, since a fraction of photons are lost by passing through the slit or grism, fewer photons reach the detector and hence it takes more observational time to reach a significant S/N, increasing the total cost per target object [Howell, 2006]. This makes spectroscopy less suited for large wide field surveys.

With modern technology, spectroscopic instruments are now capable of reaching high S/N moderately quickly for low redshift galaxies. One exemplary instrument

is the Sloan Digital Sky Survey (more in Chapter 3). However, the newer generation spectroscopic surveys such as the Prime Focus Spectrograph (PFS) currently operating on the Subaru Telescope, Big Boss which is planned to operate in 2020 and EUCLID will be capable of acquiring spectra of higher redshift galaxies in the ranges $0.6 < z < 1.6$, $0.2 < z < 3.5$ and $z < 2$ respectively with different field size and threshold S/N [Takada and Silverman,], [LBNL, 2011], [ESA, 2011].

1.9 Motivation and goals

Our goal in this study is to place a constraint on the cosmic string tension $G\mu$. Therefore we start from theoretical modeling, convert the models to observables and place constraints by comparing with observations. The main prediction is that the cosmic string wake will produce a planar overdensity and hence a unique environment for galaxies to form and grow. If this is true, signatures will be seen in spatial clustering and in the characteristic properties of the galaxies. Therefore both spatial clustering and statistical analyses of the galaxy properties will be done.

Clustering analysis can be performed using both photometric surveys and spectroscopic surveys as long as there are enough samples. However, since photometric surveys have high uncertainty in redshift, projected clustering is often analyzed. The advantage of using photometric surveys is the possibility of using galaxies at higher redshifts, where signatures of cosmic string wakes are likely to be more prominent. The advantage of using spectroscopic surveys is the availability of the 3D information since the uncertainty in redshift is negligible (for galaxies with adequate S/N which pass through the initial filtering). In addition, since the spectra will provide us a

more complete description of the galaxy, it will be easier to model the environment the galaxy is in. Combining the two, we are likely to obtain a tighter constraint on the cosmic string tension by using spectroscopic surveys.

In this study WMAP-7 cosmology (scaled Hubble constant $h = 0.702$, baryonic density $\Omega_{\rm b} = 0.0455$, matter density $\Omega_{\rm m} = 0.272$, dark energy density $\Omega_\Lambda = 0.728$, primordial spectral index of scalar fluctuations $n_{\rm s} = 0.961$, amplitude of the (linear) power spectrum on the scale of $8h^{-1}$ Mpc $\sigma_8 = 0.807$) will be used and magnitudes are quoted in the AB-system.

CHAPTER 2
Simulations

2.1 Initial Conditions: 2LPT

In linear perturbation theory, which is applicable to pressureless fluid, the density field of matter, gravitational acceleration, and the peculiar velocity of matter grows self-similarly with time. In this Lagrangian description, the growth of structure is given by the displacement and the peculiar velocity of each mass in terms of the initial position. In a procedure known as the Zel'dovich approximation, this formulation is used to extrapolate the evolution of structures into the regime where displacements are no longer infinitesimal [Zel'dovich, 1970]. In the original development of cosmological N-body simulations, the Zel'dovich approximation was commonly used to produce the initial conditions due to its computational efficiency. However, the Zel'dovich approximation was found to only reproduce correct linear growth, underestimating skewness and higher order moments at high redshifts [Grinstein and Wise, 1987]. As a result, using the Zel'dovich approximation as a starting point of a simulation leads to incorrect second and higher order growing modes, which enhances non-linear decaying modes called transients [Crocce et al., 2006]. Therefore, we utilize a code that employs a higher order calculation, namely second order Lagrangian perturbation theory (2LPT) in producing the initial conditions with the primordial non-Gaussianty generation turned off [Scoccimarro et al., 2012]. The code takes in a matter power spectrum and a

transfer function to generate initial condition at arbitrary redshifts. These input functions were generated by CAMB as shown in Figure 2–1, using WMAP-7 cosmology. 2LPT is much less affected by the aforementioned problems, and can correctly reproduce the skewness of the density field that develops even from Gaussian initial conditions [Zhao et al., 2013]. Therefore, the starting redshift of the simulation can be significantly lower relative to simulations where Zel'dovich approximations are used, while attaining similar accuracy [Scoccimarro et al., 2012] [Zhao et al., 2013].

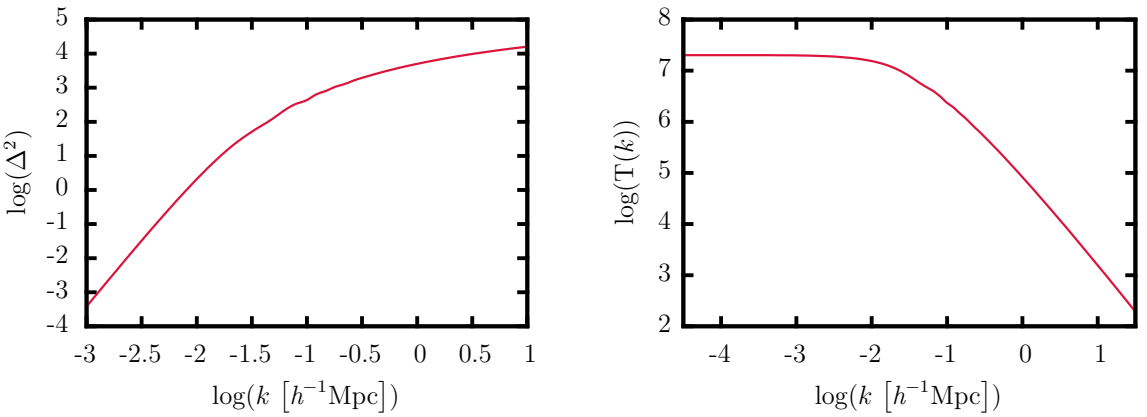

Figure 2–1: Input matter power spectrum and transfer function calculated from CAMB [Lewis and Bridle, 2002] for $z = 0$, which are used to produce the initial conditions at $z = 50$.

Although running the simulation from the time of wake's birth at $z \sim 5000$ is ideal, this is computationally inefficient. Alternatively, we could begin our simulation at a redshift at which an analytic descriptions of the wake structure are obtainable and apply these to the initial conditions. Since redshift scales with time and non-linearities in the cosmic structures become prominent at low redshifts, structures can be predicted fairly accurately at intermediate redshifts ($z \sim 50$) and higher.

Therefore our strategy is to use observed cosmological parameters to produce initial conditions at intermediate redshifts and overlay cosmic string wakes calculated analytically. We let this state evolve using a N-body simulation and compare the results with observations.

There are primarily two trivial ways dark matter particles can be distributed in the initial state of the simulation. One is using a cubic grid arrangement and the other is using a Poisson distribution. For the cubic grid arrangement, the power spectrum lacks power on all scales other than the Nyquist frequency of the particles [Baugh et al., 1995]. In contrast, a Poisson distribution has power at all scales, but contains a high degree of shot noise. The shot noise becomes a severe problem on the very small scales where the noise dominates the input spectrum. Both of these configurations will cause problems in our studies as these effects will add noise to the halo mass function. Alternatively, a glass-like distribution, where the particles are displaced with maximum separation without structure could be used. This distribution was shown to produce different voids in comparison to a simulation started from a grid, but identical power spectrum [Baugh et al., 1995]. We choose to use this glass–like configuration to avoid simulation artifacts. In practice, this glass–like configuration is produced by particles randomly placed in the simulation box and running the simulation with a negative gravitational force. Within a few time steps, a stable configuration is reached and a glass–like cell is tiled to produce the initial simulation condition.

The modeling of the cosmic string wake was done by modifying the x, y, z positions and x, y, z velocities of randomly selected samples in the simulation box to

preserve external structures. The randomly selected particles were projected onto the wake sheet until the density inside the wake was 4 times the local (defined as the region within 50 h^{-1}Mpc of the wake in the direction perpendicular to the surface of wake) background density. Additionally, a velocity kick (eq. 1.10) was also added to the velocities of particles within 50 h^{-1}Mpc of the wake, which is the length and width of the wake.

Three models were made: 1. Model with no cosmic strings $G\mu = 0$ as the control sample (colored gray through out this thesis), 2. Model with $G\mu = 10^{-5}$ to easily identify the effects induced by the string (colored red through out this thesis), 3. Model with $G\mu = 10^{-7}$ which is the current limit on the cosmic string tension (colored blue through out this thesis).

2.2 Evolution: GADGET-2

Gadget-2, a N-body code written by Volker Springel and collaborators [Springel et al., 2001] [Springel, 2005] which is widely used for cosmological simulations such as the Millennium I/II [Springel et al., 2005] [Boylan-Kolchin et al., 2009] and GiggleZ simulation [Poole, prep] was used to evolve the particles through time. Gadget-2 makes use of a TreePM algorithm to compute the gravitational forces accurately. The tree algorithm groups distant particles into larger cells and approximates their potentials using multipole expansions about the center of mass of the group [Barnes and Hut, 1986]. The algorithm is used to evolve the initial particle distribution ($z = 50$) to the final state ($z = 0$), during which 160 snapshots are taken for analysis. The time resolution was chosen to be $\Delta a = 1.0265$ for the 150 snap shots and $\Delta a = 1.0027$ for the last 10

where a is the scale factor which represents the relative expansion of the universe. The higher time resolution towards the low redshifts was chosen so that the semi-analytic model will produce galaxies accurately. Using these snapshots, merger trees and halo catalogs were produced.

2.3 Halo finding: ROCKSTAR-0.99

In locating dark matter halos in the simulation, we used the ROCKSTAR code by Behroozi, Welchsler and Wu [Behroozi et al., 2013a]. This code is based on adaptive hierarchical refinement of Friends-of-Friends (FOF) groups in six phase-space dimensions and one time dimension. It first uses a 3D Friends-of-Friends algorithm with large linking length to identify parent FOF groups. Once these are identified, subgroups in phase space are identified by reducing the FOF linking length. The FOF subgroups are then converted to halos beginning at the deepest level of hierarchy (for further details refer to [Behroozi et al., 2013a]). Comparisons of halo finding algorithms have been studied in [Knebe et al., 2011]. The parameters used are shown in Table 2–1, which were chosen from cross-calibrating with theory, observations and other simulations. The list of possible outputs are listed in appendix A. We further run the tree consistency checker [Behroozi et al., 2013b] to reject any spurious phantom halos, which are artifacts produced by discreteness of time-steps in simulations.

There are several ways to quantify the mass of a halo. One of them is to use the total number of particles times the mass of the particles. This intuitive definition encounters a problem when two halos merge where the particle's association to a

Parameters	Value
Halo definition	r_{200b}
Linking length	0.28
Friends-of-friend fraction	0.7
Unbound threshold	0.5
Host potential ratio	0.3
Double count subhalo mass ratio	1
Minimum halo particles	10

Table 2–1: Parameters used in ROCKSTAR. For detailed explanation of each parameter, refer to [Behroozi et al., 2013a].

specific halo becomes ambiguous. Another mass definition is mXb or mXc, where X refers to the total mass enclosed in a radius with the density X times larger than the background density or the critical density respectively. The choice of the value of X often depends on the application. For example, in the context of measuring X-ray mass of a cluster, m500 or m2500 is used (mass enclosed in a smaller region relative to m200) since X-rays are emitted only in the cores of clusters where the temperatures are high enough to do so [Morandi, 2008]. It is therefore more logical to use mass definitions such as m200b (which is approximately half of the total asymptotic cluster mass for dark matter halos), due to their diffuse natures [Busha et al., 2005].

2.4 Semi-analytic models: Semi-Analytic Galaxy Evolution (SAGE)

Although dark matter simulations are often used to understand the large scale structures in the universe, we require a conversion from dark matter structure to

baryonic structures to compare simulations with observations. One way to infer the baryonic content in halos is by using the conditional luminosity function (CLF), which assigns a probability of obtaining a galaxy with luminosity L inside a halo with mass $M_{\rm h}$ [Yang et al., 2003], [van den Bosch et al., 2003]. The CLF provides useful information such as the average relation between light and mass and occupation number of galaxies. However, it requires prior assumptions such as the functional form of the luminosity function and its parameter dependence on halo mass [Yang et al., 2005]. The subhalo abundance matching method (more in Chapter 4.) on the other hand, assumes a monotonic relation between the halo mass $M_{\rm h}$ and luminosity L predicting that larger dark matter halos host more luminous galaxies. Both of these simple methods only predict the luminosity of galaxies, although they do remarkably well in doing so [Mo et al., 2010].

Since there is correlation between luminosity and local density of galaxies, (which could be seen from the luminosity dependence of the correlation function) we are able to use luminosity as a first-order proxy for environment. However, this alone is not enough to discriminate between galaxies in high and low density environments. Therefore, we use more sensitive proxies such as SFR to probe the environment and the evolutionary stage the galaxy is in. In doing so, we must use galaxy formation models that provide a more complete description of galaxies produced using physical principles, namely semi-analytic models.

All semi-analytic models assume that baryons are initially distributed uniformly and that they mimic the dark matter distribution on scales above the Jean's length at

lower redshifts [Arons and Silk, 1968], [Gnedin and Hui, 1998]. Galaxies form at locations of deep gravitational potential produced by dark matter halos where baryons accumulate. Within the (sub)halos present at each snapshot of the simulation, the amount of three distinct baryonic components – hot gas, cold gas, and stars – are kept track of.

Prescriptions are then assigned to the physical processes that govern conversion between the baryonic components. Cooling mechanisms such as atomic, Compton and molecular hydrogen cooling convert hot gas into cold gas, star formation converts cold gas into stars and heating mechanisms such as photo-heating and supernova feedback converts cold gas into hot gas or increases the temperature of existing hot gas. In these conversions, the fractions of metals produced are also encoded, which affect the efficiency of cooling and heating mechanisms at later times. All semi-analytic models consider these physical mechanisms but are implemented with different prescriptions, although quite often, various models give similar results despite very different modeling due to our lack of understanding.

Similar analyses can be done by using hydro-simulations. The advantage of semi-analytic models is the computational efficiency (in terms of speed and intensity), which is orders of magnitude better than hydro-simulations, and therefore allows for rapid exploration of parameter space [Henriques et al., 2009] and model space (which allows for quick modifications to the modeling). The foremost disadvantage is the degree of assumptions and approximations required [Benson, 2010].

In our study, we utilize the "Semi Analytic Galactic Evolution" (SAGE) code developed by Croton et al. obtained through private communication [Croton, prep].

Note that observational studies of galaxies are done by analyzing the spectra of galaxies. Different physical processes emit at different wavelengths and therefore our understanding of all these processes is crucial to understanding the internal physical properties of galaxies given observational data. Modeling the emergent spectrum from a theoretically modeled galaxy requires proper formulation in the amount of baryonic material available, and radiative processes such as dust and gas emission/remission processes, which are heavily model dependent. Therefore we avoid using properties such as colour and apparent magnitude. We do however, make indirect use of luminosity to produce volume limited samples.

2.5 Lightcones

N-body simulations are often run in cubical domains with all the particles at the same redshift. Unlike such a geometry, real observational surveys do not have simple shapes, and therefore manipulations of either the data or the simulation is needed to match their geometries. This can be accounted for by:

1. Simulating in a large enough volume to encompass the survey volume.
2. Simulating in a non-cubical box.
3. Making use of the periodic boundary to crop out desired geometry.

If we wish to make comparisons between pencil beam surveys (small sky coverage with high depth) and simulations, it is clear that a cube is not an ideal simulation geometry, since covering the same depth requires an enormous simulation box, with most of the simulation volume ending up being unused. 2. often has numerical issues when the lengths of the sides are highly disproportionate in addition to the simulation

losing its applicability for other studies. 3. often does not preserve simulation volume and repetition in structure is inevitable, although it is avoided as much as possible.

Instead of using these methods, we use the remapping code developed by Carlson and White, which identifies a unique sub-volume inside tiled simulations with periodic boundary conditions, in which the volume and structure is preserved [Carlson and White, 2010]. For example, the simplest case of $L = 1 : 1 : 1 \rightarrow L = \sqrt{2} : \sqrt{2}/2 : 1$ could be done by dividing the cube into quadrants and shifting two of the quadrants to their continuous faces (refer to Figure 2–2).

We then translate the simulations into position space such that the simulation lies over the survey coordinates. This region is further sliced into shells with respect to the distance from the coordinate origin and galaxy samples are extracted from the corresponding snapshots. Finally, a fan-like geometry is extracted, making maximum use of the cuboid. The limits in right ascension and declination found this way are also used to crop out the SDSS catalogue to preserve geometry.

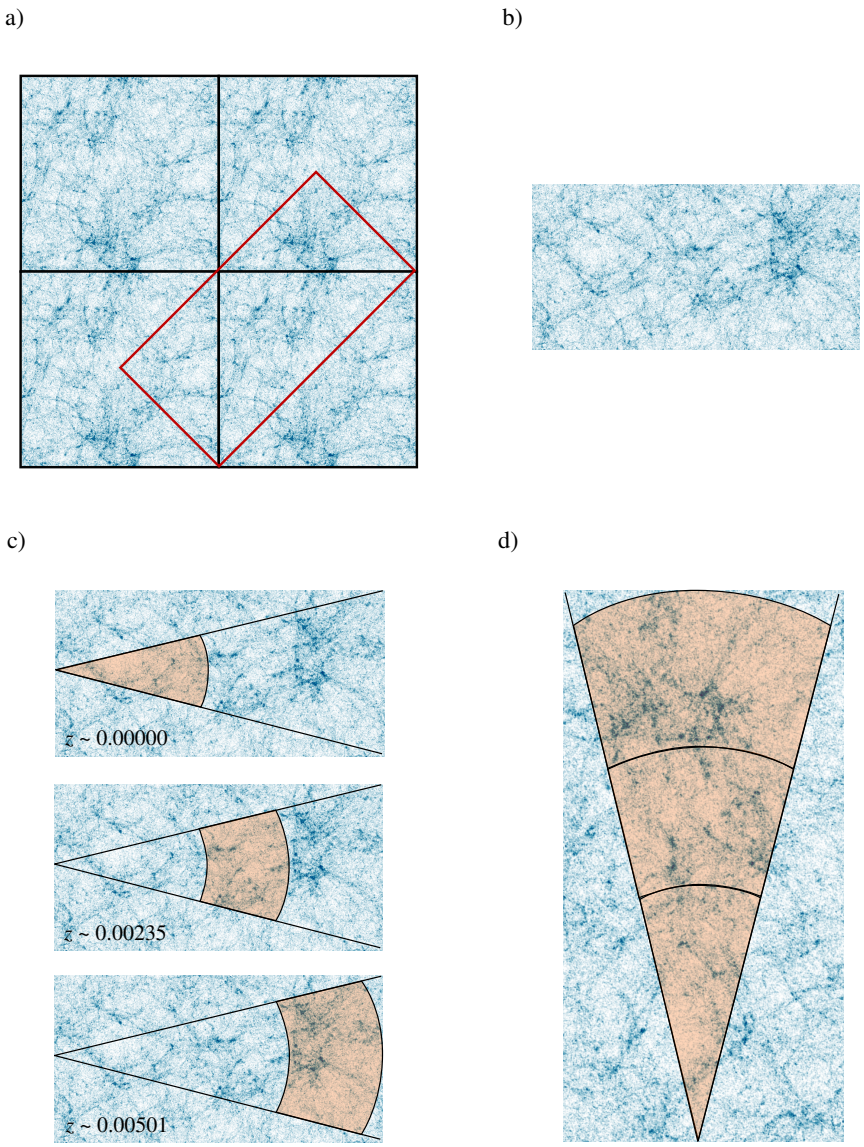

Figure 2–2: Steps in producing the lightcones. (a),(b). The simulation is tiled, making use of the period boundary conditions such that structure is continuous across the boundaries. Note that not only does this preserve structure in the simulation, but it uses *all* of the original simulation volume, and there is no repetition in the structure. (c) Appropriate regions in the simulation box are sliced in each snapshot. (d) The sliced simulations are stacked to form a lightcone.

CHAPTER 3
SDSS

The Sloan Digital Sky Survey (SDSS) [York et al., 2000], [Stoughton et al., 2002] is one of most important surveys in large scale structure studies. It uses a 2.5 m telescope [Gunn et al., 2006] and scans the high-latitude sky in the northern galactic cap. The original survey (SDSS-I) started taking data in April 2000 and it is currently in Phase III. A drift scanning mosaic CCD camera uses 5 photometric bands u, g, r, i, z [Smith et al., 2002] and currently reaches down to a limiting magnitude of $r \sim 22.5$. The imaging data are then passed through a series of data reduction pipelines, which receive astrometric calibration, photometric reduction and photometric calibration, [Pier et al., 2003], [Lupton et al., 1999], [Hogg et al., 2001], [Ivezić et al., 2004]. From these photometric images, galaxies were identified and were selected for spectroscopic follow-ups.

3.1 NYU-VAGC & MPA-JHU-VAGC

We base our galaxy samples on the New York University value added catalogue (NYU-VAGC) of large scale structure (LSS) samples [Blanton et al., 2005b] and the MPA-JHU value added catalogue [Tremonti et al., 2004]. Both these catalogues are based on the SDSS data release 7, which contains approximately 700,000 main sample galaxies and covers about 8000 deg^2 of the sky. The NYU-VAGC is a catalogue that

contains positional and basic photometric data with objects carefully selected in terms of selection functions, while MPA-JHU-VAGC is a catalogue which contains extra parameters such as stellar mass, SFR and metallicity derived from spectra. We base our catalogue on the NYU-VAGC and append extra derived parameters by cross matching z, right ascension (ra) and declination (dec) to within 0.001, 0.01 degrees and 0.01 degrees respectively. The primary information given in the NYU-VAGC are the positional information z, ra, dec, absolute and apparent magnitudes and the completeness of the sector where the galaxy is located. We use the magnitudes of the galaxy to produce volume limited samples, which ensures the completeness of our sample sets. Prior to doing this, two corrections must be made to the magnitudes in the catalogue. One is the K-correction, which is the correction made to take in account the redshifting of light. This correction is necessary to obtain the rest-frame magnitude instead of the observed magnitude such that galaxy selection will be based on the physical property of the galaxy and not its apparent property. The other is the Evolution-correction or E-correction to correct for the magnitude due to the spread in redshift in given galaxy data. The K-correction is pre-calculated using the method noted in [Blanton et al., 2003a] and the E-correction is calculated by:

$$M_{\mathrm{r,ev}} = M_{\mathrm{r}} + Q(z - 0.1), \tag{3.1}$$

where Q=1.6 and M_{r} is the absolute r-band magnitude [Blanton et al., 2003b]. This calibrates the galaxy magnitudes as if they were at $z = 0.1$, which is near the median redshift of the *full* LSS-sample. Although the mean redshift of our galaxy samples is

lower than this redshift, it is advantageous to calibrate the luminosity at this redshift so that comparisons with other studies can be made, which often employ $z = 0.1$.

Due to the design of the instrument, no spectra of two targets within 55" (which is the diameter of the fibre used to obtain spectra) of each other can be simultaneously obtained. From this limitation, approximately 7% of the galaxy spectra unobtainable. A simple method to circumvent this problem is to assign the collided galaxy a redshift of the other galaxy within the fibre diameter. Since clustering at these scales are likely due to true clustering and not line-of-sight effect, it turns out that this is a reasonable procedure for about 60% of galaxies [Blanton et al., 2003d].

With the r-band absolute magnitude corrected, we create volume-limited samples by applying cuts in magnitude. We produce 3 subsamples L1, L2, L3 with absolute magnitude ranges $-18.5 < M_{\rm r} < -17.5$, $-19.5 < M_{\rm r} < -18.5$, $-20.5 < M_{\rm r} < -19.5$ as shown in Figure 3–1. All these magnitude ranges are considered low magnitude in other studies, and the choice of these magnitude ranges originates from the depth of the simulation box we could make. Although the total number of galaxies in volume limited samples is significantly lower than that in flux limited samples, volume limited samples have an easier interpretation and are well defined. Flux limits are also applied to the samples and galaxies are selected from the range $14.5 < r < 17.6$. The faint apparent magnitude limit is applied to obtain a uniform magnitude limit across the whole survey. Below this such limit, LSS analysis will be contaminated by observational bias, where apparent clustering will be present in regions where observations are deeper (with longer exposures). The brighter flux limit is imposed to avoid spurious objects that mimic galaxies, and also to avoid

small incompletenesses associated with galaxy de-blending in the process of making the NYU-VAGC safe samples [Zehavi et al., 2011].

Stellar mass calculations in the MPA-JHU-VAGC are based on Bayesian methodology and model grids [Kauffmann et al., 2003] assume a Kroupa initial mass function [Kroupa, 2001]. However, the spectra of each galaxy does not cover the entire galaxy since the aperture is 3" in size (a typical galaxy is $\sim 2-10$" at $z < 0.2$ in r-band [Unzicker and Fabian, 2010]) and hence the model is based on u, g, r, i, z galaxy photometry only instead of 4000 angstrom break and Hδ measurements. Star formation methods are based on techniques discussed in [Brinchmann et al., 2004] with fits to star forming galaxies carried out using the same method as [Charlot and Longhetti, 2001] and with derivations of star formation rates (SFR) for other classes of galaxies such as active galactic nucleis (AGNs) and composite galaxies based on [Brinchmann et al., 2004].

While many of the other studies [Zehavi et al., 2011], [McBride et al., 2011a], [Guo et al., 2013] split blue cloud galaxies from the red sequence galaxies in colour–magnitude space, we make a cut in stellar mass – SFR space using:

$$SFR = 2 \times 10^{-8.5}\ M_*^{0.75}, \tag{3.2}$$

since stellar mass and SFR are directly calculated for each galaxy in the semi-analytic code, whereas color and magnitudes require additional stellar population modeling. In Figure 3–2, the low mass star forming field galaxies occupy the left portion of the upper group. These galaxies undergo wet merging and increase in size (move

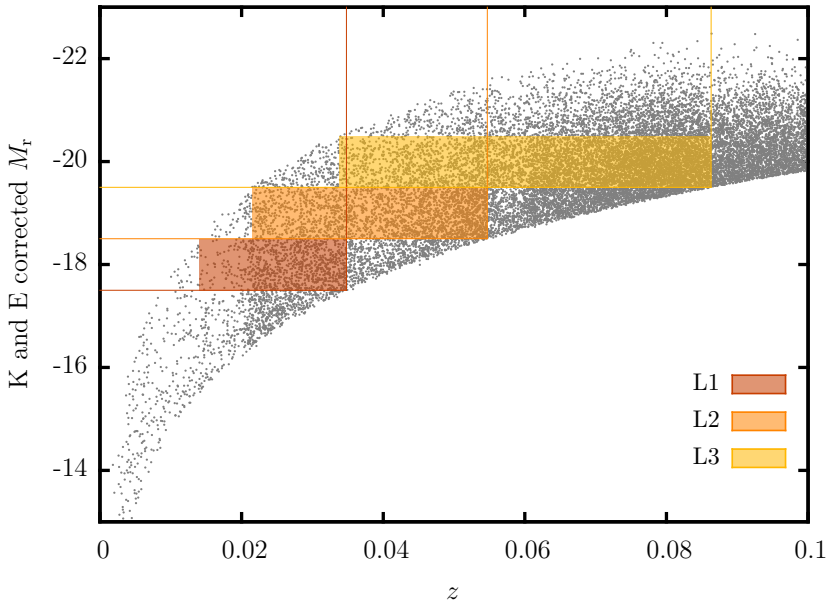

Figure 3–1: Producing volume-limited samples. Galaxies that fall in regions shaded in orange are included in the subsamples.

towards the right). These large wet merged galaxies quench gas and become non-star forming. These further evolve or dry merge, making a transition to the lower group. Since it takes time for galaxies to go through this sequence, making a cut in stellar mass – star formation rate is, to first order, dividing the population in terms of stellar age and cold gas abundance. Basing the cut in colour splits the population in a similar way but is less sensitive to the abundance of cold gas contained inside galaxies.

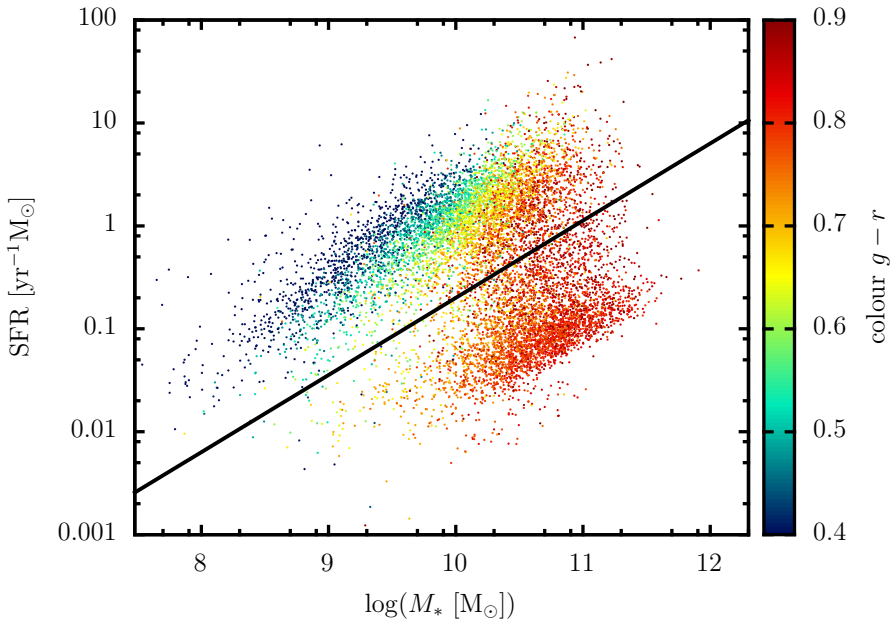

Figure 3–2: M_* against SFR of SDSS galaxies. The black line shown is eq. 3.2. The number of galaxies has been reduced for clarity. The colour scale was added to show the gradient in galaxy colour, which is another property often used to split galaxy populations based on type. Taking a cut in colour is equivalent to taking a steeper cut in the stellar mass–SFR space.

Sample	M_r range	$\langle M_\mathrm{r} \rangle$	σ_{M_r}	Number
L1	$[-17.50 : -18.50)$	-17.97	0.29	8486
L2	$[-18.50 : -19.50)$	-18.98	0.28	15442
L3	$[-19.50 : -20.50)$	-19.95	0.29	11912

Sample	$\langle M_\mathrm{r} \rangle$	Number
L1-Q	-18.01	1812
L2-Q	-19.02	4633
L3-Q	-19.98	5579
L1-S	-17.98	6674
L2-S	-18.96	10809
L3-S	-19.93	6333

Table 3–1: Absolute magnitude ranges, means, dispersions and numbers for each sub-sample. Q stands for "Quenched" equivalent to non-star forming galaxies and S stands for "Star forming" galaxies.

CHAPTER 4
Matching SAM and SDSS galaxies

4.1 SubHalo Abundance Matching (SHAM)

One of the simple approaches to inferring the luminosities of galaxies residing in halos is to assume a monotonic relation between subhalo mass and luminosity of the residing galaxy, since larger halos have deeper gravitational potential and will accumulate more baryonic mass.

First, a halo catalogue with the same geometry as the volume limited samples is produced using the lightcone procedure. Similarly, magnitudes of galaxies within the same geometry cropped from the SDSS catalogue are extracted. Both the halo catalogue and the magnitude list are then rearranged from highest to lowest mass and brightest to dimmest magnitudes, which are then matched one-to-one. We follow the works of Conroy, Weschler and Kravtov [Conroy et al., 2006] and order the halos in terms of the maximum rotational velocity $v_{\rm max}$, identified from the well defined peak of the relation $v_{\rm c}(r) = (GM(< r)/r)^{1/2}$, which serves as a natural halo mass scale [Diemand and Moore, 2011].

For isolated halos, the circular velocity is directly correlated with the baryonic mass of the galaxy residing inside. For subhalos which are accreted onto larger halos, this correlation is disrupted due to the strong tidal forces from the virialized region of the parent halo. Since the baryonic component of the galaxy is more strongly bound to the gravitational potential of the parent halo, it is less affected by tidal

disruptions. Therefore, in the case of subhalos which are accreted onto a larger halos, the luminosity of the galaxy is correlated with the circular velocity $v_{\rm c}$ at the epoch of accretion rather than the $v_{\rm max}^{\rm now}$ [Conroy et al., 2006].

Note that although SDSS data are used here, we are not calibrating our simulations to it. We are merely using it to approximate the luminosity of the semi-analytic model galaxies to make a cut in luminosity to produce volume limited samples.

CHAPTER 5
STATISTICAL ANALYSIS

To quantify the differences between the simulations, we utilize statistical tools commonly used in clustering analyses. We present both the statistical analysis for the halos and for the galaxies formed from the semi-analytic model.

5.1 Halo mass function

The halo mass function describes the number density of halos as a function of mass and it generally evolves as a function of redshift, describing the accretion history of dark matter halos and subhalos. First, we define halo mass M_Δ to be

$$\Delta = \frac{M_\Delta}{(4/3)\pi R_\Delta^3(M)\rho_{\rm m}} \tag{5.1}$$

where Δ is the over-density within a sphere of radius $R_\Delta(M)$ with respect to the mean background density of the universe at the epoch of analysis $\rho_{\rm m}(z) \equiv \Omega(z)\rho_{\rm crit}(z) \equiv \rho_{\rm m}(0)(1+z)^3$. We define $\Delta = 200$ for reasons discussed in Chapter 2. Using this definition of halo mass, the mass distribution is compared with the Tinker mass function [Tinker et al., 2008], which is commonly used for calibration and comparison in literature. In general, the mass function can be expressed in the form:

$$\frac{{\rm d}n}{{\rm dlog}_{10}M} = f(\sigma)\frac{\bar{\rho}_{\rm m}}{M}\frac{{\rm d}\ln\sigma^{-1}}{{\rm d}M} \tag{5.2}$$

where $f(\sigma)$ is a universal function which could be parametrized as:

$$f(\sigma) = A\left[\left(\frac{\sigma}{b}\right)^{-a} + 1\right] e^{-c/\sigma^2}. \tag{5.3}$$

Here, $\sigma(M)$ is the mass variance of the smoothed density field and can be written as:

$$\sigma = \int P(k)\hat{W}(kR)k^2 \mathrm{d}k, \tag{5.4}$$

where $P(k)$ is the linear matter power spectrum as a function of wavenumber k, and W is the Fourier transform of the real space top-hat window function of radius $R_\Delta(M)$. The parameters A, a, b, and c are constants to be calibrated by observations and simulations. A sets the overall amplitude of the mass function, while a and b set the slope and amplitude of the low-mass power law, respectively. c determines the cutoff scale at which the abundance of halos exponentially decreases. For the halo definition of r_{200} that we use, the fitted parameters are $A = 0.186, a = 1.47, b = 2.57, c = 1.19$ at redshift $z = 0$ [Tinker et al., 2008].

Since the mass function is dependent on the redshift and describes the growth of halo mass through accretion, we also track the evolution of the halo mass function over time.

5.2 Stellar mass function

Similar to the halo mass function, the stellar mass function measures the number density of galaxies as a function of stellar mass. Although the maximum likelihood method is preferred in the latest studies such as the SDSS-DR7 luminosity function

analysis [Blanton et al., 2003c], the so called $1/V_{\rm max}$ method will be employed to calculate the luminosity function of galaxies, which is a simpler method. Here, $V_{\rm max}$ is the maximum volume that a galaxy could be seen given its luminosity, and should not be confused with $v_{\rm max}$, which is the peak rotational velocity of the host halo. The advantage of $1/V_{\rm max}$ is the ease of computation and lack of prior assumptions such as the functional form of the luminosity function. The disadvantage is the assumption that galaxies are uniformly distributed, which is not strictly true at low redshifts.

The luminosity function of galaxies, $\Phi(L)$ is defined as:

$$\mathrm{d}n(L) = \Phi(L)\mathrm{d}L \tag{5.5}$$

where $\mathrm{d}n(L)$ is the number density of galaxies with luminosity $L \pm \mathrm{d}L/2$. Due to the magnitude limit set by the survey telescope, there is a cutoff in L for a given d_L, set by:

$$5\log[d_{\rm max}(L)/{\rm Mpc}] = m_{\rm lim} - M_\odot - 25 + 2.5\log(L/L_\odot) - K - E. \tag{5.6}$$

We then logarithmically bin the luminosity and count the number of galaxies. Therefore, for a magnitude limited sample covering a solid angle ω the expected number of galaxies with luminosities in the range $L \pm \mathrm{d}L/2$ is

$$\mathrm{d}N = \Phi(L)V_{\rm max}(L)\mathrm{d}L, \tag{5.7}$$

where

$$V_{\rm max}(L) = \frac{\omega}{3}\left[\frac{d_{\rm max}(L)}{1+z}\right]^3 \tag{5.8}$$

and thus

$$\Phi(L)\mathrm{d}L = \frac{\mathrm{d}N(L)}{V_{\max}(L)} = \sum_{i}^{N} \frac{1}{V_{\max}(L_i)}. \tag{5.9}$$

To obtain the stellar mass function, instead of binning in L, we bin in the corresponding stellar mass M_*,

$$\Phi(M_*)\mathrm{d}M_* = \sum_{i}^{N} \frac{1}{V_{\max}(M_*(L_i))}. \tag{5.10}$$

5.3 Halo occupation density

Since the clustering of dark matter halos is understood well, we can model the galaxy distribution in terms of the halo clustering. The properties of galaxy clustering could be split into two terms: the one-halo term and the two-halo term. The one-halo term is the correlation between galaxies within the same parent halo. The two-halo term on the other hand is the correlation between galaxies in different parent halos. Following the derivations from [Mo et al., 2010], these are described as follows.

Suppose there are N galaxies in a halo of mass M with average spatial distribution given by the normalized function $u(\mathbf{x}|M_\mathrm{h})$ (i.e the probability of finding another galaxy separated by $\mathbf{x}$ in a halo of mass M_h). The average number of galaxy pairs separated by $\mathbf{r} = \mathbf{x_2} - \mathbf{x_1}$ within such a halo is:

$$Nu(\mathbf{x}_1 - \mathbf{x}_0|M_\mathrm{h})(N-1)u(\mathbf{x}_2 - \mathbf{x}_0|M_\mathrm{h})\mathrm{d}^3\mathbf{x}_1\mathrm{d}^3\mathbf{x}_2 \tag{5.11}$$

where $\mathbf{x_0}$ is the center of the parent halo. The average number of galaxy pairs with separation $\mathbf{r}$ per unit volume of halo with mass $M_{\rm h}$ and N galaxies is:

$$\begin{aligned} DD^{\rm 1h}(\mathbf{r}|M_{\rm h},N) &= \int Nu(\mathbf{x}_1-\mathbf{x}_0|M_{\rm h})(N-1)u(\mathbf{x}_2-\mathbf{x}_0|M_{\rm h}){\rm d}^3\mathbf{x}_0 \\ &= \int Nu(\mathbf{x}|M_{\rm h})(N-1)u(\mathbf{x}+\mathbf{r}|M_{\rm h}){\rm d}^3\mathbf{x}. \end{aligned} \tag{5.12}$$

Since halo mass and the number of galaxies inside do not have a strict one-to-one relation, we describe this halo occupation as a probability function $P(N|M_{\rm h})$ and get:

$$\begin{aligned} DD^{\rm 1h}(\mathbf{r}|M_{\rm h}) &= \sum_N P(N|M_{\rm h})DD^{\rm 1h}(\mathbf{x}|M_{\rm h},N) \\ &= \int \langle N(N-1)|M_{\rm h}\rangle u(\mathbf{x}|M_{\rm h})u(\mathbf{x}+\mathbf{r}|M_{\rm h}){\rm d}^3\mathbf{x} \end{aligned} \tag{5.13}$$

where we have described the average number of galaxies in a halo of mass $M_{\rm h}$ as

$$\langle N|M_{\rm h}\rangle \equiv \sum_N NP(N|M_{\rm h}) \tag{5.14}$$

which is called the *halo occupation density*. Integrating over the mass function

$$DD^{\rm 1h}(\mathbf{r}) = \int n(M_{\rm h})\langle N(N-1)|M_{\rm h}\rangle u(\mathbf{x}|M_{\rm h})u(\mathbf{x}+\mathbf{r}|M_{\rm h}){\rm d}^3\mathbf{x}{\rm d}M_{\rm h} \tag{5.15}$$

gives the total number of galaxy pairs with separation $\mathbf{r}$.

On the other hand, the number of inter-halo galaxy pairs separated by $\mathbf{r}=\mathbf{x_2}-\mathbf{x_1}$ between halos with $N_1, M_{\rm h,1}$ and $N_2, M_{\rm h,2}$ per ${\rm d}^3\mathbf{x}_1{\rm d}^3\mathbf{x}_2$ is given by

$$N_1u(\mathbf{x_1}-\mathbf{x_0}|M_{\rm h,1})N_2u(\mathbf{x_2}-\mathbf{x_0'}|M_{\rm h,2}), \tag{5.16}$$

and the probability of finding two halos in small volumes $\mathrm{d}^3\mathbf{x}, \mathrm{d}^3\mathbf{x}'$ separated by a distance $\mathbf{x_0} - \mathbf{x_0'}$ is proportional to

$$\begin{aligned} P^{\mathrm{2h}} =& n(M_{\mathrm{h},1})n(M_{\mathrm{h},2})\mathrm{d}M_{\mathrm{h},1}\mathrm{d}M_{\mathrm{h},2} \\ & \times \left[1 + \xi_{\mathrm{hh}}(\mathbf{x_0} - \mathbf{x_0}'|M_{\mathrm{h},1}, M_{\mathrm{h},2})\right] d^3\mathbf{x_0} d^3\mathbf{x_0'}, \end{aligned} \tag{5.17}$$

where ξ_{hh} is the halo two-point correlation function. Therefore, the probability of finding an inter-halo galaxy pair separated by $\mathbf{r}$, which is hosted by halos with masses M_1, M_2 with separation $\mathbf{x_0} - \mathbf{x_0'}$ is the product of these two equations, and the number of galaxy pairs can be written as:

$$\begin{aligned} & DD^{\mathrm{2h}}(\mathbf{r}) \\ & = \int dM_{\mathrm{h},1} dM_{\mathrm{h},2} n(M_{\mathrm{h},1}) n(M_{\mathrm{h},2}) \langle N_1|M_{\mathrm{h},1}\rangle \langle N_2|M_{\mathrm{h},2}\rangle \\ & \times u(\mathbf{x_1} - \mathbf{x}|M_{\mathrm{h},1}) u(\mathbf{x_1} - \mathbf{x}'|M_{\mathrm{h},1}) \\ & \times \left[1 + \xi_{\mathrm{hh}}(\mathbf{x_0} - \mathbf{x_0'}|M_{\mathrm{h},1}, M_{\mathrm{h},2})\right] \mathrm{d}^3\mathbf{x_0} \mathrm{d}^3\mathbf{x_0}'. \end{aligned} \tag{5.18}$$

From knowing the number of galaxy pairs with separation $\mathbf{r}$, the two point correlation could be computed, which will be the topic of discussion in the following section.

5.4 2–point correlation function

The 2-point correlation function (2PCF) is a simple statistical method to measure the fractional increase in finding two objects inside two volumes $\mathrm{d}^3\mathbf{x_1}$,$\mathrm{d}^3\mathbf{x_2}$,

separated by r relative to a random distribution:

$$P = \bar{n}^3[1 + \xi(r_{12,i})]\mathrm{d}^3\mathbf{x_1}\mathrm{d}^3\mathbf{x_2}, \tag{5.19}$$

where $\bar{n}^3$ is the mean density. To calculate $\xi(r_{12,i})$, we use the Landy–Szalay estimator [Landy and Szalay, 1993] defined as:

$$\xi(r_{12,i}) = \frac{DD(r_{12,i}) - 2DR(r_{12,i}) + RR(r_{12,i})}{RR(r_{12,i})}, \tag{5.20}$$

where D denotes a galaxy in the data catalogue and R denotes a galaxy in a random catalogue. Therefore $DD(r_{12,i})$ represents the number of pairs of galaxies in the data catalogue with separation $r_{12,i}$ (i denoting the i-th bin) normalized over the total number of pairs. The galaxy correlation function was first approximated to be a power law taking the form:

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}. \tag{5.21}$$

However, with more recent measurements, it has been shown that it deviates away from such form, due to the strong amplitude increase from the galaxy clustering within the same halo, i.e the 1-halo term (see [Coil, 2013], [Zehavi et al., 2011]).

Separation calculated using galaxy redshifts contains systematic errors as the line of sight motions (peculiar motions) of galaxies alter the inferred redshifts. This contamination effect is called redshift space distortion and creates an effect called "Finger of god", where structures observed point directly at the observers direction (see Figure 5–1). Since our position in the universe is not a privileged one, this is an observational artifact. When analyzing the structures in the universe, this could

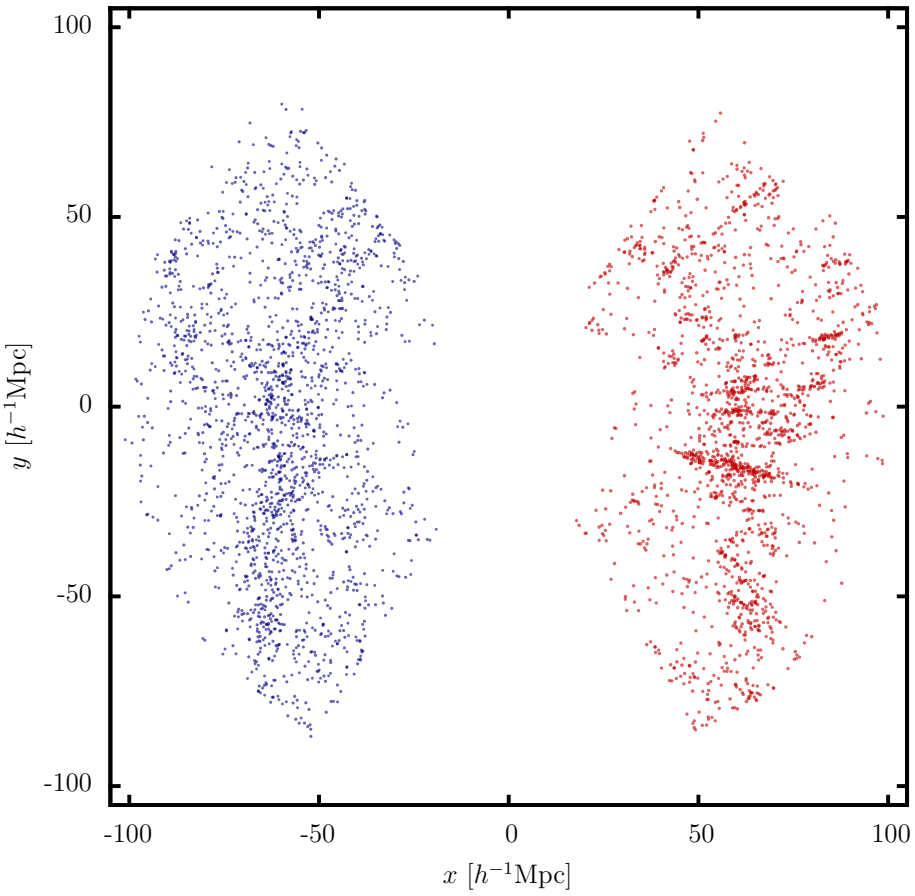

Figure 5–1: **Navy:** Star forming galaxies in sample SDSS-L1S. **Dark red:** Non-star forming galaxies in SDSS-L1Q flipped across the y-axis. The "Finger of god" effect is clearly more noticeable for the non-star forming samples. This is caused by Doppler shift from the peculiar velocity of the galaxy, which is greater for galaxies which are in cluster-like environments. Such effects are also seen in other studies (see for example [Madgwick et al., 2003]).

be avoided by using *projected* correlation functions. We first define redshift space separations $\vec{s} = \vec{v_1} - \vec{v_2}$, mean distance to the pairs $\vec{l} = \frac{1}{2}(\vec{v_1} + \vec{v_2})$ and call the line of sight separations

$$\vec{\pi} = \frac{\vec{s} \cdot \vec{l}}{|\vec{l}|}, \tag{5.22}$$

and the perpendicular separations

$$\vec{r_{\rm p}} = \sqrt{\vec{s} \cdot \vec{s} - \vec{\pi}^2}. \tag{5.23}$$

We then compute the correlation functions binned in perpendicular and parallel components:

$$w_{\rm p}(r_{\rm p}) \equiv 2 \int_0^{\pi_{\rm max}} {\rm d}\pi \xi(r_{\rm p}, \pi) \sim 2 \sum_0^{\pi_{\rm max}} \Delta\pi \xi(r_{\rm p}, \pi), \tag{5.24}$$

where ideally $\pi_{\rm max} = \infty$, but instead, we apply a limit at $\pi_{\rm max}$ to avoid signal from uncorrelated structures and use a bin width of $\Delta\pi = 1\ h^{-1}$Mpc. The value of the cut-off often ranges between $20-40\ h^{-1}$Mpc amongst various studies and thus testing is required to determine the adequate value.

The uncertainties on the measurements are computed by using the jack-knife method. This method is required since, a priori, we do not know the degree of correlation between the bins. Jack-knife subsamples of the estimator $w_{\rm p}(r_{{\rm p},i})^*_j$ are produced by masking a fraction of the catalogue, and the uncertainties are calculated by:

$$\sigma^2_{w_{\rm p}} = \frac{N-1}{N} \sum_j^N (w_{\rm p}(r_{{\rm p},i})^*_j - \overline{w_{\rm p}(r_{{\rm p},i})})^2, \tag{5.25}$$

where i and j correspond to the i-th bin and j-th jack-knife sample. In practice, this is done by dividing the field into grids and masking one [Zehavi et al., 2002].

The advantage of this approach is that it measures the variance between different spatial regions and the covariance matrix can be calculated without arduous work [McBride et al., 2011b].

Due to the finite size of the survey we must also consider the bias due to the survey size to help take into account that the survey may be sampling an over-dense or under-dense region of the universe. Observational surveys minimize this by surveying a very large fieeled and/or many distinct and well separated fields. This finite size effect could be corrected for by adding a constant known as the integral constraint (IC):

$$w_{\rm p,true}(r_{{\rm p},i}) = w_{\rm p,obs}(r_{{\rm p},i}) + {\rm IC}. \tag{5.26}$$

To estimate the integral constraint, we first assume that $w_{\rm p}(r_{\rm p})$ follows a power-law of the form $Ar_{\rm p}^{-\beta}$ at first-order. We then estimate this constant as:

$$\begin{aligned} {\rm IC} &= \frac{1}{V^2}\int_1\int_2 w_{\rm p,true}(r_{{\rm p},i}){\rm d}^3\mathbf{x}_1{\rm d}^3\mathbf{x}_2 & (5.27)\\ &= \frac{\sum_i RR(r_{{\rm p},i})w_{\rm p,true}(r_{{\rm p},i})}{\sum_i RR(r_{{\rm p},i})} & (5.28)\\ &= \frac{\sum_i RR(r_{{\rm p},i})A_{w_{\rm p}}r_{{\rm p},i}^{-\beta}}{\sum_i RR(r_{{\rm p},i})}. & (5.29) \end{aligned}$$

In practice, we let the $A_{w_{\rm p}}$ and β vary for various values in the fiducial model

$$w_{\rm p,model} = A_{w_{\rm p}}r_{\rm p}^{-\beta} - \frac{\sum_i RR(r_{{\rm p},i})A_{w_{\rm p}}r_{{\rm p},i}^{-\beta}}{\sum_i RR(r_{{\rm p},i})} \tag{5.30}$$

and use χ^2 minimization

$$\chi^2 = \sum_i \frac{1}{\sigma^2_{w_\mathrm{p}}} (w_{\mathrm{p,true}}(r_{\mathrm{p},i}) - w_{\mathrm{p,model}}(r_{\mathrm{p},i}))^2 \tag{5.31}$$

to determine the best fit parameters A_{w_p} and β which are then used to compute the integral constraint [Lee et al., 2006].

5.5 3–point correlation function

The 2PCF is the most commonly used measurement to quantify galaxy clustering in data. The next order, the 3-point correlation (3PCF) function is the lowest order to quantify non-Gaussian features in the geometric distribution of the data, which naturally arise in the nonlinear evolution of density fluctuations, even in an initially Gaussian distributed density field [Guo et al., 2013]. Since cosmic string wakes would produce non-Gaussian features, the 3-point correlation function is a useful tool in identifying these structures. The 3-point correlation functions are considerably harder to compute, as the number of possible triplets in the data increases as N^3, where N is the number of data points. Therefore, not many studies have been carried out to date.

Similar to the 2PCF, the probability of finding three objects in $\mathrm{d}^3\mathbf{x}_1, \mathrm{d}^3\mathbf{x}_2, \mathrm{d}^3\mathbf{x}_3$ is defined by:

$$\begin{aligned} P = \bar{n}^3[1 + \xi(r_{12}) + \xi(r_{23}) + \xi(r_{31}) \\ + \zeta(r_{12}, r_{23}, r_{31})]]]\mathrm{d}^3\mathbf{x}_1\mathrm{d}^3\mathbf{x}_2\mathrm{d}^3\mathbf{x}_3 \end{aligned} \tag{5.32}$$

where $\bar{n}$ is the mean number density, $\xi(r)$ is the 2PCF and ζ is the 3-point correlation function. The 3PCF has a hierarchical form

$$\zeta(r_{12}, r_{23}, r_{31}) = Q[\xi(r_{12})\xi(r_{23}) + \xi(r_{23})\xi(r_{31}) + \xi(r_{31})\xi(r_{12})] \tag{5.33}$$

with Q being some normalization parameter [Takada and Jain, 2003]. This normalization factor Q, defined as:

$$Q(r_{12}, r_{13}, r_{31}) \equiv \frac{\zeta(r_{12}, r_{13}, r_{31})}{\xi(r_{12})\xi(r_{23}) + \xi(r_{23})\xi(r_{31}) + \xi(r_{31})\xi(r_{12})}, \tag{5.34}$$

is of order unity on all scales and is commonly referred to as the reduced 3PCF (r3PCF). $\zeta(r_{12}, r_{23}, r_{31})$ can be estimated using the Szapudi & Szalay estimator [Szapudi and Szalay, 1998]

$$\zeta(r_{12}, r_{23}, r_{31}) = \\ \frac{DDD(r_{12}, r_{23}, r_{31}) - 3DDR(r_{12}, r_{23}, r_{31}) + 3DRR(r_{12}, r_{23}, r_{31}) - RRR(r_{12}, r_{23}, r_{31})}{RRR(r_{12}, r_{23}, r_{31})}, \tag{5.35}$$

where all the terms are normalized by the total number of unique triplets.

Similar to the projected 2PCF, the projected 3PCF $\Pi(r_{\rm p12}, r_{\rm p23}, r_{\rm p31})$ is defined as [Jing and Börner, 2004]:

$$\Pi(r_{\rm p12}, r_{\rm p23}, r_{\rm p31}) = \int\int \zeta(r_{\rm p12}, r_{\rm p23}, r_{\rm p31}, \pi_{12}, \pi_{23}) {\rm d}\pi_{12} {\rm d}\pi_{23} \tag{5.36}$$

and

$$Q_{\rm p}(r_{\rm p12}, r_{\rm p23}, r_{\rm p31}) = \frac{\Pi(r_{\rm p12}, r_{\rm p23}, r_{\rm p31})}{w_{\rm p}(r_{12})w_{\rm p}(r_{13}) + w_{\rm p}(r_{12})w_{\rm p}(r_{23}) + w_{\rm p}(r_{23})w_{\rm p}(r_{31})} \tag{5.37}$$

is the projected and reduced 3PCF (pr3PCF), which requires 5 parameters, where $r_{\mathrm{p}ij}$ and π_{ij} are the separations between objects i and j, in the perpendicular and line-of-sight directions. For the integral, we integrate over the line-of-sight up to r_{max} again for the two sides of the triangle, using a bin width of $\Delta\pi = 1$ h^{-1}Mpc (see Figure 5–2 for schematic illustrations).

As noted previously, computing the 3PCF is a computationally intensive task for large catalogues. Other groups have used efficient tree-searching codes such as Ntropy-npoint code [Gardner et al., 2007],[McBride et al., 2011b] or NPT [Moore et al., 2001], [Marín, 2011]. Instead of employing these codes, we use an algorithm that exclusively searches for triplets with the configuration that we seek.

Two types of binning schemes are generally employed in computing the 3PCF. The first and traditional method is to define the lengths of the two sides and bin linearly in θ to obtain the third length with a fixed bin width. Alternatively, one could do use the same binning scheme but with bin widths $r_{ij} = f \times r_{ij}$, where f is a fraction usually in the range 0.01 – 0.1. As pointed out in [Marín, 2011] using the former binning scheme allows for a larger scatter, and inevitably contains triplets with a wider range of shapes in each bin. Using the latter binning scheme however has the consequence of having a damped signal at the higher θ due to the over lap of bins (the same triangle being in multiple bins) as noted by [McBride et al., 2011b]. Since we are trying to identify the difference, this should pose no problem as long as the same binning schemes are used. However, caution must be taken when comparing results

between papers using different binning schemes [Guo et al., 2013]. To quantify the correlation between the bins, we compute the covariance matrix to probe the effect.

The covariance matrix describing the correlation between the bins can be calculated from the jack-knife samples without additional intensive work by computing:

$$\Delta_i^k = X_i^k - \overline{X_i}, \tag{5.38}$$

and then:

$$\mathcal{C}_{ij}^{(jack)} = \frac{(N-1)^2}{N}\mathcal{C}_{ij} = \frac{N-1}{N}\sum_{k=1}^{N}\Delta_i^k\Delta_j^k, \tag{5.39}$$

where X_i^k is the k-th jack-knife realization for the i-th bin of pr3PCF Q, r3PCF Q, p3PCF Π, 3PCF ζ, 2PCF ξ or p2PCF $w_{\rm p}$, and $\overline{X_i}$ is the jack-knife average of the quantity. $\mathcal{C}_{ij}^{(jack)}$ here describes the unbiased estimator of the covariance between the bins. Not only will this covariance matrix be used to illustrate the degree of correlation between the uncertainties in the bins, but it will also be used to compute the χ^2 fitting.

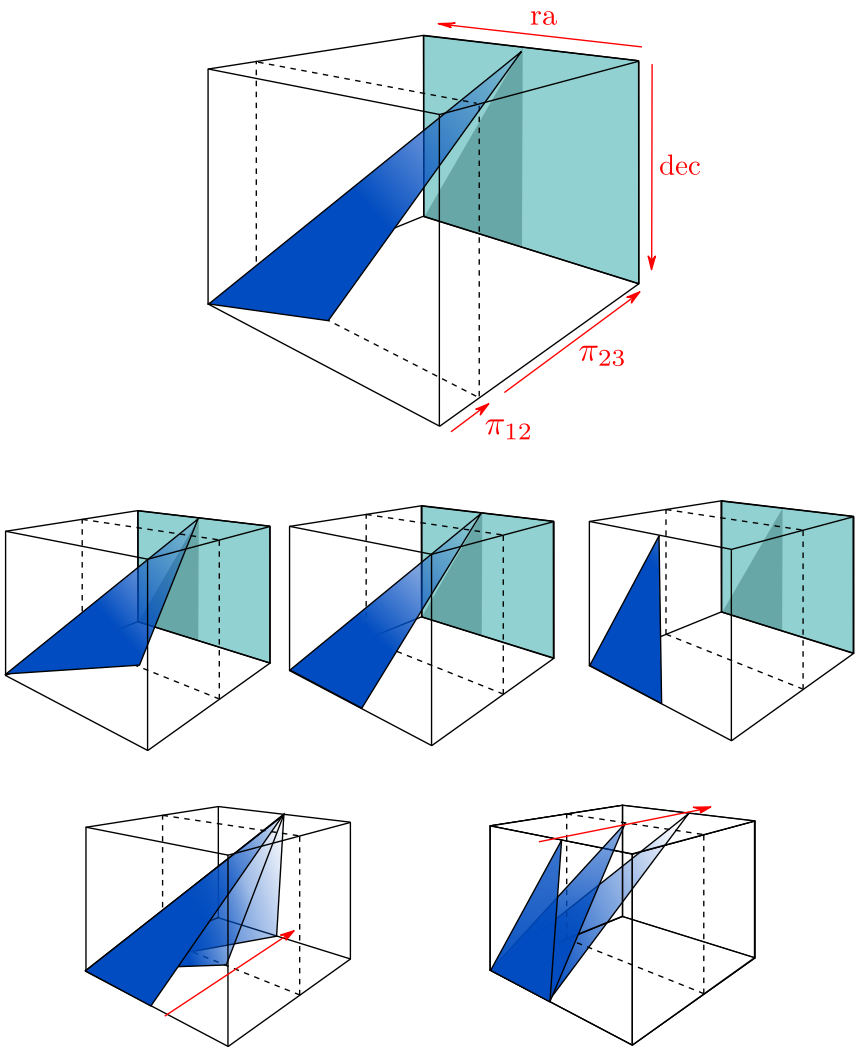

Figure 5–2: Projections taken to compute the 3PCFs. **Top row:** Direction and perspective of triangles. **Middle row:** All three triangles have the same projection (shown as a shadow on the back light shaded face) and are hence considered to possess the same configuration in projected space. **Bottom row:** Integrating over π_{12} and π_{23} for the same projected configuration.

CHAPTER 6
RESULTS

6.1 Simulations

Each simulation took approximately 24 hours using 128 cores in parallel using the supercomputer Guillimin, and the halo finding took additional 12 hours. Snapshots of three different cosmic string tensions at $z = 0$ and $z = 3$ are shown in Figure 6–1, and dark matter density contours are shown in Figure 6–2. The numbers of halos found in each of the simulations are summarized in Table 6–1. Simulations for each model were produced twice using the same parameters but different seeds to remove variance particular to the seed. In both runs, the excess of halos in the $G\mu = 10^{-5}$ model was seen at $z = 10$. However, at $z = 0$, the differences in the total number of halos between the models were within 0.1%.

6.2 Halo mass function

The halo mass functions are shown in Figure 6–3, and shows consistency between the models up to $z = 10$, at which point the $G\mu = 10^{-5}$ model finds excess halos. Since this trait is not observed at lower redshifts, it is likely that these halos halt in growth at intermediate redshifts and become accreted onto larger halos, which form and grow from fluctuations. While a cosmic string tension of $G\mu = 10^{-5}$ has already been ruled out, this suggests that at higher redshifts, the halo mass function is likely

Snapshot redshift	$G\mu = 0$	$G\mu = 10^{-7}$	$G\mu = 10^{-5}$
$z = 10$	128	124	990
$z = 5$	44323	44609	45952
$z = 3$	146623	146744	147715
$z = 0$	290303	290032	290240

Snapshot redshift	$G\mu = 0$	$G\mu = 10^{-7}$	$G\mu = 10^{-5}$
$z = 10$	120	117	962
$z = 5$	43588	43780	45276
$z = 3$	145092	145333	146280
$z = 0$	288251	288260	288577

Table 6–1: Number of halos in each simulation at different redshifts for the two different seeds used (upper and lower table).

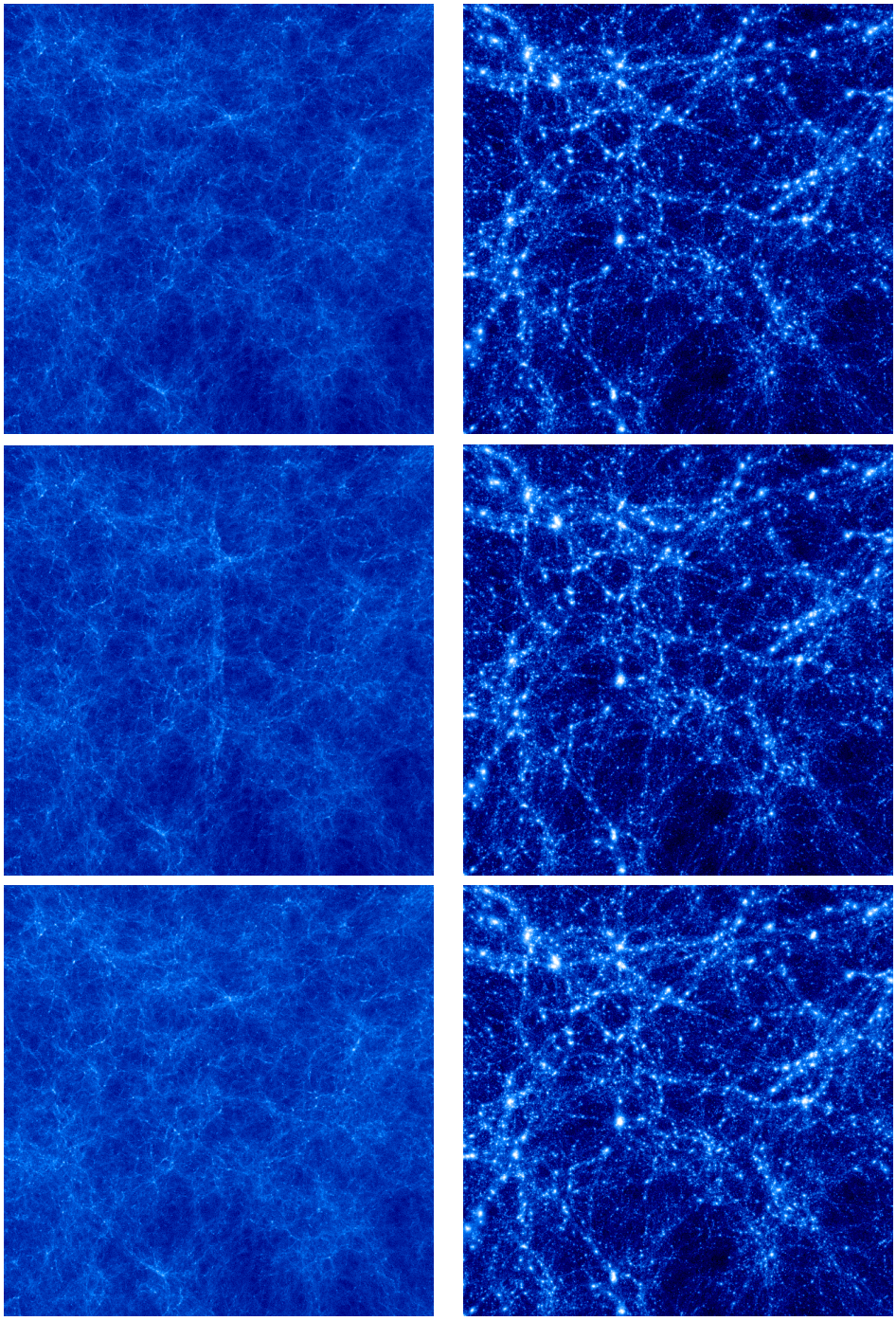

Figure 6–1: **Left 3 panels:** Simulation snapshots at $z = 3$. **Right 3 panels:** snapshots at $z = 0$ (top: $G\mu = 0$, center: $G\mu = 10^{-5}$, bottom: $G\mu = 10^{-7}$)

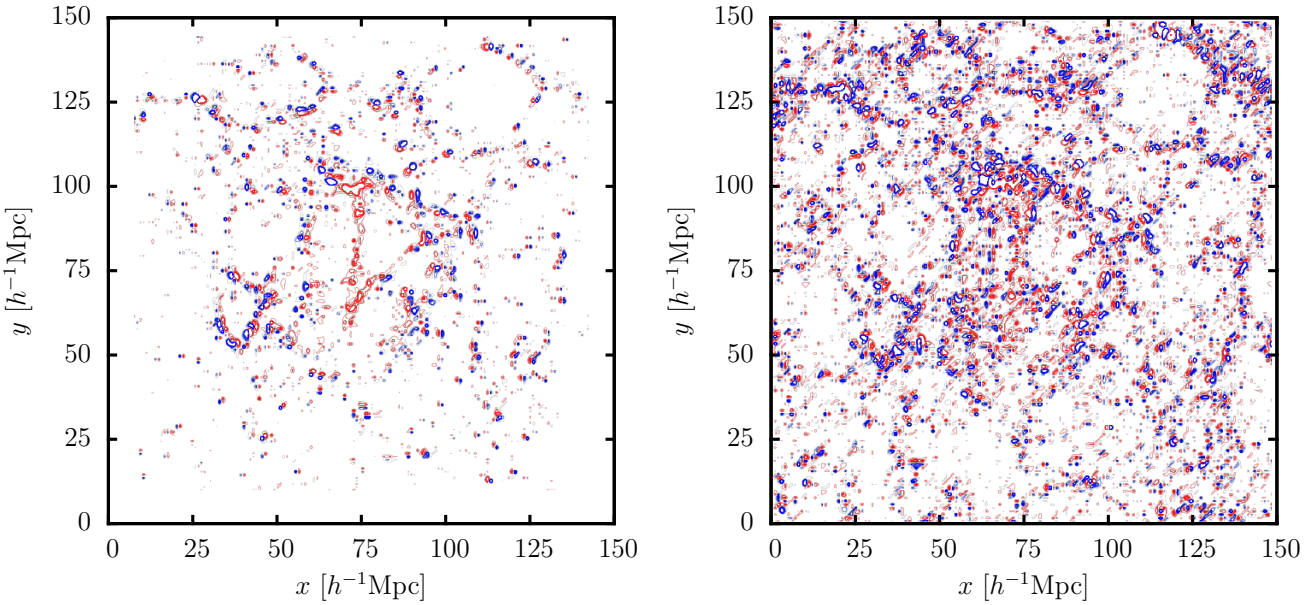

Figure 6–2: Density contours of dark matter particles at $z = 0$ for $G\mu = 10^{-5}$ (left) and $G\mu = 10^{-7}$ (right). The red contours show the regions of high density and the blue contours show the regions of low density relative to the $G\mu = 0$ model. The wake is placed at $x = 75$ h^{-1}Mpc and extends from $y = 50$ to 100 h^{-1}Mpc. Note that the density contrast has been increased up by a factor of 100 for the plot on the right.

to be different in the presence of cosmic strings with $G\mu < 10^{-5}$. However, we could expect that the difference would be small and the perturbations in structures to be small, which would require higher resolution simulations to resolve.

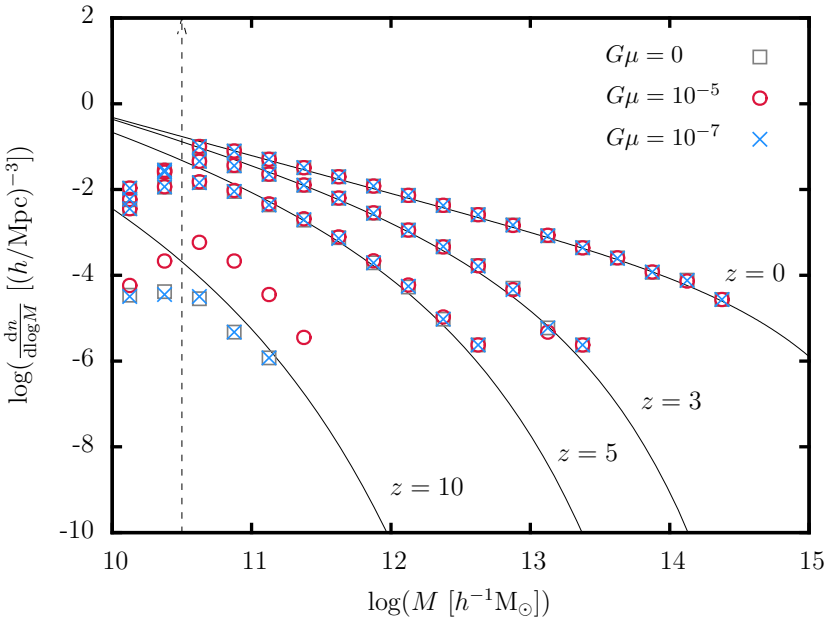

Figure 6–3: Halo mass functions at redshifts $z = 0$, $z = 3$, $z = 5$, $z = 10$. The points correspond to the different string tension and the solid black lines correspond to the Tinker 2008 mass functions [Tinker et al., 2008]. The halo finding resolution is at $\sim 10^{10.5}$ $\mathrm{M}_\odot$ (dotted line) and hence the code's ability to identify halos deceases below this limit.

6.3 Halo occupation density

The halo occupation density measures the average number of subhalos inside a parent halo. One would expect that the density would be higher for the $G\mu = 10^{-5}$ model, since the cosmic string introduces the highest and largest (in spatial dimensions) overdensity. This results in an increased probability of smaller halos

encountering another halo. At redshifts $z = 0, z = 3, z = 5$ the models are consistent with each other, although the amount of scatter increases towards higher z caused by the decrease in the number in halos present at those redshifts. As shown in Figure 6–4, the average number of subhalos in parent halos is similar at redshifts $z = 3$ and $z = 5$, with a characteristic "dip" at approximately $10^{11.5}$ $M_\odot$, which is not present at $z = 0$. Moreover, the average subhalo count *decreases* as a function of redshift.

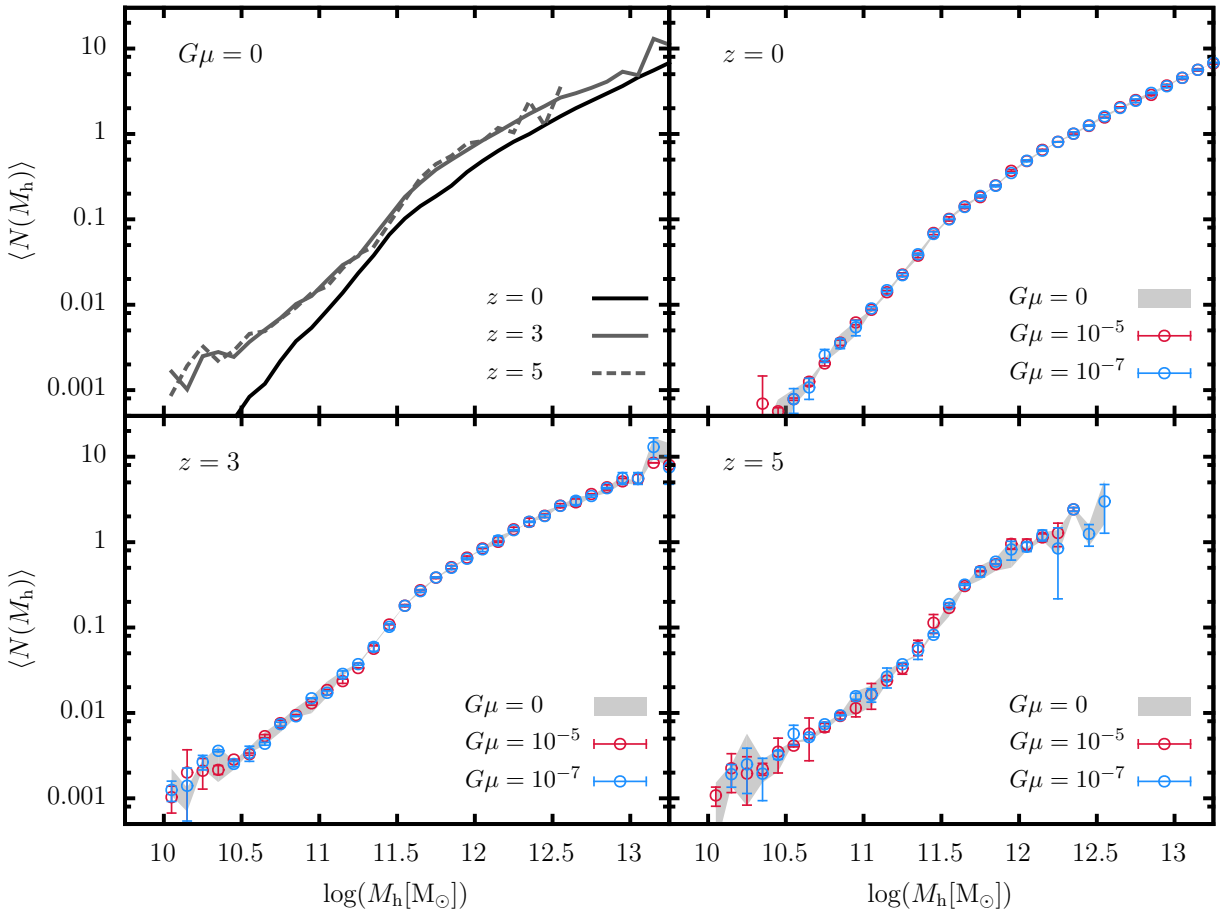

Figure 6–4: Halo occupation density at various redshifts for the three models.

6.4 Stellar mass function

The stellar mass functions were computed from the galaxies produced with semi-analytic modeling. Since there is a large uncertainty in the semi-analytic model, we used the stellar mass function to calibrate the simulation with observations. In doing so we made one assumption, namely that the stellar mass function is insensitive to the presence of cosmic strings. This assumption is fair since the difference in the stellar mass function produced from simulations with $G\mu = 0$, $G\mu = 10^{-5}$, $G\mu = 10^{-7}$ is less than the scatter from the measurement of the stellar mass function from SDSS, and that we have an existing constraint $G\mu < 10^{-5}$. Using the stellar mass function as the initial calibration, we further tuned other parameters, namely star formation rates, feedback reheating efficiency, feedback ejection efficiency and fraction of star formation mass instantaneously recycled back to cold gas. The results of such tuning are shown in appendix C and the results of the $1/V_{\max}$ calculation is shown in Figure 6–5. The galaxies were specifically chosen to be in the region $z < 0.1$, $124 < ra < 240$, $0 < dec < 56$, and the $V_{\max}$ measurements were directly extracted from the NYU-VAGC LSS *safe* samples, which determines the value by combining the luminosity of the galaxy and the specific flux limit in the direction of the galaxy.

6.5 2–point correlation function

The projected 2–point correlation functions (p2PCF) were calculated for all samples using random catalogues with 70,000 points, and averaging over 30 times. Using this catalogue size, the calculation took approximately 12 hours each using 8 cpu cores.

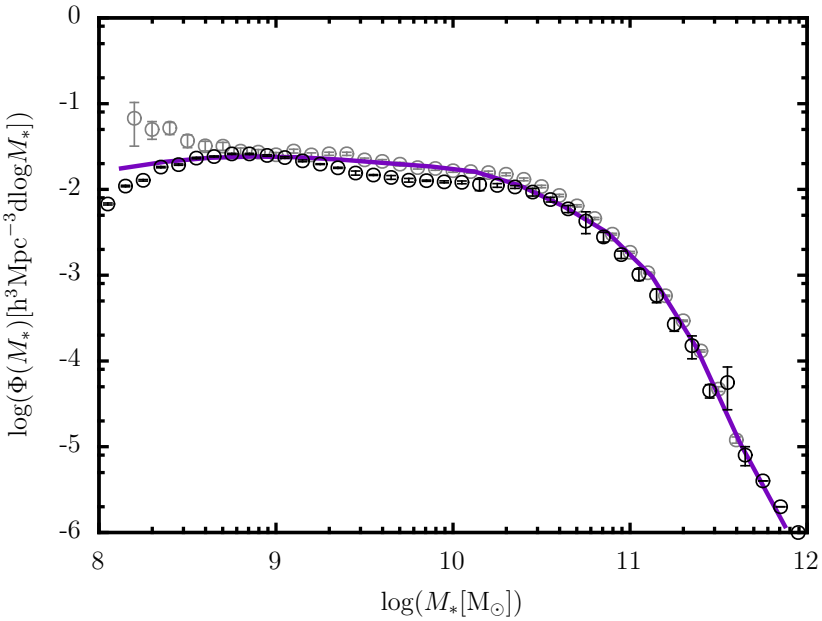

Figure 6–5: Stellar mass function at $z = 0$ for **Black:** SDSS-DR7 galaxies determined from $1/V_{\rm max}$, **Grey:** SDSS from previous data release (DR4) [Yang et al., 2009], **Magenta:** Galaxies from $G\mu = 0$ halos + SAGE. The points from Yang et al. have extra weighting assigned to each galaxy based on the completeness factor. This produces the lift in the stellar mass function at low mass scales.

We first computed the p2PCFs using different $\pi_{\rm max}$. As shown on Figure 6–6, on scales below $r_{\rm p} < 1\ h^{-1}$Mpc the choice has negligible effects. On higher scales, the differences between the cuts become apparent. We make a choice of $\pi_{\rm max} = 20\ h^{-1}$Mpc, which is long enough to include correlations between structures but not so long to include uncorrelated structures. Additionally, keeping $\pi_{\rm max}$ to the minimal value increases the computational speed due to the way our 3PCF calculation algorithm was coded. Note that it is important to employ the same $\pi_{\rm max}$ value when computing the p2PCF and the pr3PCF.

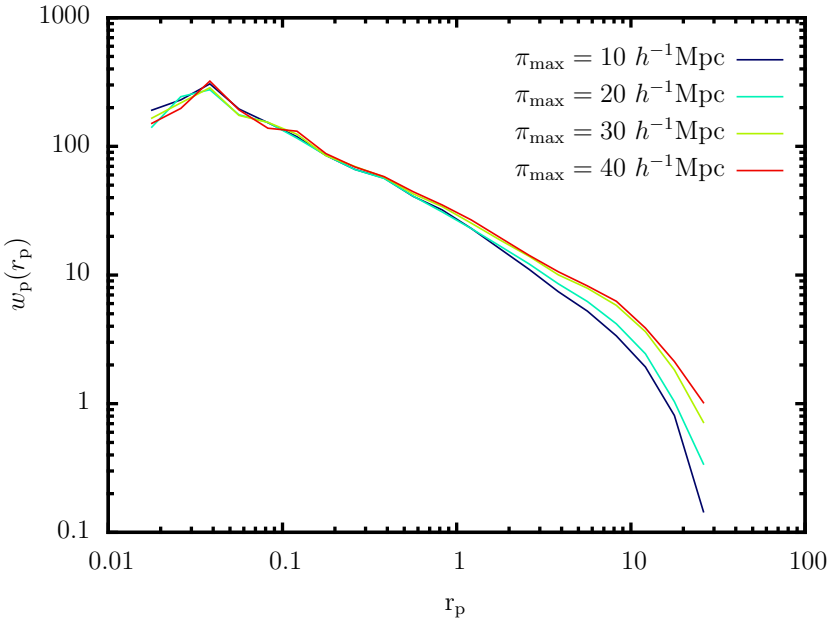

Figure 6–6: Projected 2PCFs for various $\pi_{\rm max}$.

The p2PCF for all the galaxies combined are shown in Figure 6–7. The spatial resolution of the simulations causes the uncertainties to be high for the bins

$w_{\rm p} < 0.1\ h^{-1}$Mpc. In general, galaxies with higher luminosity have higher amplitudes. This trend is very weakly seen, although it is within the range of uncertainty. In comparing the different string models, no significant discrepancies were identified. In comparing the models with the SDSS samples, the amplitudes were found to differ for the low L1 sample. However, the trend of lower correlation function amplitude for SHAM applied to simulation halos in low magnitude ranges is a known issue [Masaki et al., 2013].

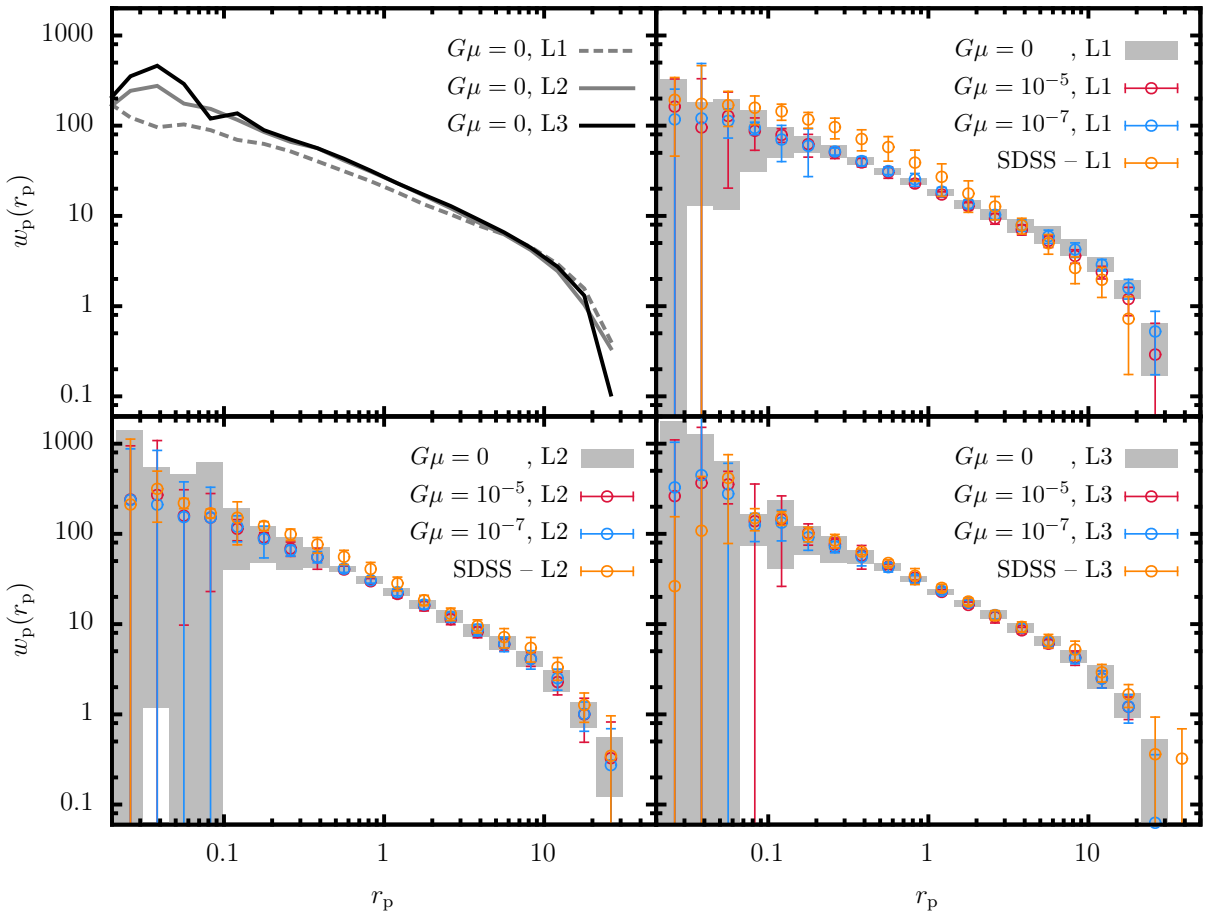

Figure 6–7: Projected 2–point correlation function of galaxies in luminosity class L1, L2 and L3.

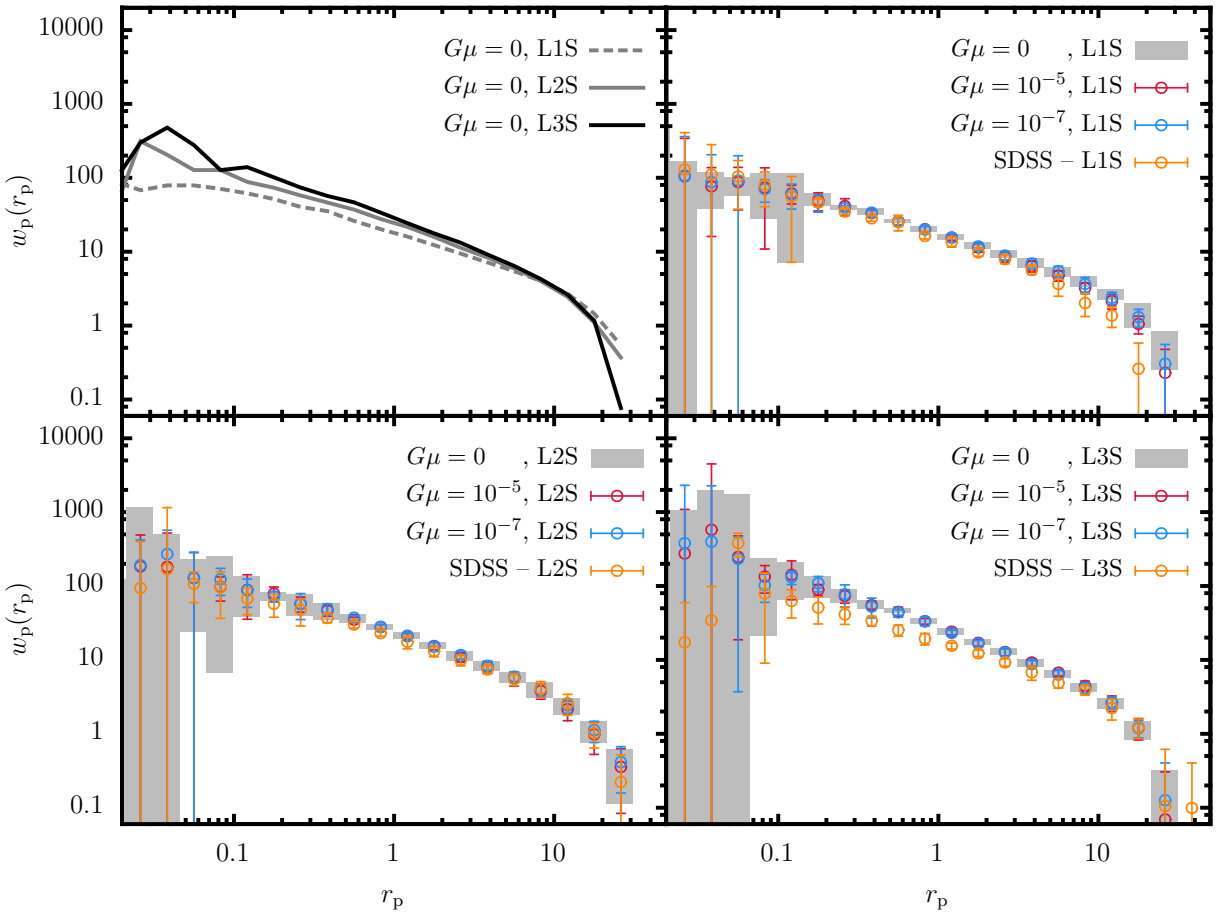

Figure 6–8: Projected 2–point correlation function of *star forming* galaxies in luminosity class L1, L2 and L3.

For the non–star forming galaxies, the string models were all in good agreement. When compared with the SDSS measurements, the amplitudes were slightly higher in the L2 and L3 samples but the overall slope was in good agreement. Therefore this

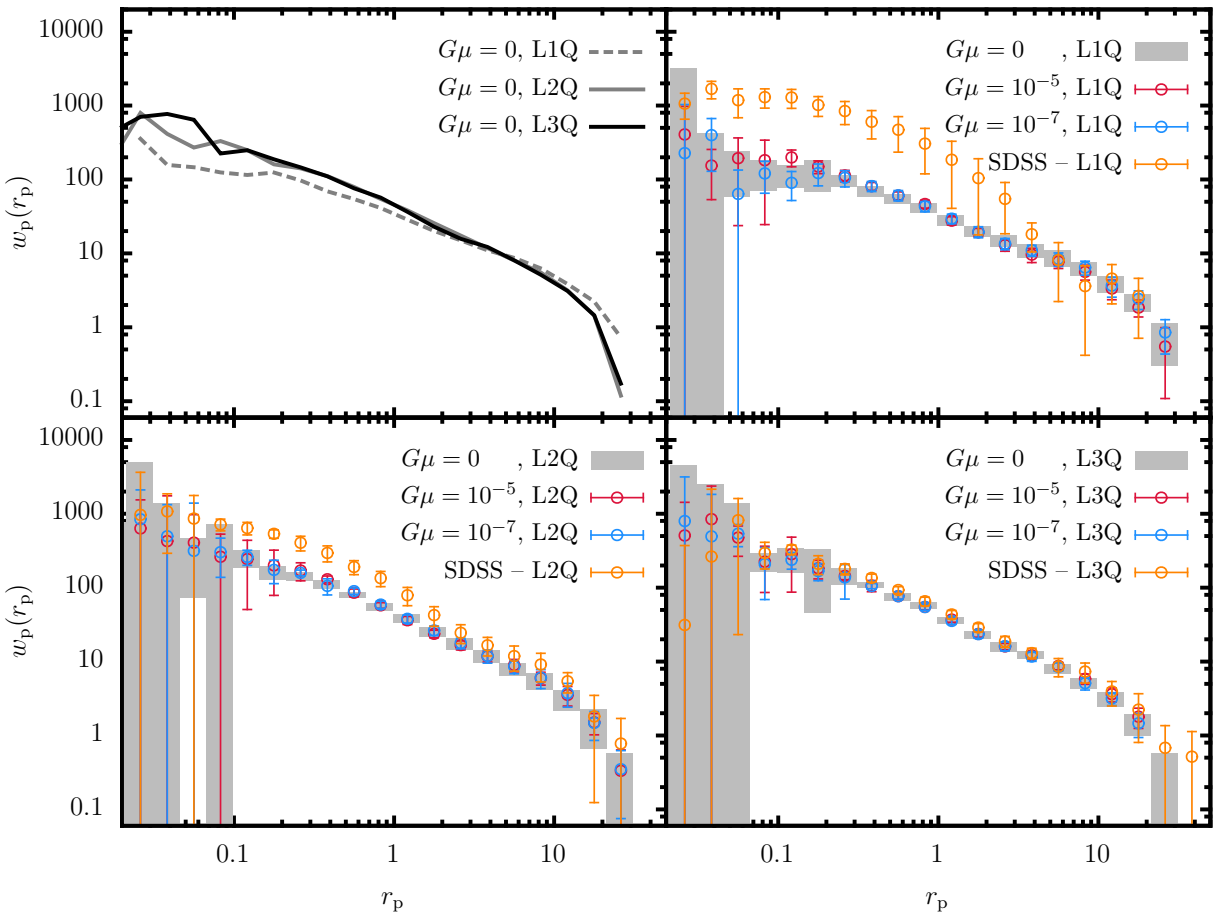

Figure 6–9: Projected 2–point correlation function of *non–star forming* galaxies in luminosity class L1, L2 and L3.

offset is likely due to the mis-approximation of the luminosity of the galaxies (SHAM is predicting a population of higher luminosities than SDSS). For the star forming galaxies, again the models were consistent with each other. However, in the luminosity classes L1 and L2, the amplitudes were lower than in the SDSS samples. The weak "hump" below 1 h^{-1}Mpc in the SDSS sample is identified in other studies also, although the origin of such clustering characteristic is not known [Masaki et al., 2013].

Since none of these measures show a significant difference between the string models, it is likely that cosmic strings do not, or very weakly affect the size of galaxy clustering.

6.6 3–point correlation function

Although no notable differences were found using the p2PCF, this does not mean that there is no preferred structural shape. For example, p2PCF is in general insensitive to the difference in structure between a connected bead like structure and a filamentary structure, since the separation between galaxies is similar in both cases.

Firstly, pr3PCF for all luminosity class and type were computed using $r_1 = 3, r_2 = 6$ h^{-1}Mpc. The calculation for non-star forming galaxies in luminosity class L1 were found to contain too much noise to comment on. In other luminosity and galaxy type subsamples, the models were found to be in fair agreement. One of the notable difference between the models is the amplitude in the star forming-L3 sample (bottom right panel of Figure 6–12), where the amplitudes of pr3PCF for $G\mu = 10^{-5}$ and $G\mu = 10^{-7}$ are slightly higher than that of $G\mu = 0$ (although within

1σ). Similarly, both the models with strings had a sharper "V" shape in the star forming L3 samples. However, due to the relatively large uncertainties (relative to 2PCF), the uncertainties must be reduced in order to constrain $G\mu$. Additionally, different configuration using $r_1 = 1, r_2 = 6\ h^{-1}$Mpc were computed. However, the results contained too much noise (much worse than the L1Q samples) and hence are not shown here.

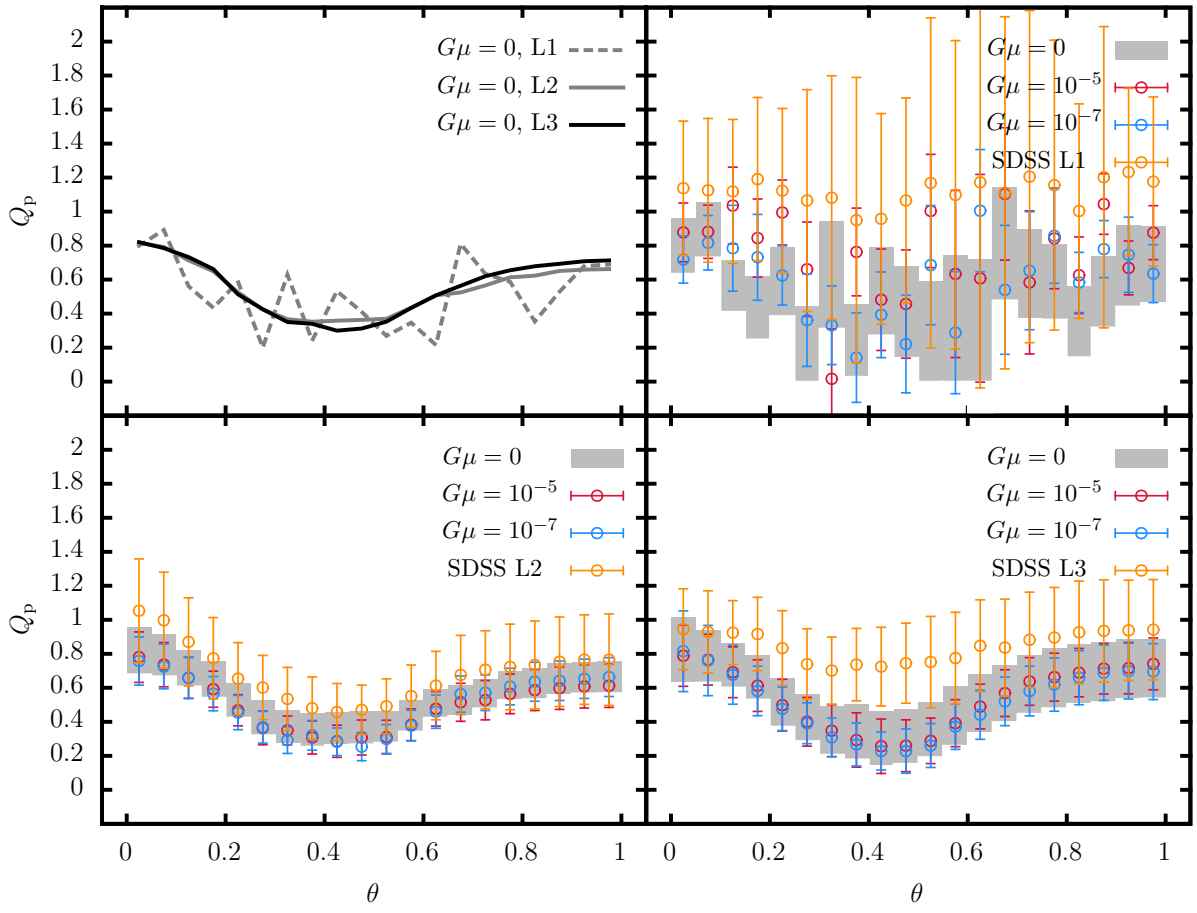

Figure 6–10: pr3PCF of *all* galaxies in luminosity class L1, L2 and L3.

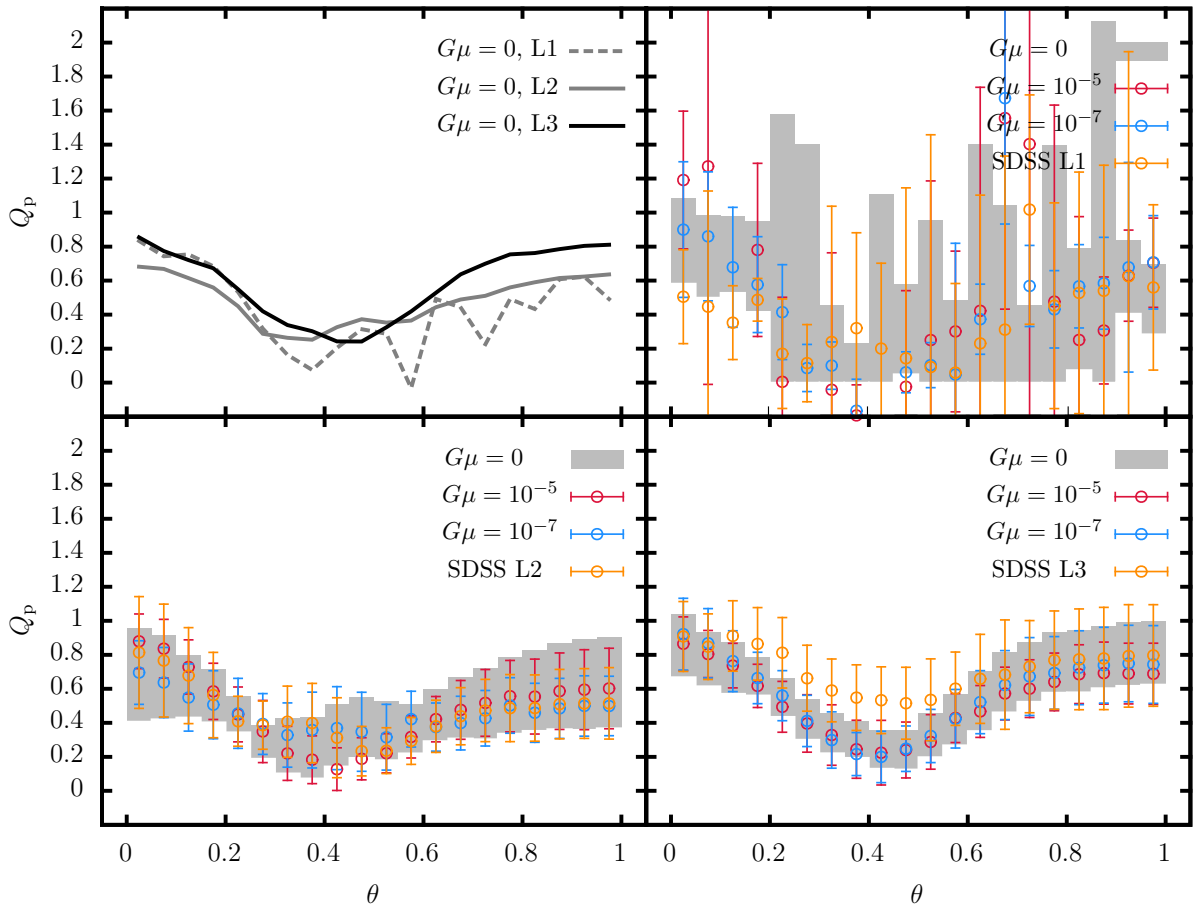

Figure 6–11: pr3PCF function of *non-star forming* galaxies in luminosity class L1, L2 and L3.

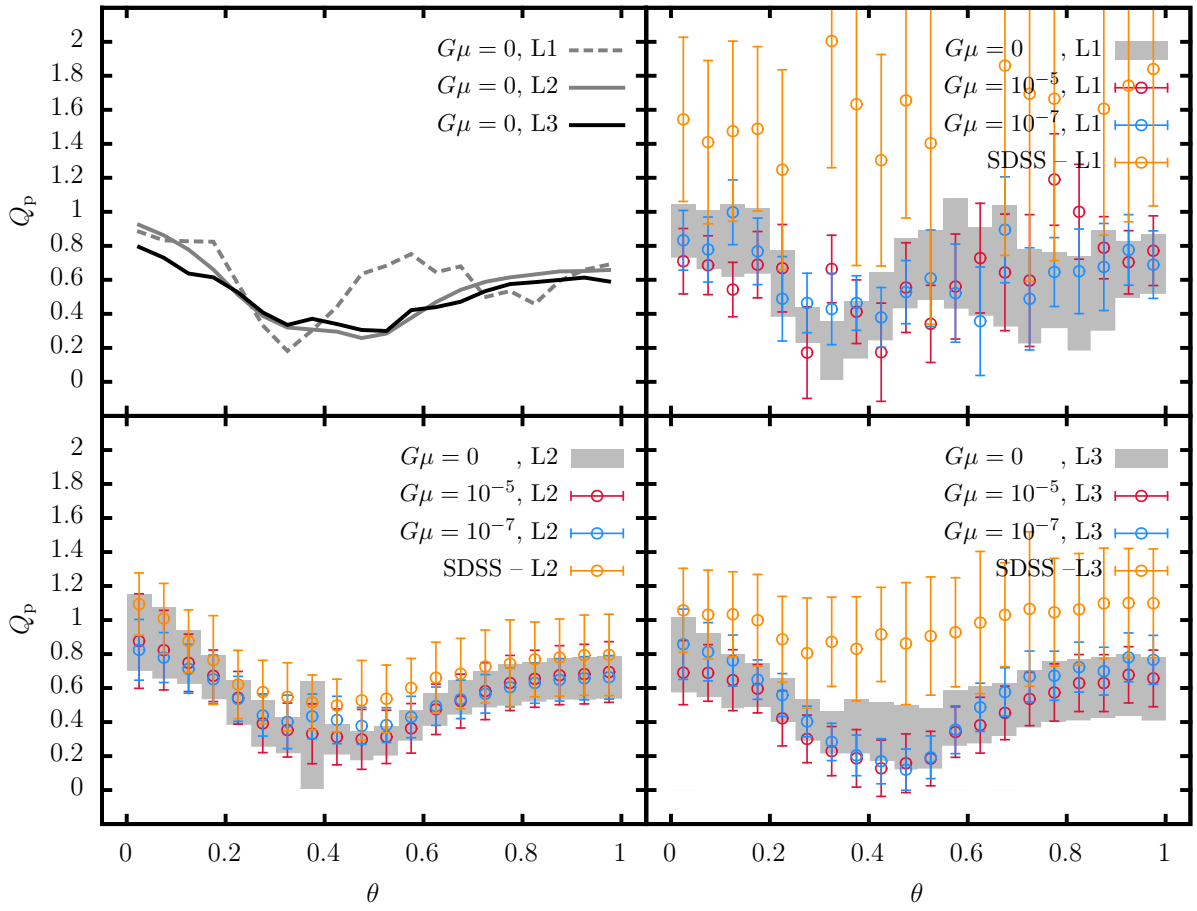

Figure 6–12: pr3PCF of *star forming* galaxies in luminosity class L1, L2 and L3.

Since the pr3PCF depends both on the projected-3PCF Π and projected-2PCF w_p it is also beneficial to only look for Π, as shown on Figures 6–13, 6–14, 6–15. The

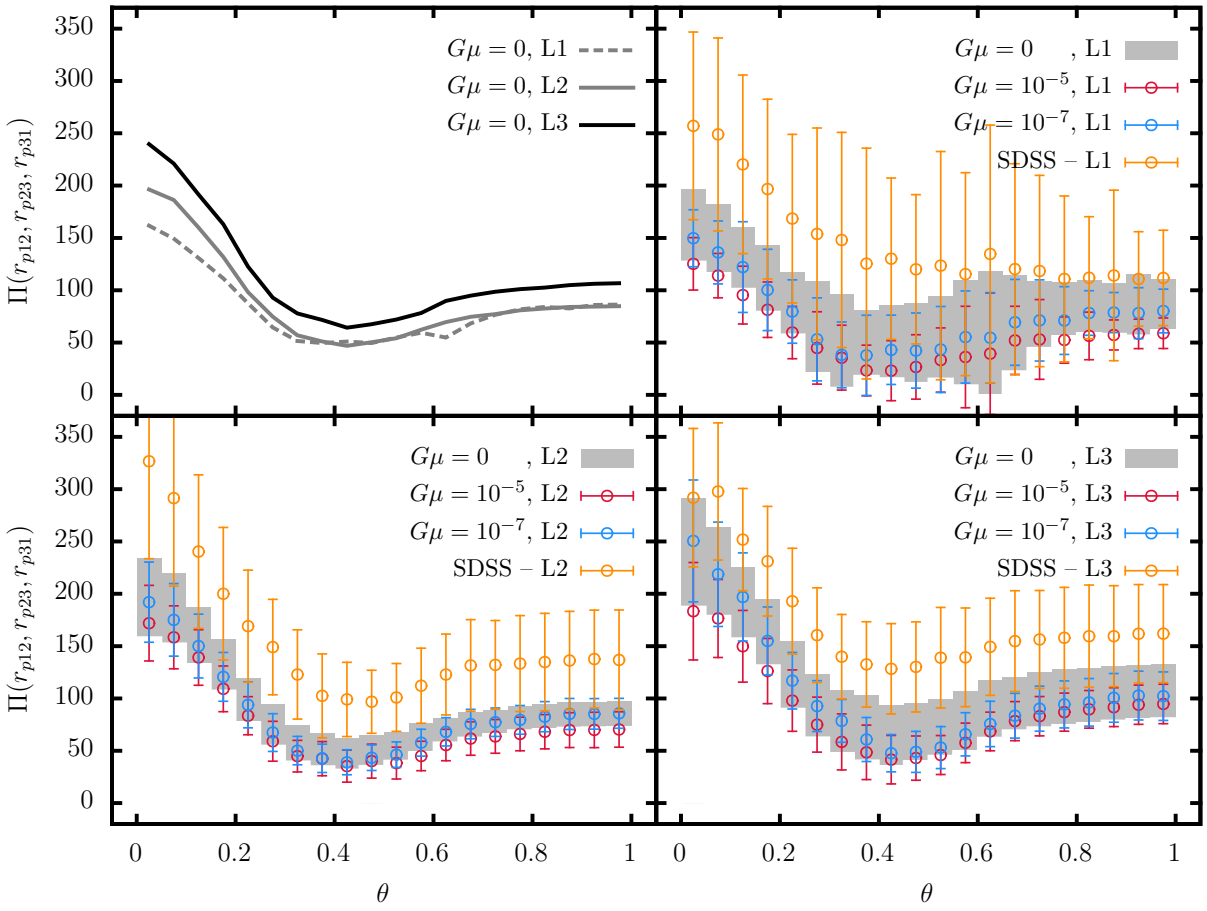

Figure 6–13: p3PCF of *all* galaxies in luminosity class L1, L2 and L3.

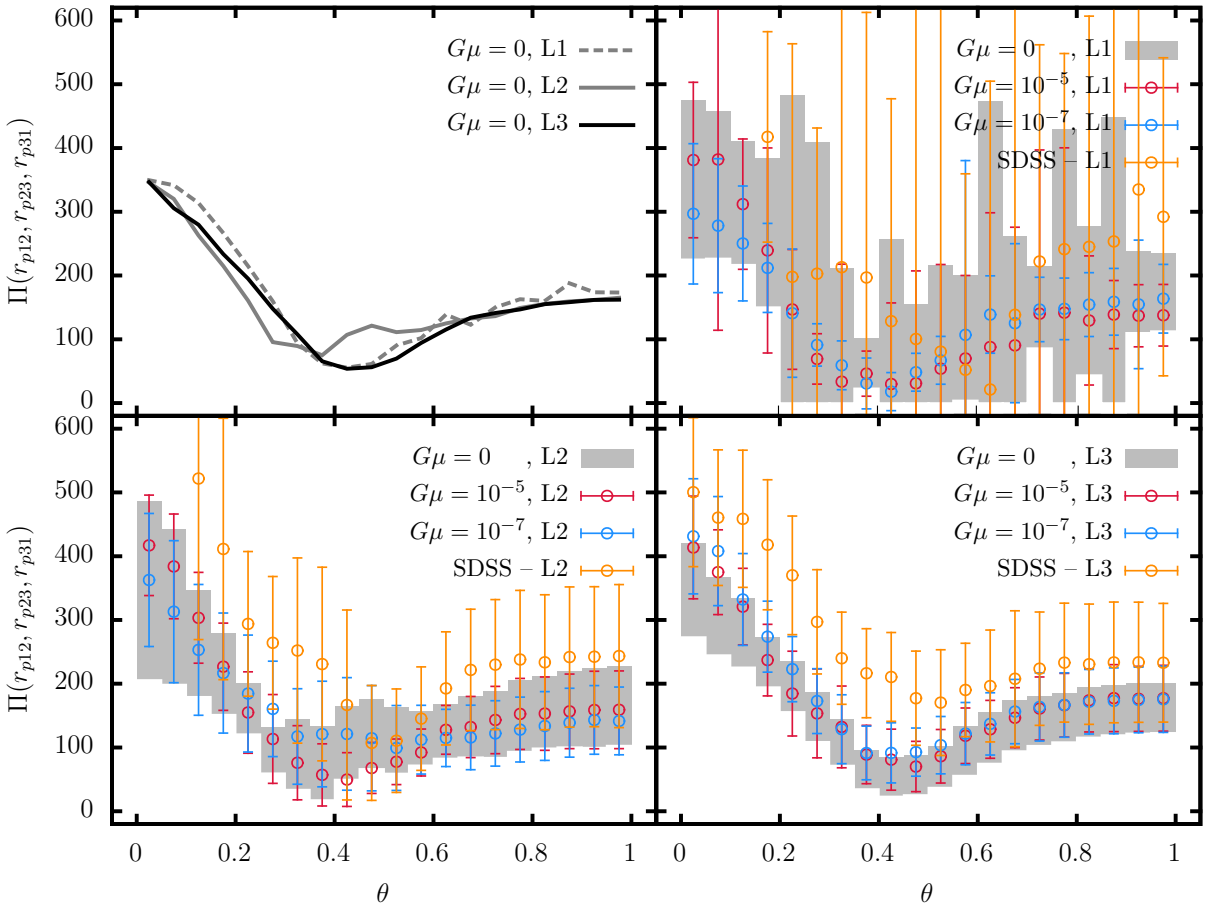

Figure 6–14: p3PCF of *non-star forming* galaxies in luminosity class L1, L2 and L3.

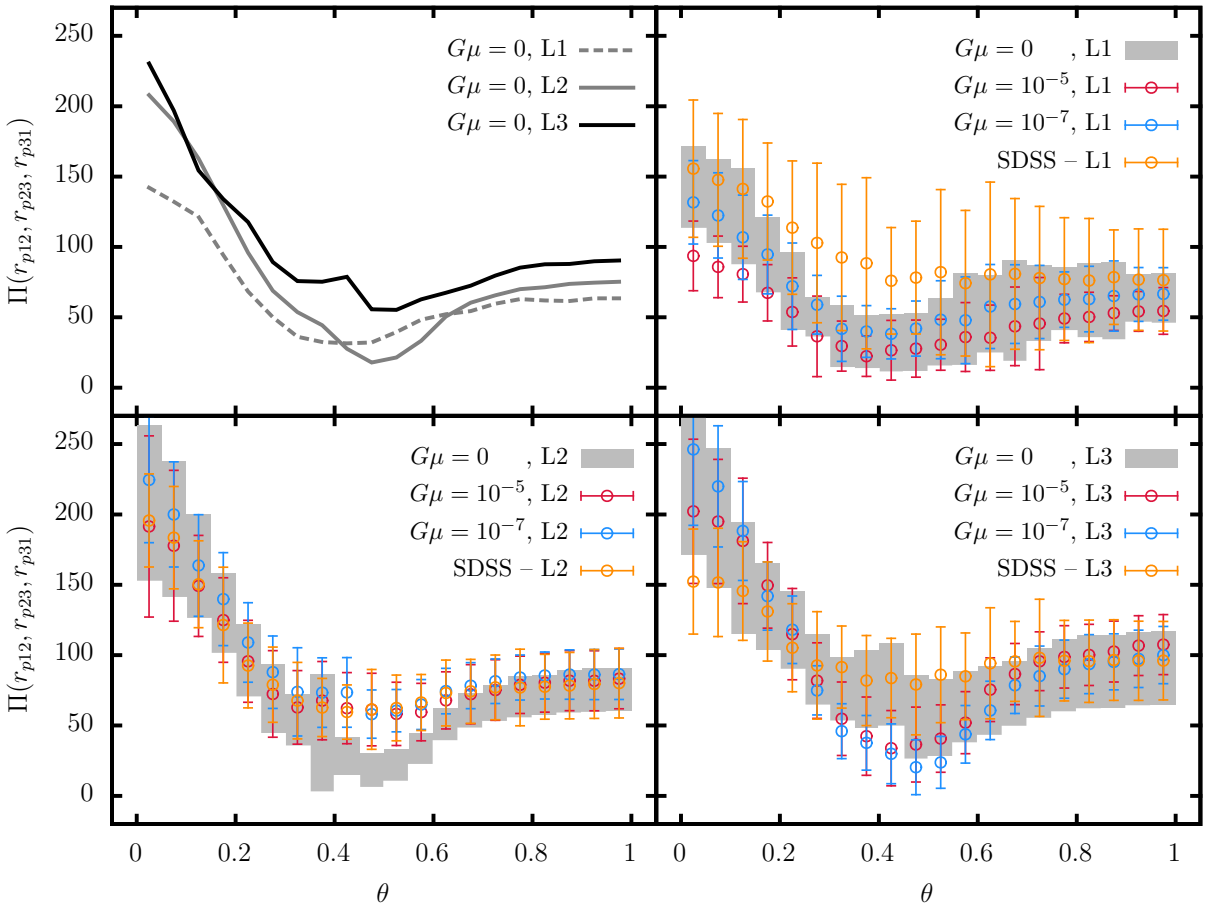

Figure 6–15: p3PCF of *star forming* galaxies in luminosity class L1, L2 and L3.

results again show that the SDSS clustering is stronger than the models. However, there are two interesting features that we observe in the results in star forming galaxies in luminosity class L2, L3 (bottom 2 panels of Figure 6–15). Firstly in L2, the $G\mu = 0$ model differs from the $G\mu = 10^{-5}$ and the $G\mu = 10^{-7}$ models between $0.4 < \theta < 0.6$, which is in agreement with the SDSS also. In L3, we see the opposite trend where the $G\mu = 10^{-5}$ and $G\mu = 10^{-7}$ are higher in amplitude than $G\mu = 0$. This suggests that cosmic strings do make a difference in the structures formed, *at least for some types of galaxies.*

Although, the measurements contained large uncertainties, we also tested for goodness of fit using the modified χ^2 statistics. If we assume that the measured Π for SDSS is the true, then :

$$\chi^2 = \sum_{ij} (\Pi_{G\mu}(r_{12}, r_{23}, r_{32,i}) - \Pi_{\rm SDSS}(r_{12}, r_{23}, r_{32,i}))^{\top} \mathcal{C}_{ij}^{-1} (\Pi_{G\mu}(r_{12}, r_{23}, r_{32,j}) - \Pi_{\rm SDSS}(r_{12}, r_{23}, r_{32,j})) \tag{6.1}$$

where $\mathcal{C}_{ij}^{-1}$ is the inverse covariance matrix (shown in Figure 6–17) determined from jackknife sampling. Such modification from the χ^2 calculation is required since the uncertainties between the bins are correlated. The χ^2 was then averaged for each string tension to test for the most realistic model. Using this, it was found that no string tension is strongly preferred as shown in Figure 6–16.

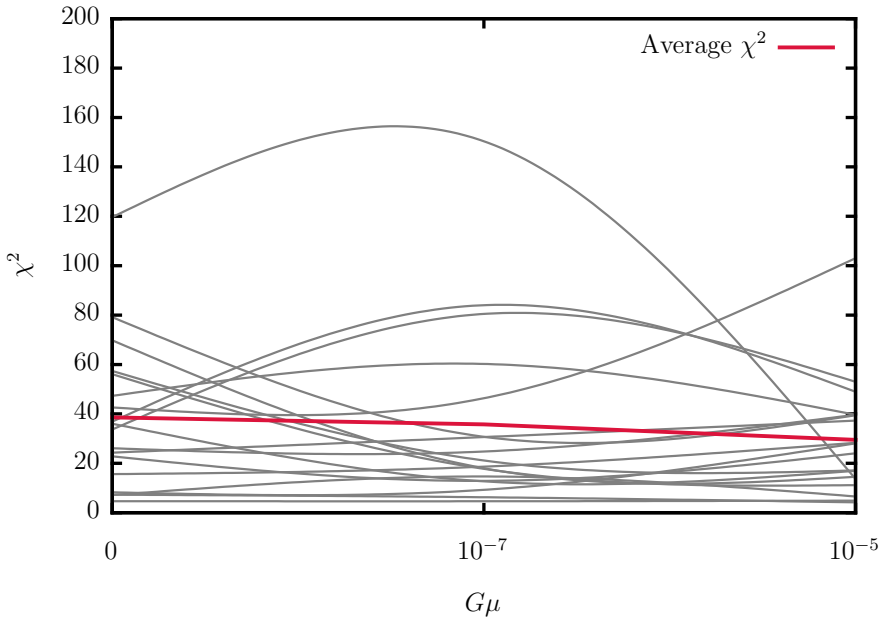

Figure 6–16: χ^2 for different subsamples treating SDSS measurements as the true model. Overall the χ^2 is high and no strong preferred value for $G\mu$ is observed.

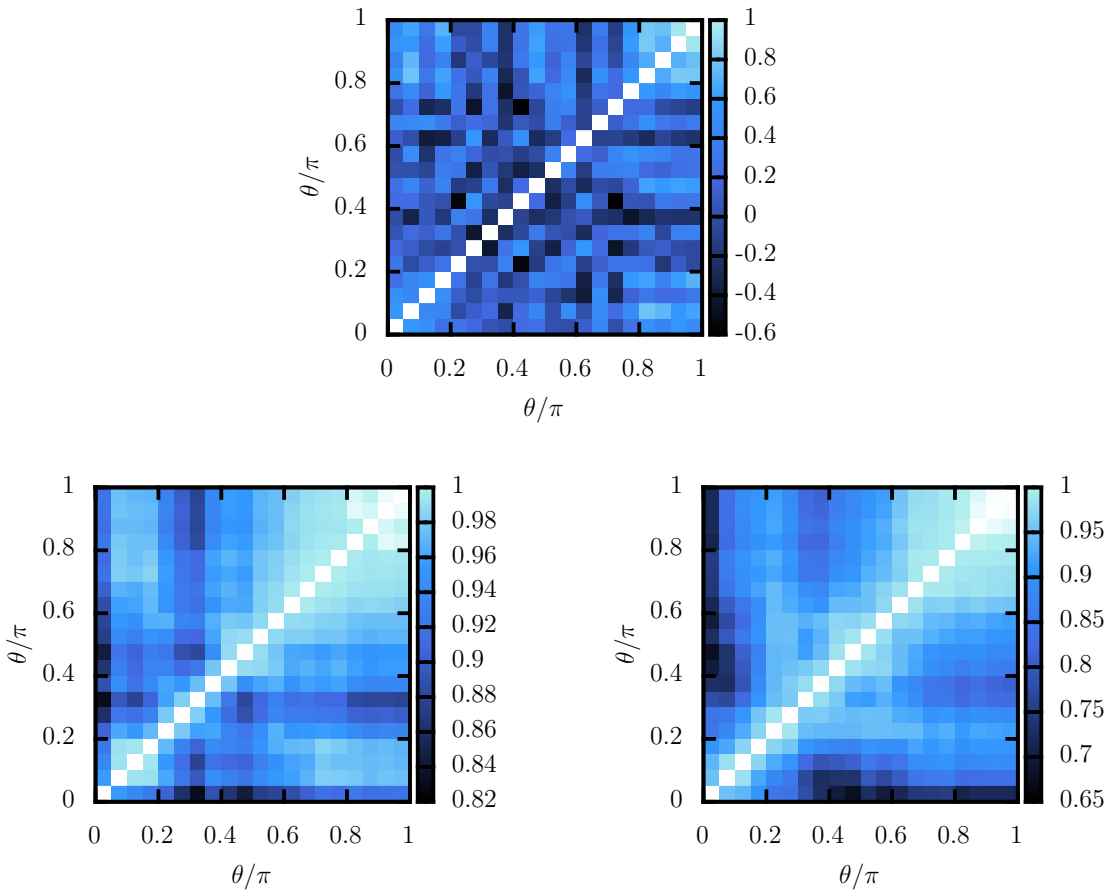

Figure 6–17: Covariance matrices for the pr3PCF calculation for $G\mu = 0$ **top:** L1, **bottom left:** L2, **bottom right:** L3. The L1 sample contains far less samples and contains more noise in the calculation and hence less correlation between the bins. For L2 and L3 samples, the bins are more smoothly correlated.

CHAPTER 7
DISCUSSION

In this chapter we discuss certain aspects of the project that need further discussion.

7.1 Fate of massive halos at $z = 10$

One of the puzzling results is the disappearance of massive halos observed at $z = 10$. Since these are not observed at lower redshifts, some physical process must "un-virialize" the halos such that the halo finding algorithm does not detect them. This prediction is supported by the evidence of *dark matter* overdensity in the wake region as shown on Figure 6–2. The history of the halos were traced through the halo catalogs and they were found to disappear abruptly between snapshots. This suggests that the physical processes that destroy these halos happen on very short timescales, much less than that of our simulation time steps $\sim$ 20 Myr at $z = 10$. This is most likely due to interaction between halos, namely, tidal destruction or harassment. However, simulations with higher time resolution and spatial resolution are required to confirm this prediction.

7.2 Uncertainties in the semi-analytic model

As of today, galaxy formation models are not fully understood due to their stochastic nature of the physical processes. Although many of the tunable parameters in the models are observationally constrained, the uncertainties are still large. Nonetheless, by using logical reasoning, we are able to predict the fundamental properties of the galaxy that resides in a halo.

One of the key assumptions that we make in this study is that although the semi-analytic model may not predict the exact star formation rates of a galaxy, it will at least distinguish between galaxies that will have star formation or not (i.e SFR > 0.001 $\mathrm{yr}^{-1}\mathrm{M}_{\odot}$ or SFR < 0.001 $\mathrm{yr}^{-1}\mathrm{M}_{\odot}$), with the use of reasonable parameters constrained through observations.

This assumption is used to make a cut between star forming and non-star forming galaxies in the semi-analytic model galaxies, which is required since the semi-analytic model does not produce the clear bi-modality present in the SDSS data. As a comparison, results from other semi-analytic models were also also tested, however, bi-modality was not present in any of the models (see Figure 7–1).

7.3 Comparison with other studies

There are 3 main studies on the SDSS 3PCF: [Guo et al., 2013], [McBride et al., 2011b], [Marín, 2011]. One major difference in results we observe is the luminosity dependence in both 2PCF and 3PCF. In other studies, it is confirmed that higher luminosity galaxies have a *higher* 2PCF and *lower* 3PCF amplitudes. In contrast to

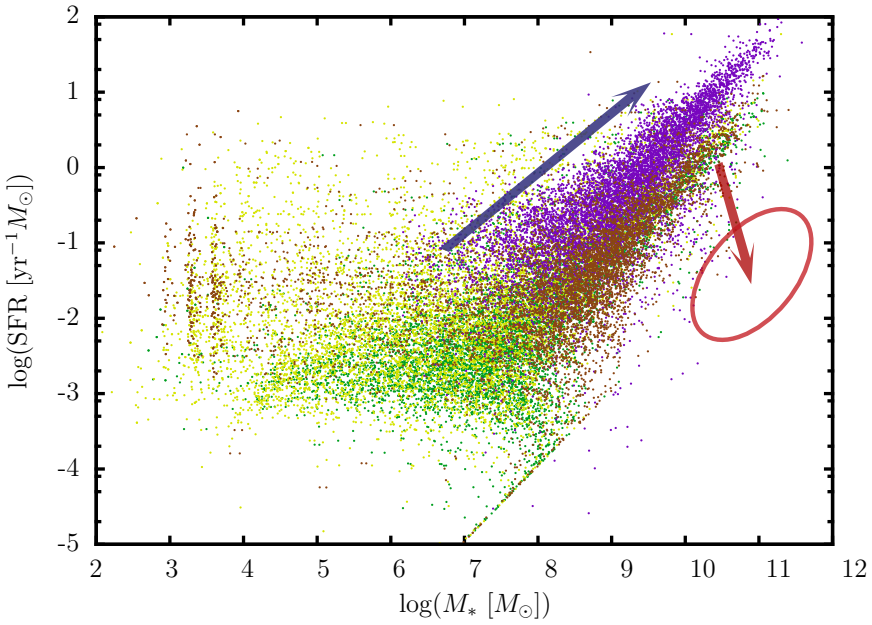

Figure 7–1: Stellar mass–SFR space for **magenta:** SAGE, **brown:** [De Lucia and Blaizot, 2007], **gold:** [Guo et al., 2013], **green:** [Guo et al., 2011] retrieved from the public Millennium database. Region encircled in red trace where the second population of galaxies in SDSS cover. Based on our current understanding of galaxy evolution, the blue arrow represents the direction of evolution for wet mergers and the red arrow shows the direction for dry mergers. Since the physical process that produce the missing population is dry mergers, it could be predicted that the modeling for such process is incorrect. However, there may exist other unknown modes of galaxy evolution that produce the missing population. *Note that the scale is logarithmic in SFR, and hence the galaxies with* $SFR = 0$ *are not shown on the plot.*

these observations, we see no amplitude differences between the correlation functions of the luminosity classes although the shapes are slightly different. This is however probably due to the low luminosity ranges $-20.5 < M_{\rm r} < -17.5$ that we are probing, which is one or two order of magnitude lower than other studies. For the 3PCF the binning scheme and the parameterization of triangles will also affect the outcome [McBride et al., 2011a].

7.4 Future work

The main limitation for this project is the simulation size, which places limits on our statistical precision and on the luminosity range we could reach. With our simulation box size of 150 h^{-1}Mpc, we were able to test and identify small effects that could potentially be used to discriminate between the string tensions. If the box size were > 500 h^{-1}Mpc, then more precise measurements could be made at the cost of losing small mass halos. Ideally, both large simulations and high resolution simulations should be run (like the Millennium I/II [Springel et al., 2005],[Boylan-Kolchin et al., 2009]), with the larger volume testing over a wide range of luminosities (for example up to $M_{\rm r} = -22$) and high-resolution simulations to study dwarf galaxies and internal structures of halos and subhalos.

Since the matter power spectrum of dark matter at $z = 0$ is the input for the simulation, there is very little uncertainty in the distribution of dark matter at this redshift (unless our knowledge of matter power spectrum is wrong). Since the halo mass function is in concordance with theoretical predictions, it is robust and hence, the dark matter sector of the simulation could be trusted. The vast majority of

uncertainty in the final mock galaxy catalogues arising from the conversion of dark matter halos to galaxies. One of the key assumptions that we have used is the validity of the SHAM method in assigning luminosities to galaxies. As shown on Figure 6–9, using a luminosity cuts based on SHAM assigned luminosities predicts a different 2PCF for non-star forming galaxies at low luminosities $-18.5 < M_{\rm r} < -17.5$ and $-19.5 < M_{\rm r} < -18.5$. This could be circumvented by matching an initial mass function (IMF) to the mock galaxy and inferring the luminosity by comparing with SEDs. Although the amount of uncertainty introduced through such procedure due to the dependence of initial mass functions needs testing, such approach would be more complete.

Another major uncertainty arises from the semi-analytic models, of which there are two origins: the dynamic physics included in the semi-analytic calculations and the parameter estimations. The physical processes encoded in the models is a developing field, which will improve in the near future. On the other hand, since the parameters used were tuned to observations "by hand", these estimates are not accurate. Improvements could be made by running a Monte-Carlo-Markov-Chain (MCMC) to fit all the parameters simultaneously. This is one of the benefits of using semi-analytic models: one could rerun halo-to-galaxy conversions quickly using different parameters to explore the parameter space efficiently. The great feature of SAGE, is its simplicity and the relatively small number of parameters (12 compared to other semi-analytic models such as GALACTICUS [Benson, 2012] which has ~ 50). Using such code, the parameter space could be explored quickly.

Although SFR was used to split the populations in this study, similar analyses could be done by using metallicity, which is also an indirect proxy for stellar age. Specifically, oxygen-to-hydrogen ratio O/H, which is included in both the MPA-JHU-VAGC, and SAGE output could be used. Unfortunately, the number of O/H measurements recorded in the catalog is $< 25\%$ of the entire catalog as of the current data release. Therefore, this may be feasible in the future versions of the SDSS data releases or other surveys with better S/N of oxygen emission lines.

Ultimately, the lack of constraints on the 3PCF measurements are due to the size of random catalogs used (although for L1 samples, the for some subsamples, the main source of uncertainty is due to the small data catalog). The code used in this study runs through the three point correlation function at $\mathcal{O}(N^3)$, meaning that increasing the random catalog by a factor of 2 will increase the computational time by a factor of 8. Using specialized codes such as N-tropy will improve this by a factor ranging from 8-30 [Gardner et al., 2007] and should give a tighter constraint.

Once improvements are made in the areas mentioned above, identical analysis should also be applied to higher redshift galaxies since the effects of cosmic strings is more prominent at higher redshifts, as the effects could be clearly be seen in Figure 6–1, at least for $G\mu = 10^{-5}$.

CHAPTER 8
CONCLUSION

This thesis serves as a first detailed study of effects of cosmic strings on large scale structures. Both statistical analysis of dark matter halos and galaxies were conducted, with more emphasis on galaxies, since galaxies are observed through observations.

First, dark matter simulations were run with cosmic string wakes placed at the center, produced from string tensions of $G\mu = 0$ (no string), $G\mu = 10^{-5}$ (large string tension to amplify the effects) and $G\mu = 10^{-7}$ (current string tension limit). Statistics including the halo mass functions and halo occupation densities were analyzed. For the halo mass functions, no notable differences were found except at $z = 10$, where the $G\mu = 10^{-5}$ model was found to have excess halos in the mass range of $\sim 10^{12} M_{\odot}$. Signatures of such excess diminished towards lower redshifts and were undetected at $z = 0$. The halo occupation density also showed statistically significant differences in the snapshots we analyzed ($z = 5, z = 3, z = 0$).

Galaxies were then assigned to halos using the semi-analytic model SAGE developed by Croton et al. The model was primarily calibrated using the stellar mass function, baryonic mass function, baryonic Tully-Fisher relation and stellar mass-metallicity relation. These tunings were performed "by hand" and thus may not, and are probably not in exact concordance with observations. Nonetheless, the stellar mass function turned out to resemble the one of SDSS remarkably well. However,

the semi-analytic model failed to produce the bi-modal distribution in stellar mass–SFR space. This is primarily due to our lack of understanding of complex dynamics that govern galactic evolution, specifically the dry merging sequence. Therefore, an assumption that galaxies that are likely to form stars will have the highest star formation rates and vice-versa was made, to split their population into star forming and non-star forming galaxies. Such a cut splits galaxies in terms of their age to first order (although there are physical phenomena that could drastically alter this relation).

Each class of galaxies was then subdivided into different luminosity ranges. This was done to determine whether cosmic strings have larger effects on large luminous galaxies or less massive dim galaxies. Since additional stellar population modeling is required for the semi-analytic model galaxies to obtain its luminosity, instead, it was approximated by using the SHAM method. Using the approximated luminosities, volume limited samples were generated and lightcones were produced.

With well defined galaxies in hand, p2PCFs and p3PCF/pr3PCFs were computed. All string tension models were consistent with each other for the 2 point statistics. In comparison to SDSS data, the L2 and L3 samples were consistent with each other. There were slight deviations for the L1 samples, whereby the amplitudes for the SDSS 2PCFs were found to posses a different slope and the amplitude was found to be higher. However, no conclusion could be made since; the cause of this effect pertains to the SHAM method as noted by other studies, especially at low luminosities.

The results of the 3PCFs contained much higher uncertainties and thus strict constraints could not be given. This owes to the fact that 3PCFs are far more computationally intensive to calculate and hence, require efficient code to sample through a large data set to make precise measurements. However, even with an efficient code, the number of the size of the random catalogues had to be relatively small compared to the random catalogues used in 2PCFs in order to keep the computational times reasonable. Therefore, special array searching algorithms developed specifically for such purposes must be employed to reduce uncertainties on the 3PCF measurements.

The averaged χ^2 for the different models compared against the SDSS measurements showed poor fit and showed no evidence for a strong preference towards one of the string tensions, which indicates that the effects we observe from cosmic strings on large scale structures is subtle. However, specific sub-classes of galaxies such as the star forming galaxies in the luminosity ranges $-19.5 < M_{\rm r} < -18.5$ and $-20.5 < M_{\rm r} < -19.5$ show an indication of slight a discrepancy between the string tension and is worth further investigation.

Appendix A - Output halo parameters from ROCKSTAR

Parameters	Description
ID	ID of halo (unique across entire simulation)
DescScale	Scale of descendant halo, if applicable
DescID	ID of descendant halo, if applicable
NumProg	Number of progenitors
Pid	Host halo ID (-1 if distinct halo)
Upid	Most massive host halo ID (only different from Pid in cases of sub-subs, or sub-sub-subs, etc.)
DescPid	Pid of descendant halo (if applicable)
Phantom	Nonzero for halos interpolated across timesteps
SAMMvir	Halo mass, smoothed across accretion history; always greater than sum of halo masses of contributing progenitors (Msun/h). Only for use with select semi-analytic models
Mvir	Halo mass (Msun/h)
Rvir	Halo radius (kpc/h comoving)
Rs	Scale radius (kpc/h comoving)
Vrms	Velocity dispersion (km/s physical)
mmp	whether the halo is the most massive progenitor or not
ScaleOfLastMM	scale factor of the last major merger (Mass ratio > 0.3)
Vmax	Maxmimum circular velocity (km/s physical)
X/Y/Z	Halo position (Mpc/h comoving)
VX/VY/VZ	Halo velocity (km/s physical)
JX/JY/JZ	Halo angular momenta ((Msun/h) * (Mpc/h) * km/s (physical))
Spin	Halo spin parameter
BreadthFirstID	breadth-first ordering of halos within a tree
DepthFirstID	depth-first ordering of halos within a tree
TreeRootID	ID of the halo at the last timestep in the tree
OrigHaloID	Original halo ID from halo finder
SnapNum	Snapshot number from which halo originated
NextCDID	Depthfirst ID of next coprogenitor
LastPDID	Depthfirst ID of last progenitor
RsKlypin	Scale radius determined using Vmax and Mvir (see Rockstar paper)
MvirAll	Mass enclosed within the specified overdensity, including unbound particles (Msun/h)
M200b–M2500c	Mass enclosed within specified overdensities (Msun/h)
Xoff	Offset of density peak from average particle position (kpc/h comoving)
Voff	Offset of density peak from average particle velocity (km/s physical)
SpinBullock	Bullock spin parameter ($J/(\sqrt{2}MVR)$)
BtoA, CtoA	Ratio of second and third largest shape ellipsoid axes (B and C) to largest shape ellipsoid axis (A) (dimensionless)
A[x],A[y],A[z]	Largest shape ellipsoid axis (kpc/h comoving)
Macc,Vacc	Mass and Vmax at accretion
Mpeak,Vpeak	Peak mass and Vmax

Table 8–1: Output halo paremeters from ROCKSTAR

Appendix B - Output galaxy parameters from SAGE

Parameters	Description
Type	Type indicating whether galaxy is at center of FOF group
GalaxyIndex	The unique identifier of this galaxy
HaloIndex	The haloId of the subhalo
FOFHaloIdx	ID of of the subhalo at the center of the FOF group containing this galaxy
TreeIdx	Unique id of galaxy formation tree containing this galaxy
SnapNum	The snapshot number where this galaxy was identified
CentralGal	The galaxy id of the central galaxy of the FOF group this galaxy is in
CentralMvir	The virial mass (as defined by $m_{crit200}$) of the FOF group the galaxy resides in
Pos	x, y, z position of galaxy
Vel	x, y, z velocity of galaxy
Spin	The spin of the galaxy
Len	Length
Mvir	Virial mass of the galaxy
Rvir	Virial radius of the FOF group
Vvir	Virial velocity of the subhalo
Vmax	Maximum rotational velocity
VelDisp	Velocity dispersion inside galaxy
ColdGas	Mass in cold gas
StellarMass	Mass in stars
BulgeMass	Mass in bulge
HotGas	Mass in hot gas
EjectedMass	The ejected mass component
BlackHoleMass	Mass of blackhole
IntraClusterStars	Mass in ICS
MetalsColdGas	Mass of metals in cold gas
MetalsStellarMass	Mass of metals in stars
MetalsBulgeMass	Mass of metals in bulge
MetalsHotGas	Mass of metals in hot gas
MetalsEjectedMass	Mass of metals ejected
MetalsIntraClusterStars	Mass of metals in intra cluster stars
Sfr	Star formation ratae
SfrBulge	Star formation rate in the bulge
SfrIntraClusterStars	Star formation rate in clusster stars
DiskRadius	Size of stellar
Cooling	Cooling rate of the galaxy
Heating	Heating rate of the galaxy
LastMajorMerger	Time since last major merger event
OutflowRate	Rate of gas outflow

Table 8–2: Output galaxy parameters from SAGE

Appendix C - SAGE fits

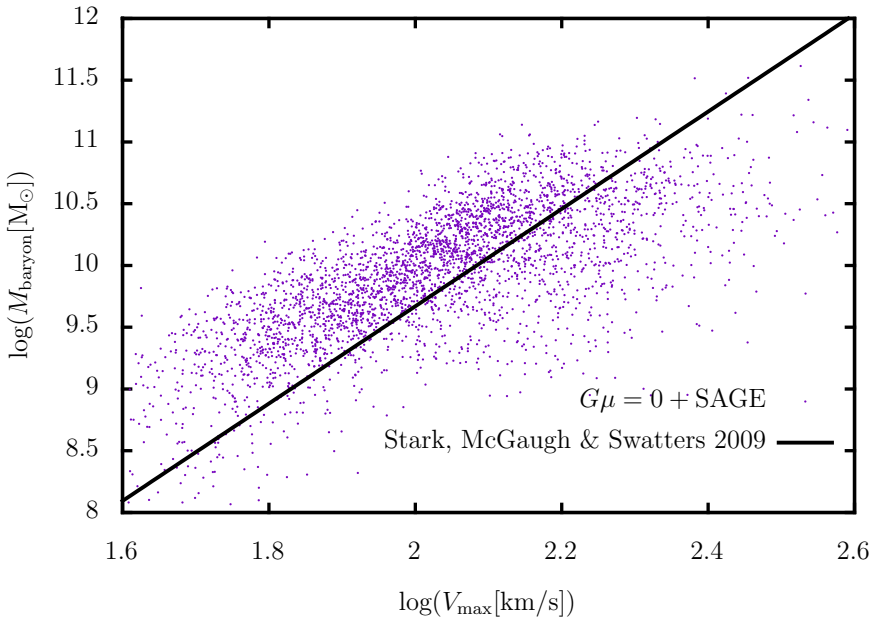

Figure 8–1: Comparison of the baryonic Tully-Fisher relation with [Stark et al., 2009]

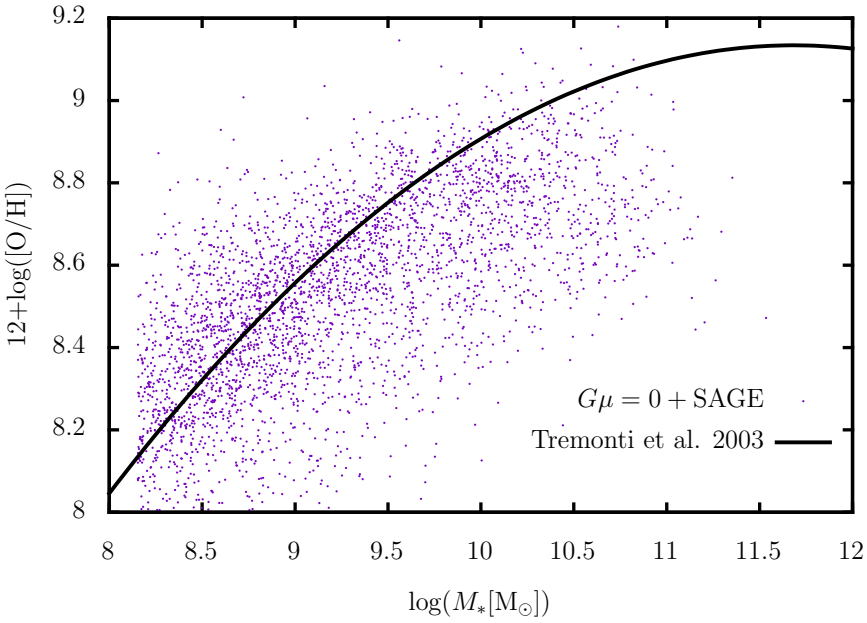

Figure 8–2: Comparison of the stellar mass – metallicity relation with [Tremonti et al., 2004]

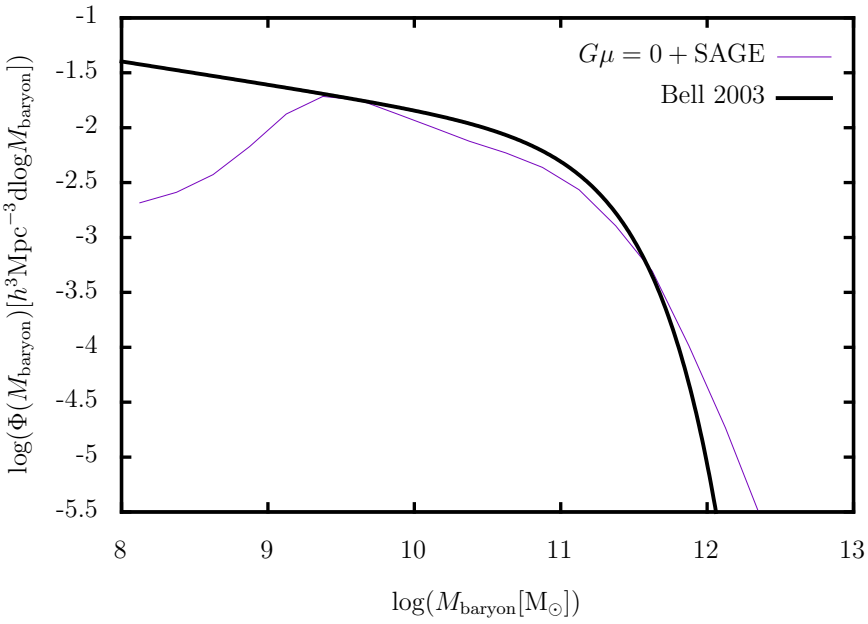

Figure 8–3: Comparison of the baryonic mass function with [Bell et al., 2003]

References

[Appenzeller, 2012] Appenzeller, I. (2012). *Introduction to Astronomical Spectroscopy*. Cambridge Observing Handbooks for Research Astronomers. Cambridge University Press. 1.8

[Aragón-Calvo et al., 2010] Aragón-Calvo, M. A., Platen, E., van de Weygaert, R., and Szalay, A. S. (2010). The Spine of the Cosmic Web. *Astrophys. J.*, 723:364–382. 1.1

[Argence and Lamareille, 2009] Argence, B. and Lamareille, F. (2009). Emission-lines calibrations of the star formation rate from the Sloan Digital Sky Survey. *Astron. Astrophys.*, 495:759–773. 1.7

[Arons and Silk, 1968] Arons, J. and Silk, J. (1968). On the Jean's criterion in relativistic cosmology. *MNRAS*, 140:331. 2.4

[Barnes and Hut, 1986] Barnes, J. and Hut, P. (1986). A hierarchical O(N log N) force-calculation algorithm. *Nature (London)*, 324:446–449. 2.2

[Barton et al., 2000] Barton, E. J., Geller, M. J., and Kenyon, S. J. (2000). Tidally Triggered Star Formation in Close Pairs of Galaxies. *Astrophys. J.*, 530:660–679. 1.6

[Barton Gillespie et al., 2003] Barton Gillespie, E., Geller, M. J., and Kenyon, S. J. (2003). Tidally Triggered Star Formation in Close Pairs of Galaxies. II. Constraints on Burst Strengths and Ages. *Astrophys. J.*, 582:668–688. 1.6

[Bassett et al., 2013] Bassett, R., Papovich, C., Lotz, J. M., Bell, E. F., Finkelstein, S. L., Newman, J. A., Tran, K.-V., Almaini, O., Lani, C., Cooper, M., Croton, D., Dekel, A., Ferguson, H. C., Kocevski, D. D., Koekemoer, A. M., Koo, D. C., McGrath, E. J., McIntosh, D. H., and Wechsler, R. H. (2013). CANDELS Observations of the Environmental Dependence of the Color-Mass-Morphology Relation at z = 1.6. *ArXiv e-prints*. 1.6

[Baugh et al., 1995] Baugh, C. M., Gaztanaga, E., and Efstathiou, G. (1995). A comparison of the evolution of density fields in perturbation theory and numerical simulations - II. Counts-in-cells analysis. *MNRAS*, 274:1049–1070. 2.1

[Behroozi et al., 2013a] Behroozi, P. S., Wechsler, R. H., and Wu, H.-Y. (2013a). The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophys. J.*, 762:109. (document), 2.3, 2–1

[Behroozi et al., 2013b] Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., Busha, M. T., Klypin, A. A., and Primack, J. R. (2013b). Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology. *Astrophys. J.*, 763:18. 2.3

[Bell et al., 2003] Bell, E. F., McIntosh, D. H., Katz, N., and Weinberg, M. D. (2003). A First Estimate of the Baryonic Mass Function of Galaxies. *Astrophys. J. Lett.*, 585:L117–L120. (document), 8–3

[Benson, 2000] Benson, A. (2000). *Galaxy formation and clustering in a hierarchical universe*. PhD thesis, Durham University. 1.1

[Benson, 2010] Benson, A. J. (2010). Galaxy formation theory. *Phys. Rep.*, 495:33–86. 2.4

[Benson, 2012] Benson, A. J. (2012). GALACTICUS: A semi-analytic model of galaxy formation. *New Astronomy*, 17:175–197. 7.4

[Binney and Merrifield, 1998] Binney, J. and Merrifield, M. (1998). *Galactic Astronomy*. Princeton paperbacks. Princeton University Press. 1.6

[Blanton et al., 2003a] Blanton, M. R., Brinkmann, J., Csabai, I., Doi, M., Eisenstein, D., Fukugita, M., Gunn, J. E., Hogg, D. W., and Schlegel, D. J. (2003a). Estimating Fixed-Frame Galaxy Magnitudes in the Sloan Digital Sky Survey. *Astronomical, J.*, 125:2348–2360. 3.1

[Blanton et al., 2005a] Blanton, M. R., Eisenstein, D., Hogg, D. W., Schlegel, D. J., and Brinkmann, J. (2005a). Relationship between Environment and the Broadband Optical Properties of Galaxies in the Sloan Digital Sky Survey. *Astrophys. J.*, 629:143–157. 1.6

[Blanton et al., 2003b] Blanton, M. R., Hogg, D. W., Bahcall, N. A., Brinkmann, J., Britton, M., Connolly, A. J., Csabai, I., Fukugita, M., Loveday, J., Meiksin, A., Munn, J. A., Nichol, R. C., Okamura, S., Quinn, T., Schneider, D. P., Shimasaku, K., Strauss, M. A., Tegmark, M., Vogeley, M. S., and Weinberg, D. H. (2003b).

The Galaxy Luminosity Function and Luminosity Density at Redshift z = 0.1. *Astrophys. J.*, 592:819–838. 3.1

[Blanton et al., 2003c] Blanton, M. R., Hogg, D. W., Bahcall, N. A., Brinkmann, J., Britton, M., Connolly, A. J., Csabai, I., Fukugita, M., Loveday, J., Meiksin, A., Munn, J. A., Nichol, R. C., Okamura, S., Quinn, T., Schneider, D. P., Shimasaku, K., Strauss, M. A., Tegmark, M., Vogeley, M. S., and Weinberg, D. H. (2003c). The Galaxy Luminosity Function and Luminosity Density at Redshift z = 0.1. *Astrophys. J.*, 592:819–838. 5.2

[Blanton et al., 2003d] Blanton, M. R., Lin, H., Lupton, R. H., Maley, F. M., Young, N., Zehavi, I., and Loveday, J. (2003d). An Efficient Targeting Strategy for Multiobject Spectrograph Surveys: the Sloan Digital Sky Survey "Tiling" Algorithm. *Astronomical, J.*, 125:2276–2286. 3.1

[Blanton et al., 2005b] Blanton, M. R., Schlegel, D. J., Strauss, M. A., Brinkmann, J., Finkbeiner, D., Fukugita, M., Gunn, J. E., Hogg, D. W., Ivezić, Ž., Knapp, G. R., Lupton, R. H., Munn, J. A., Schneider, D. P., Tegmark, M., and Zehavi, I. (2005b). New York University Value-Added Galaxy Catalog: A Galaxy Catalog Based on New Public Surveys. *Astronomical, J.*, 129:2562–2578. 3.1

[Blitz et al., 2007] Blitz, L., Fukui, Y., Kawamura, A., Leroy, A., Mizuno, N., and Rosolowsky, E. (2007). Giant Molecular Clouds in Local Group Galaxies. *Protostars and Planets V*, pages 81–96. 1.6

[Bournaud, 2011] Bournaud, F. (2011). Star formation in galaxy interactions and mergers. In Charbonnel, C. and Montmerle, T., editors, *EAS Publications Series*, volume 51 of *EAS Publications Series*, pages 107–131. 1.6

[Boylan-Kolchin et al., 2009] Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., and Lemson, G. (2009). Resolving cosmic structure formation with the Millennium-II Simulation. *MNRAS*, 398:1150–1164. 2.2, 7.4

[Brammer et al., 2008] Brammer, G. B., van Dokkum, P. G., and Coppi, P. (2008). EAZY: A Fast, Public Photometric Redshift Code. *Astrophys. J.*, 686:1503–1513. 1.8

[Brinchmann et al., 2004] Brinchmann, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., and Brinkmann, J. (2004). The physical properties of star-forming galaxies in the low-redshift Universe. *MNRAS*, 351:1151–1179. 3.1

[Bromm and Yoshida, 2011] Bromm, V. and Yoshida, N. (2011). The First Galaxies. *Annul. Rev. Astron. and Astrophys.*, 49:373–407. 1.1

[Bundy et al., 2005] Bundy, K., Ellis, R. S., and Conselice, C. J. (2005). The Mass Assembly Histories of Galaxies of Various Morphologies in the GOODS Fields. *Astrophys. J.*, 625:621–632. 1.7

[Busha et al., 2005] Busha, M. T., Evrard, A. E., Adams, F. C., and Wechsler, R. H. (2005). The ultimate halo mass in a ΛCDM universe. *MNRAS*, 363:L11–L15. 2.3

[Calzetti, 2012] Calzetti, D. (2012). Star Formation Rate Indicators. *ArXiv e-prints*. 1.7

[Carlson and White, 2010] Carlson, J. and White, M. (2010). Embedding Realistic Surveys in Simulations Through Volume Remapping. *Appl. Phys. Lett.*, 190:311–314. 2.5

[Carroll and Ostlie, 2007] Carroll, B. and Ostlie, D. (2007). *An introduction to modern astrophysics*. Pearson Addison-Wesley. 1.6

[Charlot and Longhetti, 2001] Charlot, S. and Longhetti, M. (2001). Nebular emission from star-forming galaxies. *MNRAS*, 323:887–903. 3.1

[Chomiuk and Povich, 2011] Chomiuk, L. and Povich, M. S. (2011). Toward a Unification of Star Formation Rate Determinations in the Milky Way and Other Galaxies. *Astronomical, J.*, 142:197. 1.7

[Coil, 2013] Coil, A. L. (2013). *The Large-Scale Structure of the Universe*, page 387. 5.4

[Conroy et al., 2006] Conroy, C., Wechsler, R. H., and Kravtsov, A. V. (2006). Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time. *Astrophys. J.*, 647:201–214. 4.1

[Crocce et al., 2006] Crocce, M., Pueblas, S., and Scoccimarro, R. (2006). Transients from initial conditions in cosmological simulations. *MNRAS*, 373:369–381. 2.1

[Croton, prep] Croton, D. e. a. (in-prep). In prep. 2.4

[Cuillandre and Bertin, 2006] Cuillandre, J.-C. and Bertin, E. (2006). CFHT Legacy Survey (CFHTLS) : a rich data set. In Barret, D., Casoli, F., Lagache, G., Lecavelier, A., and Pagani, L., editors, *SF2A-2006: Semaine de l'Astrophysique Francaise*, page 265. 1.8

[Danos, 2010] Danos, R. (2010). *Cosmic string searches in new observational windows*. PhD thesis, McGill University. 1.3, 1.4

[De Lucia and Blaizot, 2007] De Lucia, G. and Blaizot, J. (2007). The hierarchical formation of the brightest cluster galaxies. *MNRAS*, 375:2–14. (document), 7–1

[Diemand and Moore, 2011] Diemand, J. and Moore, B. (2011). The Structure and Evolution of Cold Dark Matter Halos. *Advanced Science Letters*, 4:297–310. 4.1

[Dressler, 1980] Dressler, A. (1980). Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. *Astrophys. J.*, 236:351–365. 1.6

[Duplessis and Brandenberger, 2013] Duplessis, F. and Brandenberger, R. (2013). Note on structure formation from cosmic string wakes. *J. Cosmo. Astro. Part. Phys.*, 4:45. 1.4

[Dvorkin et al., 2011] Dvorkin, C., Wyman, M., and Hu, W. (2011). Cosmic string constraints from WMAP and the South Pole Telescope data. *Phys. Rev. D*, 84(12):123519. 1.5

[Einasto et al., 1974] Einasto, J., Saar, E., Kaasik, A., and Chernin, A. D. (1974). Missing mass around galaxies - Morphological evidence. *Nature (London)*, 252:111–113. 1.6

[ESA, 2011] ESA (2011). Definition study report. 1.8

[Gardner et al., 2007] Gardner, J. P., Connolly, A., and McBride, C. (2007). A Framework for Analyzing Massive Astrophysical Datasets on a Distributed Grid. In Shaw, R. A., Hill, F., and Bell, D. J., editors, *Astronomical Data Analysis Software and Systems XVI*, volume 376 of *Astronomical Society of the Pacific Conference Series*, page 69. 5.5, 7.4

[Gnedin and Hui, 1998] Gnedin, N. Y. and Hui, L. (1998). Probing the Universe with the Lyalpha forest - I. Hydrodynamics of the low-density intergalactic medium. *MNRAS*, 296:44–55. 2.4

[Grinstein and Wise, 1987] Grinstein, B. and Wise, M. B. (1987). On the validity of the Zel'dovich approximation. *Astrophys. J.*, 320:448–453. 2.1

[Gunn et al., 2006] Gunn, J. E., Siegmund, W. A., Mannery, E. J., Owen, R. E., Hull, C. L., Leger, R. F., Carey, L. N., Knapp, G. R., York, D. G., Boroski, W. N., Kent, S. M., Lupton, R. H., Rockosi, C. M., Evans, M. L., Waddell, P., Anderson, J. E., Annis, J., Barentine, J. C., Bartoszek, L. M., Bastian, S., Bracker, S. B., Brewington, H. J., Briegel, C. I., Brinkmann, J., Brown, Y. J., Carr, M. A., Czarapata, P. C., Drennan, C. C., Dombeck, T., Federwitz, G. R., Gillespie, B. A., Gonzales, C., Hansen, S. U., Harvanek, M., Hayes, J., Jordan, W., Kinney, E., Klaene, M., Kleinman, S. J., Kron, R. G., Kresinski, J., Lee, G., Limmongkol, S., Lindenmeyer, C. W., Long, D. C., Loomis, C. L., McGehee, P. M., Mantsch, P. M., Neilsen, Jr., E. H., Neswold, R. M., Newman, P. R., Nitta, A., Peoples, Jr., J., Pier, J. R., Prieto, P. S., Prosapio, A., Rivetta, C., Schneider, D. P., Snedden, S., and Wang, S.-i. (2006). The 2.5 m Telescope of the Sloan Digital Sky Survey. *Astronomical, J.*, 131:2332–2359. 3

[Guo et al., 2013] Guo, H., Li, C., Jing, Y. P., and Boerner, G. (2013). Mass and color dependence of the three-point correlation function of galaxies in the local universe. *ArXiv e-prints*. (document), 3.1, 5.5, 5.5, 7.3, 7–1

[Guo et al., 2011] Guo, Q., White, S., Boylan-Kolchin, M., De Lucia, G., Kauffmann, G., Lemson, G., Li, C., Springel, V., and Weinmann, S. (2011). From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. *MNRAS*, 413:101–131. (document), 7–1

[Heiderman et al., 2010] Heiderman, A., Evans, II, N. J., Allen, L. E., Huard, T., and Heyer, M. (2010). The Star Formation Rate and Gas Surface Density Relation in the Milky Way: Implications for Extragalactic Studies. *Astrophys. J.*, 723:1019–1037. 1.6

[Henriques et al., 2009] Henriques, B. M. B., Thomas, P. A., Oliver, S., and Roseboom, I. (2009). Monte Carlo Markov Chain parameter estimation in semi-analytic models of galaxy formation. *MNRAS*, 396:535–547. 2.4

[Hernández and Brandenberger, 2012] Hernández, O. and Brandenberger, R. (2012). The 21 cm signature of shock heated and diffuse cosmic string wakes. *Astrophys. J.*, 7:32. 1.4, 1.4

[Hogg et al., 2001] Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., and Gunn, J. E. (2001). A Photometricity and Extinction Monitor at the Apache Point Observatory. *Astronomical, J.*, 122:2129–2138. 3

[Howell, 2006] Howell, S. (2006). *Handbook of CCD Astronomy*. Cambridge observing handbooks for research astronomers. Cambridge University Press. 1.8

[Ivezić et al., 2004] Ivezić, Ž., Lupton, R. H., Schlegel, D., Boroski, B., Adelman-McCarthy, J., Yanny, B., Kent, S., Stoughton, C., Finkbeiner, D., Padmanabhan, N., Rockosi, C. M., Gunn, J. E., Knapp, G. R., Strauss, M. A., Richards, G. T., Eisenstein, D., Nicinski, T., Kleinman, S. J., Krzesinski, J., Newman, P. R., Snedden, S., Thakar, A. R., Szalay, A., Munn, J. A., Smith, J. A., Tucker, D., and Lee, B. C. (2004). SDSS data management and photometric quality assessment. *Astronomische Nachrichten*, 325:583–589. 3

[Jing and Börner, 2004] Jing, Y. P. and Börner, G. (2004). The Three-Point Correlation Function of Galaxies Determined from the Two-Degree Field Galaxy Redshift Survey. *Astrophys. J.*, 607:140–163. 5.5

[Kauffmann et al., 2003] Kauffmann, G., Heckman, T. M., White, S. D. M., Charlot, S., Tremonti, C., Brinchmann, J., Bruzual, G., Peng, E. W., Seibert, M., Bernardi, M., Blanton, M., Brinkmann, J., Castander, F., Csábai, I., Fukugita, M., Ivezic, Z., Munn, J. A., Nichol, R. C., Padmanabhan, N., Thakar, A. R., Weinberg, D. H., and York, D. (2003). Stellar masses and star formation histories for 10^5 galaxies from the Sloan Digital Sky Survey. *MNRAS*, 341:33–53. 3.1

[Knebe et al., 2011] Knebe, A., Knollmann, S. R., Muldrew, S. I., Pearce, F. R., Aragon-Calvo, M. A., Ascasibar, Y., Behroozi, P. S., Ceverino, D., Colombi, S., Diemand, J., Dolag, K., Falck, B. L., Fasel, P., Gardner, J., Gottlöber, S., Hsu, C.-H., Iannuzzi, F., Klypin, A., Lukić, Z., Maciejewski, M., McBride, C., Neyrinck, M. C., Planelles, S., Potter, D., Quilis, V., Rasera, Y., Read, J. I., Ricker, P. M., Roy, F., Springel, V., Stadel, J., Stinson, G., Sutter, P. M., Turchaninov, V., Tweed, D., Yepes, G., and Zemp, M. (2011). Haloes gone MAD: The Halo-Finder Comparison Project. *MNRAS*, 415:2293–2318. 2.3

[Komatsu et al., 2011] Komatsu, E., Smith, K., Dunkley, J., Bennett, C., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker,

G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2011). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *Appl. Phys. Lett.*, 192:18. 1.1

[Kroupa, 2001] Kroupa, P. (2001). On the variation of the initial mass function. *MNRAS*, 322:231–246. 3.1

[Landy and Szalay, 1993] Landy, S. D. and Szalay, A. S. (1993). Bias and variance of angular correlation functions. *Astrophys. J.*, 412:64–71. 5.4

[Larson et al., 2011] Larson, D., Dunkley, J., Hinshaw, G., Komatsu, E., Nolta, M. R., Bennett, C. L., Gold, B., Halpern, M., Hill, R. S., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Page, L., Smith, K. M., Spergel, D. N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2011). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters. *Appl. Phys. Lett.*, 192:16. 1.1

[LBNL, 2011] LBNL (2011). Department of Energy Review rport for the BigBOSS proposal. 1.8

[Lee et al., 2006] Lee, K.-S., Giavalisco, M., Gnedin, O. Y., Somerville, R. S., Ferguson, H. C., Dickinson, M., and Ouchi, M. (2006). The Large-Scale and Small-Scale Clustering of Lyman Break Galaxies at 3.5¡=z¡=5.5 from the GOODS Survey. *Astrophys. J.*, 642:63–80. 5.4

[Lewis and Bridle, 2002] Lewis, A. and Bridle, S. (2002). Cosmological parameters from CMB and other data: a Monte- Carlo approach. *Phys. Rev.*, D66:103511. (document), 2–1

[Lupton et al., 1999] Lupton, R. H., Gunn, J. E., and Szalay, A. S. (1999). A Modified Magnitude System that Produces Well-Behaved Magnitudes, Colors, and Errors Even for Low Signal-to-Noise Ratio Measurements. *Astronomical, J.*, 118:1406–1410. 3

[Mac Low and Klessen, 2004] Mac Low, M.-M. and Klessen, R. S. (2004). Control of star formation by supersonic turbulence. *Reviews of Modern Physics*, 76:125–194. 1.6

[Madgwick et al., 2003] Madgwick, D. S., Hawkins, E., Lahav, O., Maddox, S., Norberg, P., Peacock, J. A., Baldry, I. K., Baugh, C. M., Bland-Hawthorn, J., Bridges,

T., Cannon, R., Cole, S., Colless, M., Collins, C., Couch, W., Dalton, G., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lewis, I., Lumsden, S., Peterson, B. A., Sutherland, W., and Taylor, K. (2003). The 2dF Galaxy Redshift Survey: galaxy clustering per spectral type. *MNRAS*, 344:847–856. (document), 5–1

[Marín, 2011] Marín, F. (2011). The Large-scale Three-point Correlation Function of Sloan Digital Sky Survey Luminous Red Galaxies. *Astrophys. J.*, 737:97. 5.5, 7.3

[Masaki et al., 2013] Masaki, S., Lin, Y.-T., and Yoshida, N. (2013). Modeling color-dependent galaxy clustering in cosmological simulations. *ArXiv e-prints*. 6.5, 6.5

[McBride et al., 2011a] McBride, C. K., Connolly, A. J., Gardner, J. P., Scranton, R., Newman, J. A., Scoccimarro, R., Zehavi, I., and Schneider, D. P. (2011a). Three-point Correlation Functions of SDSS Galaxies: Luminosity and Color Dependence in Redshift and Projected Space. *Astrophys. J.*, 726:13. 3.1, 7.3

[McBride et al., 2011b] McBride, C. K., Connolly, A. J., Gardner, J. P., Scranton, R., Scoccimarro, R., Berlind, A. A., Marín, F., and Schneider, D. P. (2011b). Three-point Correlation Functions of SDSS Galaxies: Constraining Galaxy-mass Bias. *Astrophys. J.*, 739:85. 5.4, 5.5, 7.3

[Mo et al., 2010] Mo, H., van den Bosch, F., and White, S. (2010). *Galaxy Formation and Evolution*. Galaxy Formation and Evolution. Cambridge University Press. 1.1, 1.6, 2.4, 5.3

[Moore et al., 2001] Moore, A. W., Connolly, A. J., Genovese, C., Gray, A., Grone, L., Kanidoris, II, N., Nichol, R. C., Schneider, J., Szalay, A. S., Szapudi, I., and Wasserman, L. (2001). Fast Algorithms and Efficient Statistics: N-Point Correlation Functions. In Banday, A. J., Zaroubi, S., and Bartelmann, M., editors, *Mining the Sky*, page 71. 5.5

[Morandi, 2008] Morandi, A. (2008). *Properties of gas and dark matter in X-ray galaxy clusters with Sunyaev Zel'dovich measurements*. PhD thesis, UNIVERSITA DI BOLOGNA. 2.3

[Padoan and Nordlund, 2011] Padoan, P. and Nordlund, Å. (2011). The Star Formation Rate of Supersonic Magnetohydrodynamic Turbulence. *Astrophys. J.*, 730:40. 1.6

[Pier et al., 2003] Pier, J. R., Munn, J. A., Hindsley, R. B., Hennessy, G. S., Kent, S. M., Lupton, R. H., and Ivezić, Ž. (2003). Astrometric Calibration of the Sloan Digital Sky Survey. *Astronomical, J.*, 125:1559–1579. 3

[Planck collaboration XXV et al., 2013] Planck collaboration XXV, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2013). Planck 2013 results. XXV. Searches for cosmic strings and other topological defects. *ArXiv e-prints*. 1.2

[Poggianti et al., 2013] Poggianti, B. M., Calvi, R., Bindoni, D., D'Onofrio, M., Moretti, A., Valentinuzzi, T., Fasano, G., Fritz, J., De Lucia, G., Vulcani, B., Bettoni, D., Gullieuszik, M., and Omizzolo, A. (2013). Superdense Galaxies and the Mass-Size Relation at Low Redshift. *Astrophys. J.*, 762:77. 1.6

[Poole, prep] Poole, G. (in-prep). In prep. 2.2

[Ryden, 2003] Ryden, B. (2003). *Introduction to cosmology*. Addison-Wesley. 1.1

[Sanidas et al., 2012] Sanidas, S. A., Battye, R. A., and Stappers, B. W. (2012). Constraints on cosmic string tension imposed by the limit on the stochastic gravitational wave background from the European Pulsar Timing Array. *Phys. Rev. D*, 85(12):122003. 1.5

[Scoccimarro et al., 2012] Scoccimarro, R., Hui, L., Manera, M., and Chan, K. C. (2012). Large-scale bias and efficient generation of initial conditions for nonlocal primordial non-Gaussianity. *Phys. Rev. D*, 85(8):083002. 2.1

[Silk and Vilenkin, 1984] Silk, J. and Vilenkin, A. (1984). Cosmic strings and galaxy formation. *Physical Review Letters*, 53:1700–1703. 1.4

[Smith et al., 2002] Smith, J. A., Tucker, D. L., Kent, S., Richmond, M. W., Fukugita, M., Ichikawa, T., Ichikawa, S.-i., Jorgensen, A. M., Uomoto, A., Gunn, J. E., Hamabe, M., Watanabe, M., Tolea, A., Henden, A., Annis, J., Pier, J. R., McKay, T. A., Brinkmann, J., Chen, B., Holtzman, J., Shimasaku, K., and York, D. G. (2002). The u'g'r'i'z' Standard-Star System. *Astronomical, J.*, 123:2121–2144. 3

[Springel, 2005] Springel, V. (2005). The cosmological simulation code GADGET-2. *MNRAS*, 364:1105–1134. 2.2

[Springel et al., 2005] Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature (London)*, 435:629–636. 2.2, 7.4

[Springel et al., 2001] Springel, V., Yoshida, N., and White, S. D. M. (2001). GADGET: a code for collisionless and gasdynamical cosmological simulations. *Nature (London)*, 6:79–117. 2.2

[Stark et al., 2009] Stark, D. V., McGaugh, S. S., and Swaters, R. A. (2009). A First Attempt to Calibrate the Baryonic Tully-Fisher Relation with Gas-Dominated Galaxies. *Astronomical, J.*, 138:392–401. (document), 8–1

[Stoughton et al., 2002] Stoughton, C., Lupton, R. H., Bernardi, M., Blanton, M. R., Burles, S., Castander, F. J., Connolly, A. J., Eisenstein, D. J., Frieman, J. A., Hennessy, G. S., Hindsley, R. B., Ivezić, Ž., Kent, S., Kunszt, P. Z., Lee, B. C., Meiksin, A., Munn, J. A., Newberg, H. J., Nichol, R. C., Nicinski, T., Pier, J. R., Richards, G. T., Richmond, M. W., Schlegel, D. J., Smith, J. A., Strauss, M. A., SubbaRao, M., Szalay, A. S., Thakar, A. R., Tucker, D. L., Vanden Berk, D. E., Yanny, B., Adelman, J. K., Anderson, Jr., J. E., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Bartelmann, M., Bastian, S., Bauer, A., Berman, E., Böhringer, H., Boroski, W. N., Bracker, S., Briegel, C., Briggs, J. W., Brinkmann, J., Brunner, R., Carey, L., Carr, M. A., Chen, B., Christian, D., Colestock, P. L., Crocker, J. H., Csabai, I., Czarapata, P. C., Dalcanton, J., Davidsen, A. F., Davis, J. E., Dehnen, W., Dodelson, S., Doi, M., Dombeck, T., Donahue, M., Ellman, N., Elms, B. R., Evans, M. L., Eyer, L., Fan, X., Federwitz, G. R., Friedman, S., Fukugita, M., Gal, R., Gillespie, B., Glazebrook, K., Gray, J., Grebel, E. K., Greenawalt, B., Greene, G., Gunn, J. E., de Haas, E., Haiman, Z., Haldeman, M., Hall, P. B., Hamabe, M., Hansen, B., Harris, F. H., Harris, H., Harvanek, M., Hawley, S. L., Hayes, J. J. E., Heckman, T. M., Helmi, A., Henden, A., Hogan, C. J., Hogg, D. W., Holmgren, D. J., Holtzman, J., Huang, C.-H., Hull, C., Ichikawa, S.-I., Ichikawa, T., Johnston, D. E., Kauffmann, G., Kim, R. S. J., Kimball, T., Kinney, E., Klaene, M., Kleinman, S. J., Klypin, A., Knapp, G. R., Korienek, J., Krolik, J., Kron, R. G., Krzesiński, J., Lamb, D. Q., Leger, R. F., Limmongkol, S., Lindenmeyer, C., Long, D. C., Loomis, C., Loveday, J., MacKinnon, B., Mannery, E. J., Mantsch, P. M., Margon, B., McGehee, P., McKay, T. A., McLean, B., Menou, K., Merelli, A., Mo, H. J., Monet, D. G., Nakamura, O., Narayanan, V. K., Nash, T., Neilsen, Jr., E. H., Newman, P. R., Nitta, A., Odenkirchen,

M., Okada, N., Okamura, S., Ostriker, J. P., Owen, R., Pauls, A. G., Peoples, J., Peterson, R. S., Petravick, D., Pope, A., Pordes, R., Postman, M., Prosapio, A., Quinn, T. R., Rechenmacher, R., Rivetta, C. H., Rix, H.-W., Rockosi, C. M., Rosner, R., Ruthmansdorfer, K., Sandford, D., Schneider, D. P., Scranton, R., Sekiguchi, M., Sergey, G., Sheth, R., Shimasaku, K., Smee, S., Snedden, S. A., Stebbins, A., Stubbs, C., Szapudi, I., Szkody, P., Szokoly, G. P., Tabachnik, S., Tsvetanov, Z., Uomoto, A., Vogeley, M. S., Voges, W., Waddell, P., Walterbos, R., Wang, S.-i., Watanabe, M., Weinberg, D. H., White, R. L., White, S. D. M., Wilhite, B., Wolfe, D., Yasuda, N., York, D. G., Zehavi, I., and Zheng, W. (2002). Sloan Digital Sky Survey: Early Data Release. *Astronomical, J.*, 123:485–548. 3

[Szapudi and Szalay, 1998] Szapudi, S. and Szalay, A. S. (1998). A New Class of Estimators for the N-Point Correlations. *Astrophys. J. Lett.*, 494:L41. 5.5

[Takada and Jain, 2003] Takada, M. and Jain, B. (2003). The three-point correlation function in cosmology. *MNRAS*, 340:580–608. 5.5

[Takada and Silverman,] Takada, M. and Silverman, J. Sciences with SuMIRe PFR Survey. 1.8

[Tempel, 2011] Tempel, E. (2011). Tracing the galaxy evolution in supercluster-void environment. In *Galaxy Formation*, page 134P. 1.6

[Tinker et al., 2008] Tinker, J. L., Conroy, C., Norberg, P., Patiri, S. G., Weinberg, D. H., and Warren, M. S. (2008). Void Statistics in Large Galaxy Redshift Surveys: Does Halo Occupation of Field Galaxies Depend on Environment? *Astrophys. J.*, 686:53–71. (document), 5.1, 5.1, 6–3

[Tremonti et al., 2004] Tremonti, C. A., Heckman, T. M., Kauffmann, G., Brinchmann, J., Charlot, S., White, S. D. M., Seibert, M., Peng, E. W., Schlegel, D. J., Uomoto, A., Fukugita, M., and Brinkmann, J. (2004). The Origin of the Mass-Metallicity Relation: Insights from 53,000 Star-forming Galaxies in the Sloan Digital Sky Survey. *Astrophys. J.*, 613:898–913. (document), 3.1, 8–2

[Unzicker and Fabian, 2010] Unzicker, A. and Fabian, K. (2010). Do Galaxies Change in Size ? An Angular Size Test at Low Redshift with SDSS Data. *ArXiv e-prints*. 3.1

[van den Bosch et al., 2003] van den Bosch, F. C., Yang, X., and Mo, H. J. (2003). Linking early- and late-type galaxies to their dark matter haloes. *MNRAS*, 340:771–792. 2.4

[Vilenkin and Shellard, 2000] Vilenkin, A. and Shellard, E. (2000). *Cosmic Strings and Other Topological Defects*. Cambridge Monographs on Mathematical Physics. Cambridge University Press. 1.2

[Woods et al., 2006] Woods, D. F., Geller, M. J., and Barton, E. J. (2006). Tidally Triggered Star Formation in Close Pairs of Galaxies: Major and Minor Interactions. *Astronomical, J.*, 132:197–209. 1.6

[Yang et al., 2005] Yang, X., Mo, H. J., Jing, Y. P., and van den Bosch, F. C. (2005). Galaxy occupation statistics of dark matter haloes: observational results. *MNRAS*, 358:217–232. 2.4

[Yang et al., 2003] Yang, X., Mo, H. J., and van den Bosch, F. C. (2003). Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies. *MNRAS*, 339:1057–1080. 2.4

[Yang et al., 2009] Yang, X., Mo, H. J., and van den Bosch, F. C. (2009). Galaxy Groups in the SDSS DR4. III. The Luminosity and Stellar Mass Functions. *Astrophys. J.*, 695:900–916. (document), 6–5

[Yoo et al., 2012] Yoo, C.-M., Saito, R., Sendouda, Y., Takahashi, K., and Yamauchi, D. (2012). Femto-lensing due to a Cosmic String. *ArXiv e-prints*. 1.5

[York et al., 2000] York, D. G., Adelman, J., Anderson, Jr., J. E., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouser, R., Bastian, S., Berman, E., Boroski, W. N., Bracker, S., Briegel, C., Briggs, J. W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Carr, M. A., Castander, F. J., Chen, B., Colestock, P. L., Connolly, A. J., Crocker, J. H., Csabai, I., Czarapata, P. C., Davis, J. E., Doi, M., Dombeck, T., Eisenstein, D., Ellman, N., Elms, B. R., Evans, M. L., Fan, X., Federwitz, G. R., Fiscelli, L., Friedman, S., Frieman, J. A., Fukugita, M., Gillespie, B., Gunn, J. E., Gurbani, V. K., de Haas, E., Haldeman, M., Harris, F. H., Hayes, J., Heckman, T. M., Hennessy, G. S., Hindsley, R. B., Holm, S., Holmgren, D. J., Huang, C.-h., Hull, C., Husby, D., Ichikawa, S.-I., Ichikawa, T., Ivezić, Ž., Kent, S., Kim, R. S. J., Kinney, E., Klaene, M., Kleinman, A. N., Kleinman, S., Knapp, G. R., Korienek, J., Kron, R. G., Kunszt, P. Z., Lamb, D. Q., Lee, B., Leger, R. F., Limmongkol, S., Lindenmeyer, C., Long, D. C., Loomis, C., Loveday, J., Lucinio, R., Lupton, R. H., MacKinnon, B., Mannery, E. J., Mantsch, P. M., Margon, B., McGehee, P., McKay, T. A., Meiksin, A., Merelli, A., Monet, D. G., Munn, J. A., Narayanan, V. K., Nash, T., Neilsen, E., Neswold, R., Newberg, H. J., Nichol, R. C., Nicinski, T., Nonino, M., Okada, N., Okamura, S., Ostriker, J. P., Owen,

R., Pauls, A. G., Peoples, J., Peterson, R. L., Petravick, D., Pier, J. R., Pope, A., Pordes, R., Prosapio, A., Rechenmacher, R., Quinn, T. R., Richards, G. T., Richmond, M. W., Rivetta, C. H., Rockosi, C. M., Ruthmansdorfer, K., Sandford, D., Schlegel, D. J., Schneider, D. P., Sekiguchi, M., Sergey, G., Shimasaku, K., Siegmund, W. A., Smee, S., Smith, J. A., Snedden, S., Stone, R., Stoughton, C., Strauss, M. A., Stubbs, C., SubbaRao, M., Szalay, A. S., Szapudi, I., Szokoly, G. P., Thakar, A. R., Tremonti, C., Tucker, D. L., Uomoto, A., Vanden Berk, D., Vogeley, M. S., Waddell, P., Wang, S.-i., Watanabe, M., Weinberg, D. H., Yanny, B., Yasuda, N., and SDSS Collaboration (2000). The Sloan Digital Sky Survey: Technical Summary. *Astronomical, J.*, 120:1579–1587. 3

[Zehavi et al., 2002] Zehavi, I., Blanton, M. R., Frieman, J. A., Weinberg, D. H., Mo, H. J., Strauss, M. A., Anderson, S. F., Annis, J., Bahcall, N. A., Bernardi, M., Briggs, J. W., Brinkmann, J., Burles, S., Carey, L., Castander, F. J., Connolly, A. J., Csabai, I., Dalcanton, J. J., Dodelson, S., Doi, M., Eisenstein, D., Evans, M. L., Finkbeiner, D. P., Friedman, S., Fukugita, M., Gunn, J. E., Hennessy, G. S., Hindsley, R. B., Ivezić, Ž., Kent, S., Knapp, G. R., Kron, R., Kunszt, P., Lamb, D. Q., Leger, R. F., Long, D. C., Loveday, J., Lupton, R. H., McKay, T., Meiksin, A., Merrelli, A., Munn, J. A., Narayanan, V., Newcomb, M., Nichol, R. C., Owen, R., Peoples, J., Pope, A., Rockosi, C. M., Schlegel, D., Schneider, D. P., Scoccimarro, R., Sheth, R. K., Siegmund, W., Smee, S., Snir, Y., Stebbins, A., Stoughton, C., SubbaRao, M., Szalay, A. S., Szapudi, I., Tegmark, M., Tucker, D. L., Uomoto, A., Vanden Berk, D., Vogeley, M. S., Waddell, P., Yanny, B., and York, D. G. (2002). Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data. *Astrophys. J.*, 571:172–190. 5.4

[Zehavi et al., 2011] Zehavi, I., Zheng, Z., Weinberg, D. H., Blanton, M. R., Bahcall, N. A., Berlind, A. A., Brinkmann, J., Frieman, J. A., Gunn, J. E., Lupton, R. H., Nichol, R. C., Percival, W. J., Schneider, D. P., Skibba, R. A., Strauss, M. A., Tegmark, M., and York, D. G. (2011). Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity. *Astrophys. J.*, 736:59. 3.1, 3.1, 5.4

[Zel'dovich, 1970] Zel'dovich, Y. B. (1970). Gravitational instability: An approximate theory for large density perturbations. *Astron. Astrophys.*, 5:84–89. 2.1

[Zhao et al., 2013] Zhao, X., Li, Y., Shandera, S., and Jeong, D. (2013). The Effects of Local Primordial Non-Gaussianity on the Formation and Evolution of Galaxies. *ArXiv e-prints*. 2.1