Spatial Description of Large Scale Structure and Reionisation

Philipp Busch

SIGILL·UNIVERSITAT·LUDOVIC·MAXIM·

Garching 2019

Spatial Description of Large Scale Structure and Reionisation

Philipp Busch

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Philipp Busch
aus Berlin

München, den 24.04.2019

Erstgutachter: Prof. Dr. Simon D. M. White
Zweitgutachter: Prof. Dr. Volker Springel
Tag der mündlichen Prüfung: 28.06.2019

Contents

List of Figures

List of Tables

Introductory Remarks

Many of the results presented in this thesis have been or will be published in separate scientific articles. The work on the Tessellation Level Tree in chapter 2 will be published in Busch and White (2019c) and Busch and White (2019a). The third chapter about morphology of reionisation is based on a draft (Busch et al., 2019) also waiting to be published, while the contents of chapter 4 were part of Busch and White (2017).

This thesis is my own work. I chose to use the first person plural pronoun "we" to conform with the typical writing style in our field.

Synopsis | Zusammenfassung

This thesis is the result of a PhD project that tried to investigate and find new descriptions of entities arising in large scale structure based upon their spatial configuration. For this we analyse N-body simulations of gravitational collapse in a cold dark matter universe with cosmological constant (ΛCDM) and Monte Carlo ray-tracing radiative transfer (MCRTRT) simulations of reionisation. We also use an N-body simulations to investigate possible problems with observational results connected to large scale clustering.

In the first part of this thesis we develop a novel technique to characterise the density field in cosmological N-body simulations based upon a density estimate and the connectivity between particles obtained from a Voronoi tessellation of their positions. We use this estimate to find a hierarchical set of peaks in the Millennium and Millennium II simulations. This hierarchy completely decomposes the particle load of the simulations into nodes in a single tree structure we call “Tessellation Level Tree” (TLT).

We investigate the properties of these peaks and concentrate on two novel aspects: the percolation of the connected set of peaks above densities of a few $(6-7)$ along the cosmic web and the very strong assembly bias effect if peaks are split by saddle point density. This assembly bias effect is the strongest ever obtained from quintiles in a local property of the dark matter distribution in simulations.

The second part of the thesis investigates the morphology of ionised bubbles in hydrogen and helium during reionisation. For this we use MCRTRT on regular grids and create binary representations of the ionisation fields using a threshold. We then apply techniques of mathematical morphology to extract a hierarchy of bubbles ordered by local diameter.

We show the shift in the global bubble size distribution throughout reionisation and how the ionised volume is more and more unequally distributed among the bubbles as they grow and overlap. The overlap also results in a percolation process we identify at $z \gtrsim 8$ that increasingly delocalises the reionisation process. Finally, we connect the bubbles to the properties of the underlying density field. For the first time we show that the largest bubbles in the post-overlap regime are not densest in the centre are very strongly biased with respect to the large scale matter distribution. We also quantify how ionisation reaches the most underdense parts of the universe last, reconfirming the inside-out scenario of reionisation.

In the final part of the thesis we test the assembly bias and splashback radius measurements claimed by previous publications using clusters obtained with the optical cluster finder redMaPPer. For this we develop a mock-version of the algorithm that incorporates the core aspects of the cluster identification and apply it to a semi-analytic galaxy population of the Millennium simulation. We show that the claimed concentration differences in the optically selected clusters are most likely stemming from projection effects that arise more in overdense regions, leading to a coupling between concentration and large scale clustering and therefore a false positive assembly bias detection. The claimed splashback radius identification that is inverse in connection with cluster properties compared to the results of numerical simulations is shown to be an artifact of the circular mask of the selection algorithm.

Diese Arbeit ist das Ergebnis eines Dissertationsprojektes, das versuchte, neue Beschreibungen von Entitäten in großräumigen Strukturen basierend auf ihrer räumlichen Konfiguration zu finden und zu untersuchen . Dazu analysieren wir N-Körper-Simulationen des gravitativen Kollapses in einem Universum gefüllt kalter dunklen Materie mit kosmologischer Konstante (ΛCDM) und Monte-Carlo-Raytracing Strahlungstransfer (MCRTRT) Simulationen der Reionisation. Wir verwenden auch eine N-Körper-Simulation, um mögliche Probleme mit Beobachtungsergebnissen im Zusammenhang mit großräumigem Clustering zu untersuchen.

Im ersten Teil dieser Arbeit entwickeln wir eine neuartige Technik zur Charakterisierung des Dichtefeldes in kosmologischen N-Körper-Simulationen, die auf einer Dichteabschätzung und der Konnektivität zwischen Teilchen basiert, die aus einer Voronoi-Tessellierung ihrer Positionen gewonnen wurden. Wir verwenden diese Schätzung, um einen hierarchischen Satz von Dichtespitzen in der Millennium- und Millennium-II-Simulationen zu finden. Diese Hierarchie zerlegt die Partikelbelastung der Simulationen vollständig in Knoten in einer einzigen Baumstruktur, die wir "Tessellation Level Tree" (TLT) nennen.

Wir untersuchen die Eigenschaften dieser Dichtespitzen und konzentrieren uns auf zwei neuartige Aspekte: die Perkolation des verbundenen Satzes von Dichtespitzen über Dichten von wenigen $(6-7)$ entlang des kosmischen Netzes und den sehr starken Assembly-Bias-Effekt, wenn Dichtespitzen nach der Sattelpunktdichte getrennt werden. Dieser Assembly-Bias-Effekt ist der stärkste, der je aus Quintilen in einer lokalen Eigenschaft der Verteilung der Dunklen Materie in Simulationen erhalten wurde.

Der zweite Teil der Arbeit untersucht die Morphologie ionisierter Blasen in Wasserstoff und Helium während der Reionisierung. Dazu verwenden wir MCRTRT auf regulären Gittern und erstellen binäre Darstellungen der Ionisationsfelder mithilfe eines Schwellenwerts. Wir wenden dann Techniken der mathematischen Morphologie an, um eine Hierarchie von Blasen zu extrahieren, die nach lokalem Durchmesser geordnet sind.

Wir zeigen die Verschiebung der globalen Blasengrößenverteilung während der Reionisierung und wie das ionisierte Volumen immer ungleichmäßiger unter den Blasen verteilt ist, wenn sie wachsen und sich überlappen. Die Überschneidung führt auch zu einem Perkolationsprozess, den wir bei $z \gtrsim 8$ identifizieren, der den Ionisationsprozess zunehmend delokalisiert. Schließlich verbinden wir die Blasen mit den Eigenschaften des darunter liegenden Dichtefeldes. Wir zeigen erstmals, dass die größten Blasen nach der Überlappung nicht im Zentrum am dichtesten sind und einen sehr starken Bias in Bezug auf die großräumige Materieverteilung aufweisen. Wir quantifizieren auch, wie die Reionisierung zuletzt die am wenigsten dichten Teile des Universums erreicht, und bestätigen damit das Inside-Out-Szenario der Reionisierung.

Im letzten Teil der Arbeit testen wir die Montage Bias- und Rückfallradiusmessungen, die von früheren Publikationen unter Verwendung von Clustern, die mit dem optischen Clusterfinder redMaPPer erhalten wurden, beansprucht wurden. Dazu entwickeln wir eine Mock-Version des Algorithmus, der die Kernaspekte der Clusteridentifikation berücksichtigt und auf eine semi-analytische Galaxienpopulation der Millennium-Simulation anwendet. Wir zeigen, dass die behaupteten Konzentrationsunterschiede in den optisch ausgewählten Clustern höchstwahrscheinlich auf Projektionseffekte zurückzuführen sind, die eher in überdichten Regionen auftreten, was zu einer Kopplung zwischen Konzentration und großflächigem Clustering und damit zu einer falsch-positiven Verzerrung der Baugruppe führt. Die beanspruchte Rückfallradius-Identifikation, die im Zusammenhang mit Cluster-Eigenschaften im Vergleich zu den Ergebnissen numerischer Simulationen invers ist, erscheint als Artefakt der Kreismaske des Auswahlalgorithmus.

Ponder knew he should never have let Ridcully look at the invisible writings. Wasn't it a basic principle never to let your employer know what it is you actually do *all day?*

— Terry Pratchett, *The Last Continent*

Chapter 1

Introduction

Our work investigates a number of physical phenomena which we would like to introduce in the following chapter. They can generally be split into two groups: those directly connected with the gravitational collapse in the collisionless fluid of dark matter (DM) and the process of reionisation in the early universe. We finish this chapter with a basic introduction of the Delaunay and Voronoi tessellations. Each of the main chapters will also contain a more specialised introduction.

1.1 Results of Gravitational Collapse

Of the four fundamental forces gravity is the most important one for shaping the universe we observe on large scales today. It is of infinite reach (unlike weak and strong force which are mediated by massive particles) and its source field is not vanishing on large scales, in contrast to electromagnetism which is limited by large scale charge neutrality. It sets the stage on which all processes that light up the universe unfold, from the large scale distribution of galaxies, over their merging to their internal dynamics. In the context of this thesis mostly the first of these aspects is of importance as we want to characterise the large scale density field in relation to its small scale structure. We therefore want to introduce a few key concepts in the description of the large scale density variations in the universe in their relation to galaxies and the DM haloes that host them.

1.1.1 Haloes and Galaxies

In the current standard paradigm for cosmological structure formation, the concordance ΛCDM model, cold dark matter (CDM) dominates the cosmic mass budget and gravity drives structural evolution from the low-amplitude, Gaussian fluctuation field visible in the cosmic microwave background radiation to today's highly structured, nonlinear network, the cosmic web (Shandarin and Zeldovich, 1989; Bond et al., 1996). At late times this evolution occurs within a universe where the expansion is being accelerated by dark energy in the form of an effective cosmological constant, hence the Λ in ΛCDM. The cosmic web is built of overdense filaments and sheets which link dense, centrally concentrated structures called haloes. These form through anisotropic gravitational collapse and are the birthplaces and current hosts of galaxies (White and Rees, 1978). In the inner regions of haloes, dark matter densities reach values exceeding the mean by many orders of magnitude (e.g. Pandey et al., 2013).

This hierarchy of structures, subhaloes embedded in larger haloes which are in turn embedded in the cosmic web, is usually investigated with the help of cosmological simulations (see Bagla, 2005; Trenti and Hut, 2008; Frenk and White, 2012, for reviews). In recent years such simulations have increasingly included hydrodynamical modelling in order to treat the evolution of the baryonic components in addition to that of the dark matter (Schaye et al., 2015; Vogelsberger et al., 2014; Khandai et al., 2015; Dubois et al., 2016; Pillepich et al., 2018). A wide variety of algorithms have been used to identify galaxies, galaxy clusters and the cosmic web within such simulations. In particular, since dark matter haloes play such a central role, a large number of halo-finders have been developed. While all have the same goal, they differ significantly in approach; the intrinsic complexity of cosmic structure results in each identifying a halo population with somewhat different characteristics. For example, two of the oldest and most basic halo-finders are the friends-of-friends (FOF) (Davis et al., 1985) and spherical overdensity (SO) (Lacey and Cole, 1994) algorithms. The former often links almost disjoint haloes with low-density bridges which may sometimes reflect discreteness noise rather the true cosmic web. Such composite "haloes" are much less prominent in catalogues constructed with the SO algorithm, but these are geometrically biased by the spherical boundary which it imposes.

Most more modern halo finders explicitly address halo complexity by attempting to identify all subhaloes within each halo, where a subhalo is defined to contain a single significant local density peak. Subhaloes may defined in 3D configuration space, as in algorithms such as SUBFIND (Springel et al., 2001) and ADAPTAHOP (Aubert et al., 2004) and in the Tessellation-Level-Tree (TLT) studied here, or in 6D phase space as in ROCKSTAR (Behroozi et al., 2013). These algorithms are often supplemented by additional criteria such as requiring subhaloes to be gravitationally self-bound (e.g. Springel et al., 2001; Behroozi et al., 2013) or temporally persistent (e.g. Han et al., 2012, 2018). A more complete discussion of these issues and others can be found in Knebe et al. (2013)

1.1.2 Bias

Per unit mass haloes form preferentially in overdense regions. This simple fact is of profound consequence for cosmology, as the distribution of galaxies does not simply correspond to the distribution of matter in the universe. Why this is the case was first understood by Kaiser (1984). In our exposition we follow Coles and Lucchin (2002).

We begin with the assumption of a Gaussian random field with variance σ and correlation function $\xi(r)$ and its normalised form $w(r) = \xi(r)/\sigma^2$. In this field density peaks collapse under self-gravity if they exceed a critical density of $\delta_c \approx 1.68$ (Gunn and Gott, 1972). We can then parametrise the height of a peak as $\nu = \delta_c/\sigma$.

The probability of finding two collapsed points at a separation of r is given by the integral

$$Q_2(r) = \int_{\delta_c}^{\infty} \int_{\delta_c}^{\infty} \mathcal{P}_2(\delta_1, \delta_2, r) \mathrm{d}\delta_1 \mathrm{d}\delta_2 \tag{1.1}$$

of the multivariate Gaussian

$$\mathcal{P}_2(\nu_1, \nu_2, r) = \frac{1}{2\pi} \frac{1}{\sqrt{1 - w^2(r)}} \exp\left(-\frac{\nu_1^2 + \nu_2^2 - 2w(r)\nu_1\nu_2}{2[1 - w^2(r)]} \right). \tag{1.2}$$

The correlation of these points is then given by

$$\xi_{\nu_c} = \frac{Q_2}{Q_1^2} - 1 \tag{1.3}$$

with Q_1 the probability of finding point above δ_c.

Evaluation of the integrals above gives us a relation between the correlations of matter and collapsed peaks in the form of

$$\xi_{\nu_c} = b^2 \xi. \tag{1.4}$$

on large scales. Analytic approximations show a general proportionality of the form $b \propto \nu$. One example is given by Mo and White (1996):

$$b(M, z) \approx 1 + \frac{\nu^2 - 1}{\delta} \tag{1.5}$$

The dependence on $\nu(M)$ shows that more massive haloes cluster more strongly than matter.

This simple bias model has been expanded to include many more effects incorporated in higher order bias parameters. See Desjacques et al. (2018) for an extensive review.

Throughout this work we adopt a range of $6h^{-1}$ Mpc to $20h^{-1}$ Mpc to measure bias as the minimal radius is sufficiently large to be still pretty much linear and the large end does not take up too much of a box size in the simulations we consider. We also use the object-matter cross-correlation instead of the object auto-correlation as this gives us many more pairs and a better correlation function estimate, even for low object numbers and small radii.

1.1.3 Assembly Bias

Many additional properties of haloes related to their formation history, but unrelated to halo mass, are known to further separate haloes of equal mass in subsets with differently strong clustering (Gao et al., 2005; Wechsler et al., 2006; Gao and White, 2007; Faltenbacher and White, 2010; Lazeyras et al., 2017). This effect is therefore called *assembly bias*. There is still no detailed theoretical understanding of its origin, and our inability to measure the structure of individual dark haloes directly has made it difficult to identify observationally. Although there is no comprehensive model of assembly bias aspects of the phenomenon have been explained from the statistics of Gaussian fields (Dalal et al., 2008) or tidal effects (Hahn et al., 2009).

1.1.4 The Cosmic Web and Its Definitions

The cosmic web is a space-filling network of gravitationally collapsed regions, most strikingly, in the shape of linear filaments (collapsed in two dimensions) that meet at fully collapsed nodes and which encompass planar sheets, collapsed in one dimension. These sheets in turn surround underdense regions from which matter flowed out to undergo said collapse. While this behaviour was already visible in early work following the seminal publication of the eponymous approximation by (Zel'dovich, 1970) which itself predicted sheets (also called "pancakes"). It was not until the works of Shandarin and Zeldovich (1989) and Bond et al. (1996) that the terminology of a "cosmic web" was coined.

Just as for halo-finders, there is a plethora of algorithms which identify variously defined versions of the cosmic web (see Libeskind et al. (2018) for a discussion). While the detailed definitions differ, the primary aim of all of these algorithms is to define a space-filling filamentary network from the matter density field. In many cases, the classification also extends to find nodes of the network, as well as walls spanning between filaments, and voids surrounded by the network. Unlike the TLT we present below, these algorithms typically use a subsample of the simulation particles and/or a gridded or otherwise smoothed density field.

Knowledge of the cosmic web and its morphology has two main uses. On small scales it allows the environment of galaxies to be characterised, and thus furthers our understanding of the interplay between environment and galaxy formation from both observational (eg. Kraljic et al., 2018) and theoretical (eg. Borzyszkowski et al., 2017) points of view. On larger scales quantifying the morphology of the web may give information about cosmological parameters and the initial conditions for structure formation (see, e.g. Shim et al., 2014; Lee and Hoyle, 2015; Massara et al., 2015; Kreisch et al., 2018).

1.1.5 Splashback Feature

The splashback feature is produced by the outer caustics defined by material that is just reaching apocentre after its first passage through the inner cluster. The caustic radius is sharply defined for spherical infall models (e.g. Fillmore and Goldreich, 1984; Bertschinger, 1985; Lithwick and Dalal, 2011; Adhikari et al., 2014; Shi, 2016) but is significantly blurred, even in self-similar models, by realistic deviations from spherical symmetry (e.g. Vogelsberger et al., 2009). In a ΛCDM universe, these outer caustics give rise to a sudden steepening of the spherically averaged mean density profile before it flattens out at larger radii due to contributions from neighbouring haloes. This behaviour was studied in some detail by Diemer and Kravtsov (2014) who showed it to depend on halo mass, redshift and recent halo growth. Halo growth histories are intimately connected to their concentration, so Diemer and Kravtsov (2014) also looked for a systematic dependence of splashback signal on concentration. They found that the steepest slope attained by the mean density profile should become shallower and the radius at which it is attained should become larger as halo concentration increases. These results have been further improved upon in Diemer et al. (2017).

1.2 Reionisation

In the standard model of cosmology reionisation is a phase change from neutrality to ionisation in the gas (inter galactic medium, IGM) in the early universe by early sources of radiation. It is strongly theoretically (Barkana and Loeb, 2001) and observationally motivated (e.g. by spectra of high redshift quasars (Gunn and Peterson, 1965; Becker et al., 2001; Fan et al., 2006), although the exact conditions under which this process occurs are still unknown (Bowman et al., 2018). There are a number of experiments underway to shed light on this matter, mostly focusing on the 21cm-emission of neutral hydrogen (see Section 1.2.2.3), from global measurements, over interferometric measurements of the power spectrum (LOFAR Collaboration, 2013), to those capable of delivering actual images (Mellema et al., 2013). With the latter it will be possible to recover the spatial distribution of hydrogen, commonly known as 21cm-tomography (Loeb and Zaldarriaga, 2004). There are also bounds on the time of reionisation from the constraints on scattering of CMB photons by free electrons from experiments like WMAP (Komatsu et al., 2009) and Planck (Planck Collaboration, 2018).

1.2.1 The First Galaxies

The first galaxies form in exceptionally early collapsing dark matter haloes and are the first places in which gas has a chance to cool and undergo collapse (White and Rees, 1978; Ciardi and Ferrara, 2005). In the following we will give a short overview of the sources of the photons responsible for reionisation which all reside within these galaxies.

1.2.1.1 The First Stars

With the exception of some direct-collapse black holes (Pacucci et al., 2016; Wise et al., 2019) the formation of stars stands at the beginning of every source of radiation conceivably responsible for reionisation. They are therefore a vital component in our understanding of this process.

1.2.1.1.1 Population III The very first stellar population in the universe is called Population III (short Pop III, see Bromm and Larson, 2004 and references therein). While as of now there is no direct observational evidence beyond reasonable doubt, there is strong evidence of their existence. These stars would form in early collapsing haloes from completely pristine gas. As no cooling via metals is possible under these circumstances they need to be very massive to overcome gas pressure and can therefore only form in exceptional density peaks. As these peaks are quite rare, the Pop III stars are too rare to have an appreciable impact on reionisation, despite their extremely high individual luminosity and hard spectrum. Nevertheless they might play an important role in the formation of early massive black holes.

1.2.1.1.2 Population II The second and oldest observed population of stars, Pop II, forms after the Pop III stars in only slightly metal-enriched gas within the first galaxies and are hence of rather low metallicity ($-3 \lesssim [Z/H] < -1$). While only the lowest mass stars of this population survived to today, we can deduce that their high-mass brethren must have been numerous enough to drive reionisation. These more massive stars have much higher temperatures and therefore harder spectra. Their lower metallicity even increases its hardness.

1.2.1.1.3 Binarity Traditional stellar population models of cosmological simulations assumed single stellar populations (Bruzual and Charlot, 2003). More recent observations of star forming regions such as 30 Doradus in the Large Magellanic Cloud indicate that a large fraction of massive stars are actually part of binaries (Sana et al., 2012). This alters their evolution and vastly increases their output of ionising photons (Ma et al., 2016). While our simulations in this work are not yet appreciating this fact, new simulations are underway or even have been run, confirming the strong effect of an increased binarity on reionisation (e.g. Rosdahl et al., 2018).

1.2.1.1.4 X-Ray Binaries (XRBs) Even when not including the different evolution of high-mass stars in binaries it is important to take the remnants of such systems into account as only these branches of stellar evolution emit large amounts of X-rays. In these systems one star will undergo supernova first and turn into a compact object (neutron star (NS) or black hole (BH)). During its further evolution the other partner will increase its radius and at one point fill its own Roche lobe. At this point material from its atmosphere will rain down on the compact object and form an accretion disk around it. As falling into a gravitational well of such a compact object is the energetically most efficient process in existence ($\Delta E/mc^2 \gg 0.01$) the accretion disk is heated up to very high temperatures and emits a blackbody spectrum reaching well into the X-rays. This hard radiation has a very different effect on Hydrogen in the IGM and therefore reionisation than the UV emitted by the stellar progenitors (see Section 1.2.2.2).

1.2.1.2 The First Quasars

Stars aren't the only sources present during reionisation. We have observational evidence of high redshift quasars in the EoR (Becker et al., 2001; Fan et al., 2006). How these quasars are formed is still a much debated topic (Smith et al., 2017), as is their role and impact during reionisation (Madau and Haardt, 2015). In the reionisation simulations that we are working with (Eide et al., 2018a) they have a negligible effect on HII and HeII abundances but influence the temperature of the IGM, although possibly not as much as more numerous hard sources. Their hard spectra make them very important for fully ionising helium (see chapter 3).

1.2.2 Hydrogen Reionisation

Hydrogen is the largest contributor to mass and particles in the IGM and at the same time has a much lower first ionisation energy than the other major constituent, helium. This means that hydrogen will be ionised first and also leads to an almost completely ionised IGM on its own. It is therefore not uncommon to see treatments of reionisation neglect helium completely. For the same reason our analysis in chapter 3 concentrates on the ionised bubbles in hydrogen.

1.2.2.1 UV Ionisation

The primary process of ionisation of hydrogen in the IGM during reionisation appears to be UV radiation of massive early stars (Bromm and Larson, 2004; Eide et al., 2018a). These stars emit large amounts of photons that have very short mean free paths (Rahmati and Schaye, 2018) and therefore ionise the local IGM around the first galaxies to a very high degree. At the same time the heating due to UV photons is limited compared to that of X-rays and they are not able to ionise HeII (see next section).

1.2.2.2 X-Rays

X-rays have a twofold effect on the hydrogen in the IGM. First, they have a much lower interaction cross section for ionising atoms compared to an UV-photon and therefore have much longer mean free paths and deposit their energy over a much longer length scale. Second, they deliver much more energy to an electron if they do interact. This extra energy is deposited in the surrounding medium via collisions and possibly secondary ionisation events as a result of these collisions. Hence the IGM can be "pre-heated" ahead of the ionisation front formed by the UV radiation which changes the 21cm-signal one should expect (see Section 1.2.2.3). As we will see in chapter 3 we find X-Rays in our models to be insignificant for hydrogen reionisation and to only have an influence on the appearance of HeIII.

1.2.2.3 The 21cm Signal

A prime observable to study the state of hydrogen in the early universe is the emission from the hyperfine-transition in neutral hydrogen, in which the spin of the electron and the proton in a hydrogen atom are either aligned (↑↑) in a triplet state or anti-parallel (↑↓) in a singlet state. The difference in energy between these two states is

$$\Delta E_{21cm} = E_{\uparrow\uparrow} - E_{\uparrow\downarrow} \approx 5.874\,\mu\text{eV} \tag{1.6}$$

which corresponds to a frequency and wavelength of the emitted photon of

$$\nu_{21cm} \approx 1.42\,\text{GHz} \qquad\qquad \lambda_{21cm} \approx 21.11\,\text{cm} \tag{1.7}$$

which places it in the range of radio waves.

Even when one is able to clear the considerable foregrounds at these frequencies (Liu et al., 2009; Mertens et al., 2018) one is only able to observe a signal from this spin-flip transition if the spin-temperature of the system is different from the CMB temperature at the point of emission. Here the spin-temperature is simply a shorthand to characterise the population difference between the two states.

$$\delta T_b = 28\,\mathrm{mK}(1+\delta)x_{\mathrm{HI}}\left(1-\frac{T_{CMB}}{T_{spin}}\right)\left(\frac{\Omega_b h^2}{0.0233}\right)\sqrt{\frac{1+z}{10}\,\frac{0.24}{\Omega_m}}\left(\frac{H(z)/(1+z)}{\mathrm{d}\nu_{\|}/\mathrm{d}r_{\|}}\right) \tag{1.8}$$

1.2.3 Helium Reionisation and the 4.5cm Signal

1.2.3.1 Helium Reionisation

Helium is the second major constituent of the IGM and also undergoes a reionisation process. Due to its higher ionisation energies (24.6 eV for HeII and 54.4 eV for HeIII) the process of full ionisation is much slower, although the intermediate first ionisation closely follows hydrogen reionisation. Primordial helium contributes roughly $\frac{1}{7} \approx 14\%$ of electrons to the IGM.

The very high ionisation energy of HeIII also means that its reionisation can only be facilitated by hard sources beyond the range of stellar emission such as the accretion disks around X-ray binaries (see 1.2.1.1) and quasars (see 1.2.1.2). A simple calculation using Planck's law of the ratio of the intensities of a black body at the energies of the HII, HeII and HeIII ionisations reveals that a star has to have a surface temperature of $\gtrsim 3 \cdot 10^4$K to produce 1 HeI-ionising photon for each HI ionising one, while for HeII ionisation a temperature of $\gtrsim 9 \cdot 10^4$K is needed. While the former is attained in very high-mass main sequence stars as present during reionisation, the latter is not. A more thorough analysis given in Eide et al. (2018a) depicted in their Figure 2 shows that the drop-off is even more extreme when considering radiative transport in these stars.

The process of helium reionisation could theoretically be observed in radio frequencies very similar to the 21cm-line of hydrogen as is shown below.

1.2.3.2 The 4.5cm signal

Completely analogous to the 21cm-signal there is a similar hyperfine-split between two states of singly ionised ^{3}He whose nuclear spin does not vanish in contrast to the much more abundant ^{4}He. Then the electron spin has can couple to this nuclear spin and form a singlet or triplet state again. The energy ΔE, frequency ν and wavelength λ of the transition between these two states are (Weinreich and Hughes, 1954):

$$\Delta E \approx 27.8\,\mu\mathrm{eV} \qquad \lambda \approx 4.45\,\mathrm{cm} \qquad \nu \approx 6.739\,\mathrm{GHz} \tag{1.9}$$

Due to the much smaller abundance of ^{3}He (as indicated by the scale of 10^{-5} in Equation 1.10) the differential brightness temperature for HeII is four orders of magnitude lower than that of hydrogen:

$$\begin{aligned}\delta T_{b,\mathrm{HeII}} =& 1.614\,\mathrm{\mu K}(1+\delta)x_{\mathrm{HeII}}\frac{[^3\mathrm{He/H}]}{10^{-5}}\\ &\left(1-\frac{T_{CMB}}{T_{spin}}\right)\left(\frac{\Omega_b h^2}{0.0233}\right)\sqrt{\frac{1+z}{10}\frac{0.24}{\Omega_m}}\left(\frac{H(z)/(1+z)}{\mathrm{d}\nu_{\|}/\mathrm{d}r_{\|}}\right)\end{aligned} \tag{1.10}$$

1.3 Tessellations

A tessellation is the complete covering of a space with a set of finite, disjoint subspaces called tiles or cells. Simple examples are a Cartesian grid or, slightly more involved, a honeycomb. Both of these tessellations are monohedral, which means that they contain a single shape of cell that repeats to cover the space; their cells are congruent. This need not be the case for every tessellation and in the following we will only consider two special tessellations that are not monohedral: the Delaunay tessellation (or triangulation) and its dual, the Voronoi tessellation. Both of these tessellations have a long history in Astronomy, with an example of a Voronoi tessellation given by Descartes in 1644 in a description of the solar system. Where not indicated otherwise, we base our description of both of these tessellations below on Okabe et al. (2000), a very comprehensive resource on the topic for even everything the so-inclined reader might not have wanted to know about them. We will furthermore restrict ourselves to the simplest case of these tessellations in a Euclidean space of n dimensions.

1.3.1 Delaunay Tessellation

The Delaunay tessellation (DT) of a point set in a Euclidean n-dimensional space, also called the generators of the tessellation, is a complete covering of the convex hull of the generators, or the whole space in the case of periodic boundary conditions as in our applications, with a set of n-dimensional simplices (a polytope defined as the convex hull of $n+1$ points). It can be defined by a very simple requirement: the circumsphere of a given simplex with $n+1$ of the points at its corners must be empty. As it turns out this defines a unique covering with simplices if the generators are in “general position”, which means that not more than $i+1$ points lie on a i-simplex embedded in this n-space for $1 \leq i \leq n$.

An illustration of this procedure and its result is given in Figure 1.1. Here we show all steps in the procedure in one picture: the black generators define the grey empty circles with their red centres. The Delaunay simplices indicated by the black lines follow from the connection of these points.

This tessellation has many applications in data analysis in general and astronomy and astrophysics in particular as it gives a natural neighbourhood for points. This can be used to easily interpolate between values only known at certain positions, the most prominent example of which in astronomy is the Delaunay Tessellation Field Estimator (DTFE, Schaap and van de Weygaert, 2000). The naturality of the neighbourhoods can be better understood by turning to the Voronoi tessellation in the next section.

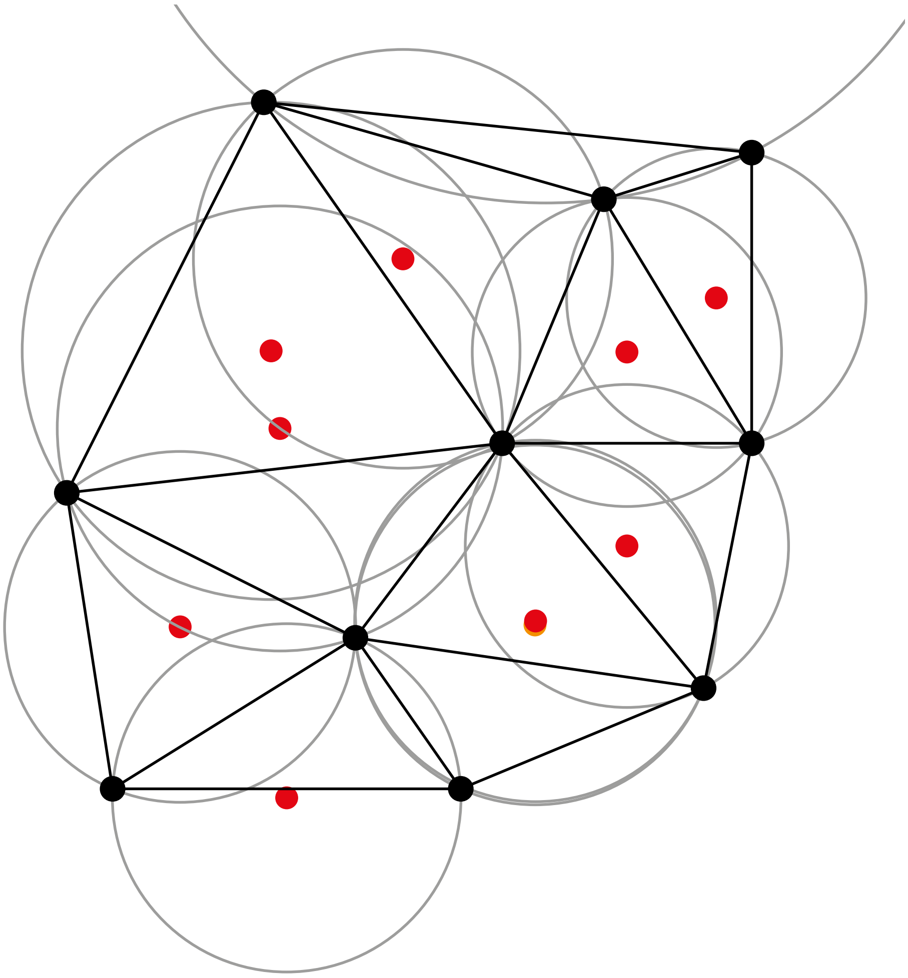

Figure 1.1: Example of a Delaunay tessellation in 2D showing the generators as black points, the circumcircles in grey, their centers in red and the Delaunay simplices (here triangles) using black lines.

1.3.2 Voronoi Tessellation

The n-dimensional Voronoi tessellation (VT) is very closely connected to the Delaunay tessellation and considered its dual. Instead of segmenting space between the generators it finds cells around each generator. In its simplest form these cells are defined as all the volume around a given generator to which it is the closest generator using the Euclidean norm.

To understand the connection to the Delaunay tessellation we recall the criterion which dictates that the ball with $n + 1$ generators on its surface must be empty. Therefore the center of the ball is equidistant only to these $n + 1$ generators. Departure from this point in any direction brings us closer to at most n and at least 1 of these generators. If we are equidistant to more than 1 generator we are on the ridge of the distance field (cf. Section 2.4.3 and 3.3.3). Due to the nature of the Euclidean norm, these ridges must be free of curvature. We can therefore construct them by first connecting the Delaunay circumcentres with lines if their accompanying simplices share a face (we stop here in 2D) as shown in Figure 1.2. In 3D we then find furthermore a planar face delimited by all of these connections of pairs that share a Delaunay edge. This means that neighbours connected by edges in the DT are connected via faces in the VT. This also means that the DT neighbours are those pairs that share faces in the VT, making them natural neighbours as there is an extended distance ridge between them.

The VT can also be constructed by placing the normal-plane of each Delaunay edge through its midpoint. The intersections of the half-spaces on either side of these planes form convex cells around each generator that have a face for each Delaunay edge originating from it. See Section 2.2.1.1 for a more technical description.

Direct construction of the VT is much more computationally expensive. It is therefore usually obtained from a DT just as in our case where we use the tessellation engine of AREPO (Springel, 2010).

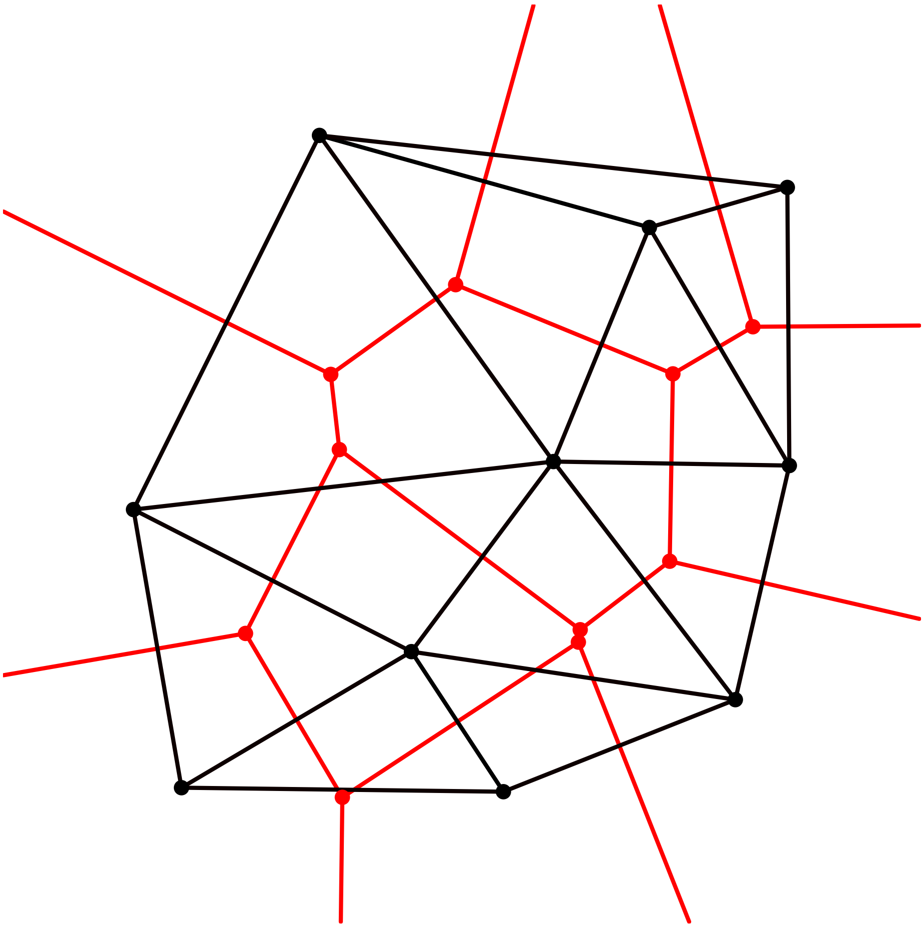

Figure 1.2: Example of a Voronoi tessellation in 2D and its dual Delaunay one. Just as in Figure 1.1 we show the generators as black points, the Delaunay simplices with black lines, the Voronoi vertices (which are the Delaunay circumcircle centers in red) and the faces of the Voronoi cells using red lines.

Chapter 2

The Tessellation Level Tree

This chapter is to appear in the upcoming publication Busch and White (2019c) and Busch and White (2019a).

2.1 Introduction

The described hierarchy of structures hosted within larger structures as discussed in 1.1 is usually investigated with the help of cosmological N-body simulations (see Bagla (2005); Trenti and Hut (2008) for reviews) that increasingly include hydrodynamical modelling (Schaye et al., 2015; Vogelsberger et al., 2014; Khandai et al., 2015; Pillepich et al., 2018). Different algorithms are employed to identify galaxies, haloes and the cosmic web within their outputs.

As the formation of galaxies in haloes is such a central problem to the study of structure formation with simulations, a large number of halo-finders has been developed. While all of them have the same goal there are very different approaches, the two most widespread of which are friends-of-friends (FOF) (Davis et al., 1985) and spherical overdensity (SO) (Lacey and Cole, 1994) halo finders. While the former suffers from spurious overlinking of disconnected haloes, the latter is hampered by its assumption of a spherical mass distribution. Some problems of classical halo finders can be alleviated by considering the binding state of particles (e.g. in SUBFIND (Springel et al., 2001)) and additionally extending the analysis to the full 6D phase space (e.g. Behroozi et al., 2013) instead of only the 3 spatial dimensions. This can be even further refined when taking temporal continuity into account (Han et al., 2018).

An alternative halo finder is VOBOZ (Neyrinck et al., 2005) which uses a Voronoi tessellation on the simulation particles to estimate their density. It then identifies density peaks and binds particles above a density threshold to them in a percolation step. An equivalent void identification algorithm called ZOBOV (Neyrinck, 2008) has been very successful and is part of the VIDE (Sutter et al., 2015) void identification toolkit. These structure finders share some commonalities with the algorithm presented in this work in that they estimate the density of each particle using a Voronoi tessellation and find clusters by percolation on a unstructured grid from the same tessellation.

It is common to all of these halo finders that they only consider the mass distribution above some density threshold far above the typical density of the cosmic web and therefore clearly distinguish between haloes and their environment. While this is sufficient to describe the peaks and by extension even the underlying matter distribution to a certain degree (Sheth and Tormen, 1999), it is neither able to, nor tries to, capture the morphology of the transition to the cosmic web or of the web itself (see 1.1.4).

We aim to connect the two regimes of haloes and the cosmic web by tracing the connectivity of the density field from the highest to the lowest densities in N-body simulations. For this we estimate each particle's density using its Voronoi cell volume and group particles to objects by connecting them over shared Voronoi cell faces.

This chapter is organised in two major parts: first we will describe the methodology to define a tessellation based density field and find a hierarchy of peaks in it, the Tessellation-Level-Tree (TLT), and then investigate its properties and those of its peaks. The method and additional methodology is presented in Section 2.2. We then use the TLT to investigate the percolation behaviour of matter above a density threshold in Section 2.4. Afterwards we present the abundance of peaks in this field and compare it with that of FOF objects in Section 2.5. Finally we shortly showcase two applications of the new methodology that will be further discussed in follow-up papers: the study of the mass-density distribution in haloes of varying mass in Section 2.6 and the strong assembly bias signal of the immediate environment of the peaks in Section 2.7.

2.2 Methodology

The methodology developed in this chapter consists of two parts: the construction of the Tessellation-Level-Tree and the description of its products are described in Section 2.2.1. This sub-section ends with a quick review of the terminology of objects as used in the remainder of this chapter. As some of the found peaks are the results of noise connected with the random sampling of the density field (as can be assumed in an evolved state of the simulations) we remove them using heuristics obtained from Poisson distributed points with a procedure described in Section 2.2.2. In Section 2.2.3 we describe how a set of peaks is used to construct catalogues of objects, some closer, some farther from usual halo definitions. We compare a catalogue of objects with similar construction to a classical FOF definition of haloes. Various properties, some of them novel, can be defined and computed for the objects obtained from the TLT. These are outlined in Section 2.2.4.

2.2.1 The Tessellation-Level-Tree

The aim of the Tessellation-Level-Tree (TLT) is to decompose the matter distribution in an N-body simulation into a hierarchy of peaks in the density field. A peak is defined as a local density maximum, the peak particles, and all the particles within the isodensity surface corresponding to the highest saddle point between this peak and a higher peak that do not belong to a sub-peak.

To obtain a density estimate for each particle we use a Voronoi tessellation generated from the particle positions. The inverse of the volume around the cell of each particle gives serves as our density estimate. The faces of the cells give us the connectivity on this unstructured grid. The tessellation is performed by the routines of the `AREPO` code (Springel, 2010).

A schematic of the method is presented in Figure 2.1. From the particle positions we obtain the tessellation structure (Section 2.2.1.1). This structure provides us with a density estimate for each particle and a neighbourhood in the form of a list of neighbours from which we construct the hierarchical set of peaks (Section 2.2.1.2).

2.2.1.1 Tessellation

The basis of the proposed analysis technique is the unweighted Voronoi tessellation (VT) $\mathcal{T}$ (see Section 1.3.2) in position space of the simulation particles P whose positions act as its generators. The cells of the tessellation are the regions in this space to which a given generator is closest in Euclidean distance. This construction leaves us with cells in the shape of convex polytopes. As we are using periodic boundary conditions all these polytopes will be of finite extent.

For this set of polytopes P we find the volumes V and shared faces to define particle densities

$$\rho_i = \frac{m_i}{V(p_i)}, \text{ for } p_i \in P \tag{2.1}$$

and a set of connections E. Particles are connected if their cells share a face.

For the following we impose a strict density ordering on the full particle set of the simulation. While in practice it is very unlikely to find two particles with exactly the same attributed volume, due to finite precision the probability is not zero. In case we do find two particles with same density we rank them randomly among themselves. While we account for the occurrence of this degeneracy, we do not expect it to appear in any real applications as the available state space in units of its granularity due to numerical precision is simply too large, especially when using double precision floating point values.

2.2.1.2 Peak Tree Construction

We traverse the density ranked list of particles in a decreasing fashion from the highest density particle in the box to construct a set of peaks. This procedure can also be reversed and started from the least dense particle. Then we would obtain a hierarchy of voids, similar to ZOBOV (Neyrinck, 2008).

A peak $\pi_i = (i, \pi_j, k)$ is an object with a first particle i, from which it inherits its rank, a parent peak π_j (initially π_i itself) and a last particle k (initially i). For each particle we keep two numbers, the peak π_i it belongs to and the next particle n_i in this peak. Together

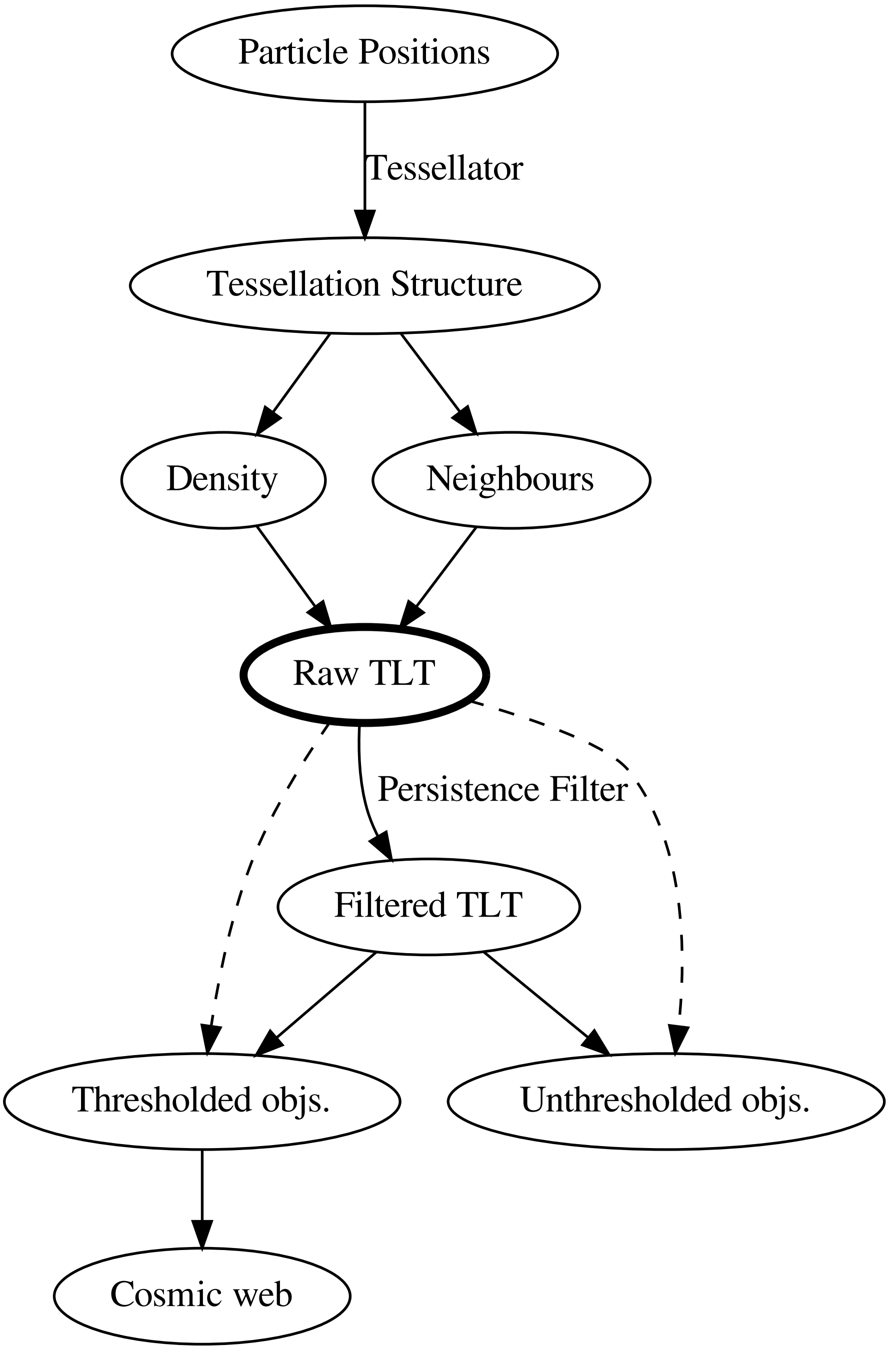

Figure 2.1: Overview of the pipeline for the TLT.

with the first and last particle entries of a given peak, the next particle entry allows to traverse the particles in a peak in the fashion of a singly linked list. One could also keep the rank of the previous particle in the list to make this a doubly linked list.

We traverse all particles in the simulation once, from highest to lowest rank. For each particle i we examine the ranks of the neighbours. For each particle one of two cases applies:

1. If the rank of this particle is higher than these of all its neighbours, we create a new peak object π_i which represents a local maximum and is identified by its first particle i.

2. Otherwise, there is one or more neighbours with a higher rank. These particles will have been processed before the current one and will have been assigned to a peak. We assign the current particle to the highest-ranked of these peaks. We then set this particle as the next particle in the chain for all higher neighbours which do not already have this set by another particle. This leaves us then again with two possible cases:
   (a) If all higher ranked neighbours belong to the same peak, we continue with the next particle.
   (b) If the higher neighbours belong to different objects, the current particle represents a saddle between these peaks. By virtue of the strict ordering there is always a first saddle that will also be processed first. In that case the peak with the higher density rank will incorporate the lower peak. Every particle processed afterwards which is connected to particles identified with either of the two peaks will be identified with the higher object. The lower peak can now be considered a sub-peak.

Each of the mentioned peak look-ups for a particle is a recursive operation on the peaks that follows the chain of parent peak entries until the peak is its own parent peak, i.e. a currently independent peak. As the process starts percolating, fewer and fewer peaks remain until finally all peaks are (possibly indirect) child peaks of the global maximum, the peak particle of the simulation so to speak.

An example of the described hierarchical segmentation for a 1D distribution is given in Figure 2.2. The accompanying peak tree is given in Figure 2.3.

In addition to the mentioned properties of peaks we can calculate physical properties on them. The simplest ones are the mass $M(\pi_i)$, volume $V(\pi_i)$ and the resulting mean

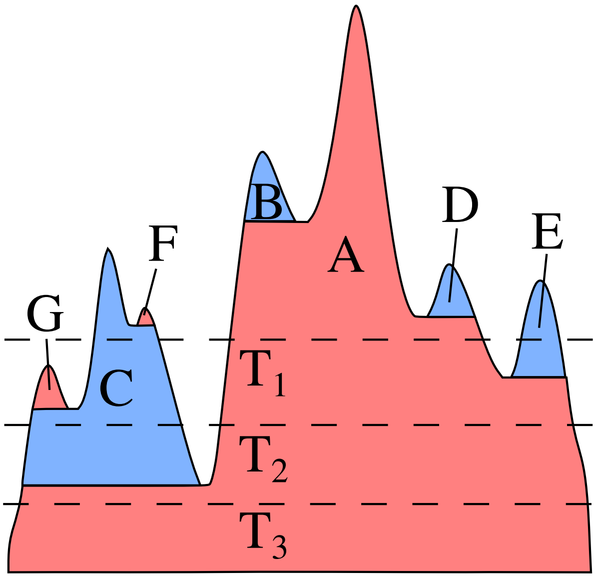

Figure 2.2: Schematic of the decomposition of a 1-D density distribution with peaks labelled alphabetically in decreasing peak density order and three thresholds T_1 through T_3. The resulting tree structure is shown in Figure 2.3.

density $\rho(\pi_i)$ of the peak π_i:

$$M(\pi_i) = \sum_{j\in\pi_i} m_j \tag{2.2}$$

$$V(\pi_i) = \sum_{j\in\pi_i} V(p_j) \tag{2.3}$$

$$\rho(\pi_i) = \frac{M(\pi_i)}{V(\pi_i)} \tag{2.4}$$

Here the mass and volume of a peak do not include their counterparts in the sub-peaks. The quantities including the sub-peaks, can be easily found by using the tree structure. One just adds the mass and volume of all sub-peaks and to that belonging to the peak. This structure also allows for the very quick calculation of quantities above a given threshold. For each peak whose range in density brackets the threshold one just adds the contributions of all sub-peaks that are joined above the threshold and then simply follows the next-particle chain until the threshold is reached. The above will be used when we present different halo definitions in 2.2.3.

Summarizing this section, we construct a decomposition of the set of particles in an N-body simulation into disjoint *peaks*, each a set of particles. Each peak consist of all particles that are reached first from a given *peak particle*, a particle of higher density than all its neighbours, when traversing the list of particles in descending density order. Each peak has a range in densities from the density of the respective peak particle down to the density of the first particle with an ascending path to a higher density peak particle, the

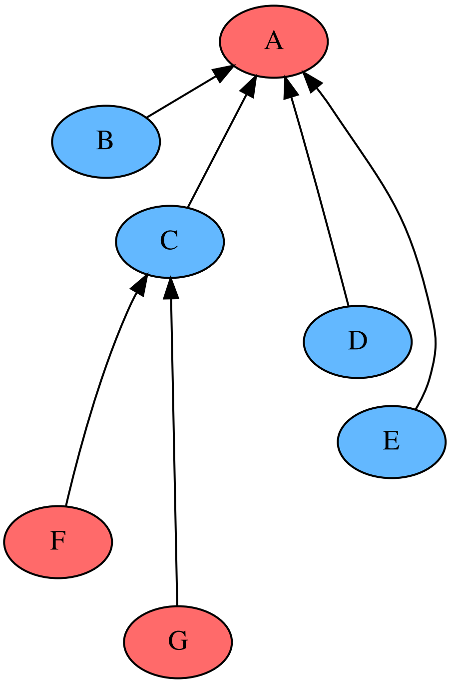

Figure 2.3: The peak tree of the example in Figure 2.2

saddle particle. Its density establishes a boundary on the range of densities in the peak and is therefore called the *limiting density* ρ_{lim}. The peak the saddle particle belongs to becomes the parent peak of the peak under consideration. The thus created hierarchy can be used to find many different halo definitions as detailed in 2.2.3.

2.2.2 Persistence and the Choice of the Density Estimator

The discretisation of a smooth density field with particles in the course of an N-body simulation leads to problems when estimating the density by the spatial distances of neighbouring particles. The particles trace the underlying density distribution only statistically and are therefore prone to exhibiting local density maxima which are purely results of the sampling.

In low density, single stream regions we expect the glass-like initial conditions to somewhat reduce this problem as the noise from the glass sampling is sub-Poissonian. Additionally, the dynamics in single stream regions will not create new peaks.

The problem therefore mostly arises in multi-stream, high density regions. The stream crossing and increasing importance of force smoothing for small scales transform the sampling of the density field by the particles towards a more Poissonian process. For an upper limit on the importance of the sampling effects we look at the local density maxima in a Poisson sampling of the field.

We tested two different estimators for the density at the particle positions, one based on the Voronoi tessellation (Equation 2.1), the other on its dual the Delaunay tessellation.

The Delaunay estimator as introduced by Schaap and van de Weygaert (2000) is very similar, but calculates a particle's volume as a quarter of the sum of the volumes of the adjacent Delaunay tetrahedra, V_D:

$$\rho_D = \frac{m_P}{V_D} = m_P \left(\sum_{c \in \mathcal{C}} \frac{1}{4} V(c) \right)^{-1}, \tag{2.5}$$

where $\mathcal{C}$ is the set of Delaunay tetrahedra incident on the particle. As each tetrahedron/Delaunay simplex is spanned by 4 particles this distributes the complete volume of the simulation on the particles.

We define the persistence r of a peak as the ratio between the densities of the highest and the lowest density particle. We can filter the TLT by a persistence criterion. If a peak does not pass a persistence threshold r_{th} it is removed from the hierarchy and its particles are grouped under its parent peak. Under the assumption that the physical peaks in the distribution are usually of higher persistence than the Poisson noise peaks we can filter the latter and retain the former.

To estimate the filter needed for a reliable Poisson peak filtering we show the ratio between the number of peaks above a given persistence threshold and the total number of peaks found in Figure 2.4. As in this case there should be exactly one particle of maximum density this directly translates into the false-positive peak identification rate.

We find that the Delaunay density estimator gives less reliable results in this Poisson case. For both estimators we find an asymptotically power law-like drop in the probability $P(r' > r)$ of a noise peak with persistence r' exceeding a given persistence value r. While the probability drops roughly as $r^{-2.2}$ for the Delaunay estimator its Voronoi counterpart drops as $r^{-4.6}$. From the results in Figure 2.4 we adopt a threshold of $r_{th} = 10$, which translates into a probability of $P(r' > 10) \sim 5 \cdot 10^{-4}$ of retaining a spurious peak in this extreme case. We repeat the experiment with varying numbers of particles, but as is to be expected from the local nature of the density estimation, the results are well converged.

The difference in behaviour can be explained if one takes a look at the distribution of estimated densities and pairwise density ratios as given in Figure 2.5. Here we plot the distribution of particle pairs (ρ_h, ρ_l) of higher and normalised lower density over $p = \rho_l / \langle\rho\rangle$ and the pair density ratio $q = \rho_h/\rho_l$. We find that the Delaunay density estimate varies much more, both among individual particles and between neighbours. For the same number of points the Delaunay densities vary over 3.5dex while the Voronoi ones have a range of 2dex.

This result might be confusing at a first glance since the relative standard deviation among the Voronoi cell volumes $\sigma(V_V)/\langle V_V\rangle \approx 0.42$ is higher than the expected standard deviation of the sum of the 27.1 ± 6.7 independent and identically distributed tetrahedron volumes that gives the Delaunay density estimate: $\sigma(V_D)/\langle V_D\rangle \approx 0.29$. The latter number assumes that the Delaunay tetrahedra incident on a given vertex are independent, which is clearly not the case. Therefore the relative standard deviation of the Delaunay estimate is much closer to the value of a single Delaunay tetrahedron $\sigma(V_{DT})/\langle V_{DT}\rangle \approx 0.83$, consistent with our finding of a poorer performance of the Delaunay density estimate.

These results are not in disagreement with previous results using the DTFE as given in Pandey et al. (2013). There the authors sampled each Delaunay tetrahedron with a test particle. This increases the number by a factor of $(24/35)\pi^2 \approx 6.77$ for the Poisson case (Okabe et al., 2000, p. 391) and further mediates the values by interpolating between the Delaunay vertex values.

A problem with this filtering approach is that we will also have a lot of false-negative events where a legitimate peak in the density field will be filtered out. We further investigate and quantify this issue in Section 2.5.2 for the simulation data. At this point the only remedy for this problem is the choice of a moderate persistence filter. We consider our choice of $r_{th} = 10$ as an acceptable compromise between false-positive and false-negative filtering.

2.2.3 Halo Definitions

The TLT allows for at least two definition of haloes, one of which is close to the typical `SUBFIND` substructure, while the other comes close to the FOF-definition by combining all particles of objects connected above a given threshold into one single object. This object then has a fixed bounding density, which corresponds to the limit of infinite resolution for FOF objects on the Voronoi density field at linking lengths corresponding to chosen density threshold. We can include or exclude substructures in the definitions below. If we exclude

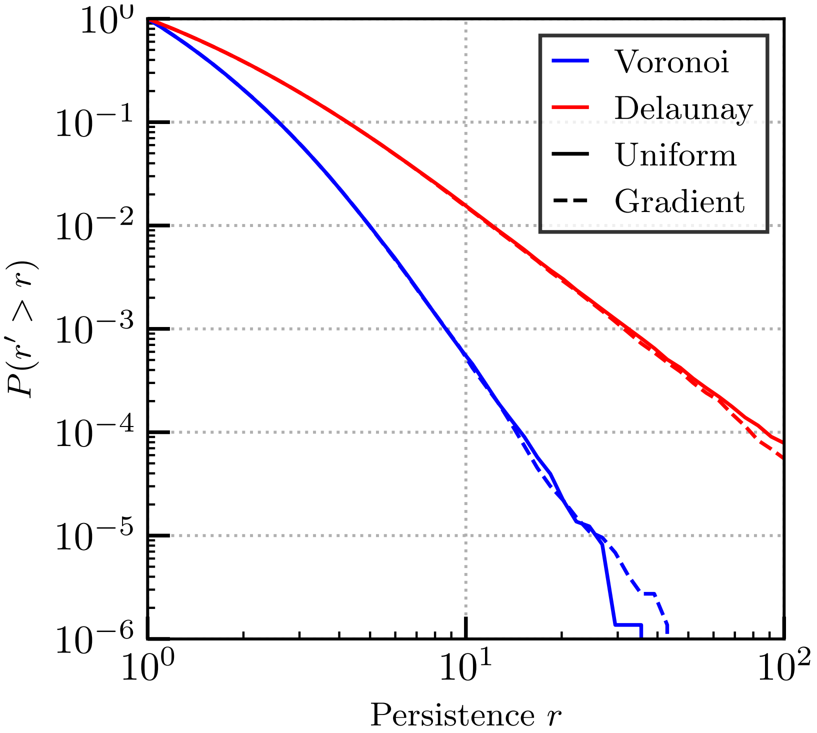

Figure 2.4: Results of an application of the Tessellation-Level-Tree on a Poisson distribution of particles. We show the probability $P(r' > r)$ of a given peak to be a false positive identification after applying a given persistence threshold r. The results are well converged for varying numbers of particles in a box.

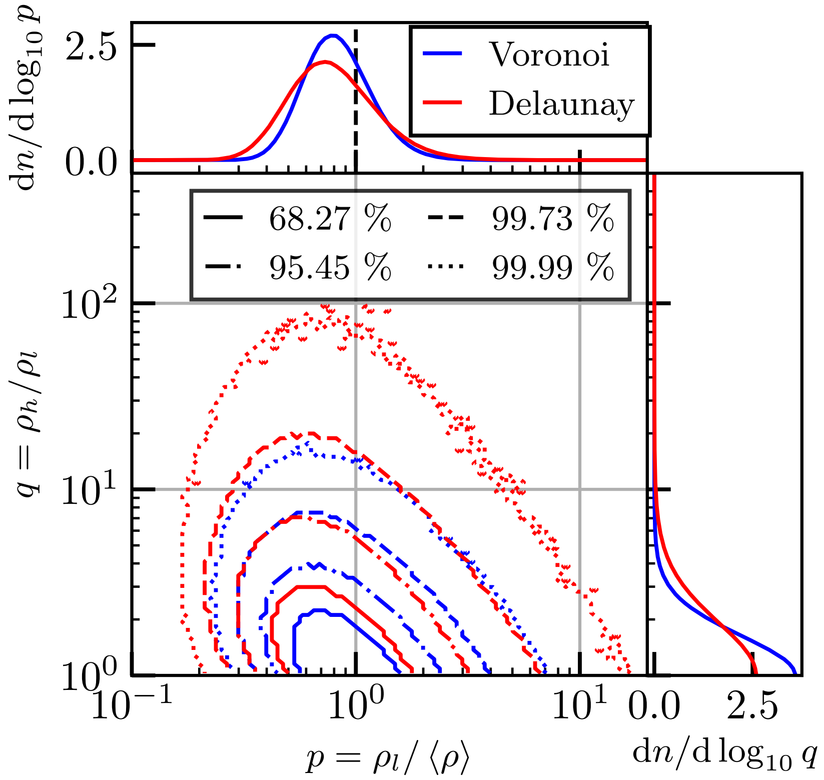

Figure 2.5: Distribution of the ratio of densities q between neighbouring points as a function of normalised density p of the lower density point in the pair for the 10^7 particle Poisson samples. Blue and red contours show the distribution of the Voronoi and Delaunay densities, respectively. The four different line styles signify the contours that enclose {68.27,95.45,99.73,99.99} per cent of the pairs. The two panels on the x- and y-axis show the marginalized normalised distribution of the pairs in logarithmic bins. The densities of the lower density particles in pairs vary much more using the Delaunay estimator than the Voronoi one as evident in the wider range on the abscissa for all but the highest density contour. The Delaunay estimator also leads to larger differences between neighbouring particles and therefore much larger q values. This test on a uniform Poisson-sample shows that the Voronoi estimator is more robust against this sampling noise.

them we denote this with a superscript "−" and if we explicitly include them with a "+". By default we include substructure.

2.2.3.1 Unthresholded Haloes

Each unthresholded halo (UH) corresponds to a node in the TLT and is therefore the entirety of particles connected to a local maximum above its limiting density ρ_{lim} at the first saddle point to a higher peak. As the only limitation of the object arises from the saddle point to the parent peak and no external threshold is imposed, we call these objects unthresholded.

A given UH^+consists of particles that are directly identified with the given peak as they were reached by it first and of indirect contributions by all the sub-peaks that are grouped below it. Their masses and volumes are the sums of the masses and volumes of all peaks in the subtree under the given peak in the peak-tree. The tree in Figure 2.3 contains seven UHs in both definitions. Each UH^+contains all peaks below it in the tree, so G is an UH of its own (as it is lacking substructure $UH^+$and $UH^-$are identical in this case) but is also included in the UH^+'s C and A. The substructure-less version $UH^-$is just a bare peak as defined in 2.2.1.2. In Figure 2.2 each single colored area would be one UH^-. The set of all $UH^-$is therefore a unique partition of the simulation particles.

These objects can exist anywhere in the whole density range of the simulation, apart from certain constraints coming from restrictions on their persistence and effects of the gravitational softening as discussed in 2.2.2. This means that an UH can be anything from a substructure in a classical halo, a classical halo itself to even a slightly less underdense region in a larger more underdense region. They are mainly of interest when looking at the structure of the segmented density field as a whole.

2.2.3.2 Thresholded Haloes

Thresholded haloes (THs) consist of all particles under a peak above a given threshold density, i.e. all particles above the threshold that are directly grouped under the peak and for the TH^+case additionally all particles in sub-peaks that are joined above the threshold. This definition establishes objects as the content of a connected bounding isodensity surface following the faces of the Voronoi cells in between neighbouring particles on different sides of the threshold.

As an illustrative example the peak structure in Figure 2.2 would give the thresholded peak sets (and therefore TH^-sets) {A,C,E}, {A,C} and {A} for the thresholds T_1, T_2 and T_3, respectively. If we include the substructures to obtain the TH^+definitions the sets would be {{A,B,D},{C, F},{E}}, {{A,B,D,E},{C,G,G}} and {{A,B,C,D,E,F,G}}. In each of these cases only the part above threshold would be included for peaks which extend beyond it.

As the FOF algorithm also tries to find objects with a given bounding density, its results come close to definition of a $TH^+$for suitable thresholds, but due to the nature of the percolation algorithm, no single threshold can be found for a given linking length as was shown by More et al. (2011) and is described for our case in 2.5.1.1.

The peak mass (Equation 2.2) and volume (Equation 2.3) are replaced with sums that do not extend over all particles in the peak but only these above ρ_{thresh}. The mean density in Equation 2.4 also uses these modified quantities.

2.2.4 Derivative Quantities

The structure of the peak tree and the knowledge about the density distribution of the simulation particles allows us to characterise the identified objects in a number of new ways in addition to more usual quantities such as halo mass. We mention these quantities in this work as part of the introduction of the Tessellation-Level-Tree but will further flesh them out and investigate them in a forthcoming publication.

The characterisation of substructure in haloes is a direct product of the TLT as the substructures are just child peaks. Their distribution and mass fractions can be readily extracted from the hierarchy of peaks.

Concerning dark matter halo density profiles we can easily distinguish between the mass in substructures and the main halo (just as, e.g. in SUBFIND). As an alternative to the usual approach of expressing density profiles as density as a function of spatial coordinate $\rho(\boldsymbol{x})$, we can reformulate the profile as cumulative mass as a function of decreasing density $M(>\rho)$. This formulation is directly reflecting the high density regions in substructures which are usually averaged over in radial profiles that use spherical or ellipsoidal shells. No assumption of any particular shape is made in our mass-density profiles.

We can also find new concentration definitions in accordance with the above approach to density profiles. These involve the ratio of masses above two different densities. The higher the share of the mass that resides at the higher density, the more concentrated the mass distribution. This approach has the advantage that it easily allows for the disentanglement of contributions from the main halo and substructure and directly captures the often implicit assumption of a representation of the density distribution of all matter in the system by the concentration parameter.

The validity of this assumption can be readily be tested by calculating the parameters of an NFW profile (Navarro et al., 1995) corresponding to the given mass-density profile. As the NFW profile has only two free parameters we only need two masses at different densities to determine it.

We can also characterise the shapes of our peaks at any given density using any of the shape estimators for haloes (e.g. the reduced tensor of inertia (Allgood et al., 2006)). This allows us to directly follow the change of shape with changing density.

2.2.4.1 Density Profiles

Traditionally a density profile describes the density of matter in radial or ellipsoidal bins around a chosen centre of a chosen object, either for all particles or for particles identified as not belonging to substructure. This way one records a density profile $\rho(\boldsymbol{x})$ in the position-density space, where the position is often just the radial distance. While this approach is very intuitive and useful for many applications, it also has its shortcomings. One of these is the smoothing of substructures over the spherical shells they reside in.

While we can still record these $\rho(\boldsymbol{x})$ profiles for all our objects, we can also record their mass above all the densities present in them. This gives us profiles of $M(\rho)$ as we descend through the density ranks. These profiles can be either recorded including the sub-peaks or without them and either until the limiting density is reached or down to a certain density threshold or even overdensity threshold. A given density profile and the mass-density profiles can be directly translated into one another as we show for the NFW profile in Section 2.2.4.2. We repeat this exercise for a different expression of the NFW which allows us to obtain the virial mass M_V and concentration c_V using the ratio of the masses of an object at any given pair of densities in Section 2.A.

While the particles directly belonging to the peak of interest are always treated the same and added to the profile as the density descend reaches them, there are two possible ways of treating sub-peaks if they are to be included in the mass. The first adds their total mass once their limiting density is reached; we call the result of this a *pseudo-radial profile* (PRP) $M_{pr}(\rho)$ as this resembles the treatment of substructure in the context of a simple radial density profile. For a peak π_i

$$M_{pr,i}(\rho) = \sum_{j \in P_i(\rho)} m_j + \sum_{\pi_k \in C_j(\rho)} M(\pi_k) \tag{2.6}$$

where $P_j(\rho)$ is the set of all particles directly connected to π_j at densities above ρ with m_i the mass of particle i. $C(j, \rho)$ is the set of all children of π_j with $\rho_{lim,j} \geq \rho$ and the peak mass $M(\pi_k)$ as defined in (2.2). This definition is equivalent for unthresholded and thresholded objects. Here and in the following the possible densities are of course limited by $\rho \geq \rho_{lim,i}$ for unthresholded and $\rho \geq \rho_{th}$ for thresholded objects with threshold ρ_{th}.

The second approach includes all particles in sub-peaks directly in the profile of all levels of parent peaks, so that their $M(\rho)$ profiles always reflect the total mass above a given density in this object. This is called the *total-mass profile* (TMP) $M_{tot}(\pi_i, \rho)$ for peak π_i. It can be expressed as:

$$M_{tot}(\pi_i, \rho) = \sum_{j \in P_i(\rho)} m_j + \sum_{\pi_k \in C_{S,j}} \sum_{l \in P_k(\rho)} m_l \tag{2.7}$$

where the same denominations as in (2.6) apply, with the additionally set $C_{S,j}$ of peaks in the subtree under π_i. In the case of thresholded objects with threshold ρ_{th} this definition has to be altered as only the subtree set $C_{S,j}(\rho_{th})$ of children with $\rho_{lim} > \rho_{th}$ is included. This profile is especially of interest for the prediction of the as of yet hypothetical self-annihilation radiation of dark matter.

If one refrains from including the mass in sub-peaks we obtain the *central-peak profile*

(CPP):

$$M_{cent}(\pi_i, \rho) = \sum_{j \in P_i(\rho)} m_j, \tag{2.8}$$

notation as in (2.6). This is also equivalently defined for thresholded and unthresholded objects.

Lastly one can also record the mass only in sub-peaks, giving the *sub-peak profile* (SPP) $M_{sub}(\pi_i, \rho)$. The SPP is hence simply the difference of TMP and CPP:

$$\begin{aligned} M_{sub}(\pi_i, \rho) &= M_{tot}(\pi_i, \rho) - M_{cent}(\pi_i, \rho) & (2.9) \\ &= \sum_{\pi_k \in C_{S,j}} \sum_{l \in P_k(\rho)} m_l & (2.10) \end{aligned}$$

with the same notation as explained under (2.7).

2.2.4.2 The NFW Mass-Density Profile

For all the different mass definitions from the previous section we can obtain mass-density profiles. This allows us to compare the mass obtained from these definitions with those of other halo finders by expressing them in a more traditional and therefore accessible picture. For this we need to express a given halo profile in the new variables $M(>\rho)$, the mass M above a given bounding density ρ, instead of $\rho(r)$, the local density at a given radius r from the centre. We then simply fit the profile to any of the $M(\rho)$ tracks obtained from the simulations as shown above.

The fitting formula used can be obtained by the lengthy, but elementary, inversion of the usual NFW-profile (Navarro et al., 1995), preferably executed by a computer algebra system. This leaves us with the following expressions:

$$\begin{aligned} t_1 &= r_s^3 \rho^2 (2\rho + 27\rho_0) \\ t_2 &= r_s^6 \rho^4 \rho_0 (4\rho + 27\rho_0) \\ t_3 &= (t_1 + 3\sqrt{2t_2})^{\frac{1}{3}} \\ t_4 &= 2 r_s \rho t_3 \\ t_5 &= 2^{\frac{4}{3}} r_s^2 \rho^2 \\ t_6 &= 2^{\frac{2}{3}} t_3^2 \\ t_7 &= 2^{\frac{1}{3}} \rho \\ M(\rho, \rho_0, r_s) &= \frac{4}{3} \pi r_s^3 \rho_0 \left(-\frac{(3(t_5 - 2t_4 + t_6))}{t_5 + t_4 + t_6} \right. \\ &\quad \left. + 3 \ln \left(\frac{r_s + (t_7 r_s^2)/t_3 + t_3/t_7}{3 r_s} \right) \right). \end{aligned} \tag{2.11}$$

Here r_s is the scale radius, ρ_0 the normalisation of the profile and ρ the density coordinate. For the inversion we have to assume all of the above numbers to be strictly positive reals, but this does not impose any restrictions.

2.2.4.3 Substructure Fraction

With the profiles introduced in the previous section we can define a measure of the substructure fraction for all densities at which a given peak exists. The ratio of SPP and TMP gives simply the substructure fraction f_{sub} above a given density:

$$f_{sub}(\pi_i, \rho) = \frac{M_{sub}(\pi_i, \rho)}{M_{tot}(\pi_i, \rho)} \tag{2.12}$$

One has to be careful when comparing this substructure fraction with a traditional definition which would be better approximated by

$$f'_{sub}(\pi_i, \rho) = 1 - \frac{M_{cent}(\pi_i, \rho)}{M_{pr}(\pi_i, \rho)}. \tag{2.13}$$

In this definition the mass of a substructure is counted all at once when it joins the main object, i.e. when the main object's density profile has reached the saddle point between them. If one were to convert densities to radii and compare the substructure fraction evolution over radii between radial bins and the converted mass-density profiles the result would closest resemble f'_{sub}, albeit under the assumption of a vanishing radial extent of the individual substructures.

The total versions $f_{sub,tot}(\pi_i)$ and $f'_{sub,tot}(\pi_i)$ of these quantities for the unthresholded object π_i is the value at the its limiting density $\rho_{lim,i}$ and thresholded objects at the threshold density ρ_{th}.

2.2.4.4 Concentration Definitions

In a very general sense the concentration of a dark matter halo is the relative amount of matter residing at high densities. In the model of a monotonic density profile this is connected to the steepness of said profile.

We can use the mass-density profiles introduced in the previous section to define a group of closely related concentration measures as mass ratios of which each has a focus on a slightly different application of concentration. The following section will show how one can use the mass values at two different densities to obtain an equivalent NFW profile which gives a classical measure of concentration.

A measure of the total mass above a given density in an object is

$$c_{tot}(\pi_i, \rho_1, \rho_2) = \frac{M_{tot}(\pi_i, \rho_1)}{M_{tot}(\pi_i, \rho_2)} \tag{2.14}$$

with $\rho_1 > \rho_2$ and M_{tot} as defined in (2.7). This measure is especially interesting for the expected signal of DM annihilation as it not only measures the mass in the central high density region as expected from a monolithic density profile but also in all substructures.

The central concentration

$$c_{cent}(\pi_i, \rho_1, \rho_2) = \frac{M_{cent}(\pi_i, \rho_1)}{M_{cent}(\pi_i, \rho_2)} \tag{2.15}$$

using M_{cent} from (2.8) only measures the concentration of the density profile of the central object under exclusion of all substructure. This characterises the underlying mass distribution in which the substructures move.

2.2.5 Halo Bias

The bias parameter b is a measure of the relative clustering of two fields. In this work the two fields are the matter and halo overdensity.

There are multiple possibilities of finding the bias parameter. Instead of comparing the auto-correlations of the two fields (Gao et al., 2005, c.f.) we use the auto-correlation of the matter particles and the cross-correlation between the halo positions and the matter particles (Gao and White, 2007). This has the advantage that the latter uses many orders of magnitude more pairs than the former technique and makes it therefore feasible to calculate bias parameters for very small numbers of objects. We are basically only limited by the sufficient sampling of the environment distribution of the objects of interest instead of their sufficient sampling in pairwise distance.

Following Gao and White (2007), we find a b that minimises

$$\sum_{i=1}^{4} \left(\log \xi_{hm,i} - \log b\xi_{mm,i}\right)^2 \tag{2.16}$$

in 4 spherical bins with radius $6 < r/\left(h^{-1}\,\mathrm{Mpc}\right) < 20$. We calculate the matter-halo cross-correlation ξ_{hm} and matter auto-correlation ξ_{mm} on cubic grids with 512^3 cells using simple nearest grid point deposition. This grid therefore has a cell length of $\Delta x_{MSI} \approx 0.977h^{-1}\,\mathrm{Mpc}$ in the MSI and $\Delta x_{MSII} \approx 0.195h^{-1}\,\mathrm{Mpc}$ in the MSII. Both, deposition and cell spacing are sufficient to resolve the scales relevant for our undertaking.

2.3 The Simulations

We calculated the structure of peaks on two different dark matter only cosmological simulations with the same cosmology: the Millennium and Millennium II simulations. Their parameters are given in Table 2.1. The latter is a box of 1/125 of the volume of the former using the same number of particles and hence 125 times better mass resolution.

The combination of both, a larger box with better large-scale statistics and a smaller one with better mass resolution and accompanying density contrast, allows us in the following to investigate the properties of resolved peaks at very different scales. At the same time we can check the convergence between the two simulations in the intermediate regime of masses and length scales that both boxes are able to represent.

We acknowledge that the cosmology of the simulations is close to but not in agreement with current measurements of the cosmological parameters from the CMB as given in Planck Collaboration (2018) or from the Dark Energy Survey Data Release 1 (DES Collaboration, 2018) (although we also note that the adopted value of h is very close to that

Table 2.1: Parameters of the simulations used in this work: Millennium I (MSI) and Millennium II (MSII).

	MSI	MSII
Ω_{dm}	0.205	
Ω_b	0.045	
Ω_Λ	0.75	
h	0.73	
σ_8	0.9	
n_s	1	
N_{part}	2160^3	
$M_{part}/\left(h^{-1}\,\mathrm{M}_\odot\right)$	$8.61 \cdot 10^8$	$6.88 \cdot 10^6$
$L_{box}/\left(h^{-1}\,\mathrm{Mpc}\right)$	500	100
$\epsilon/\left(h^{-1}\,\mathrm{kpc}\right)$	5	1

found by Riess et al. (2018)). Nevertheless, this does not change the qualitative nature of our findings. We can also still compare the quantitative strength of our assembly bias signal with previous assembly bias investigations using the Millennium simulations.

2.4 Matter Above a Density Threshold

Using the data structure for the dark matter density field and its connectivity as described in the previous sections, a first exercise is to investigate the properties of this field as a whole in an evolved state at $z = 0$. While others investigated the properties of the one-point density distribution (Pandey et al., 2013; Stücker et al., 2018) we want to concentrate on the properties of the content of isodensity surfaces.

The simplest properties of of the particle ensemble above a given threshold ρ_{thresh} are the total masses and volumes of the particles. These quantities develop with a changing ρ_{thresh} as the peaks in the density field merge or split when ρ_{thresh} is increasing or decreasing, respectively. These are given in Figure 2.6 and Figure 2.7. Here we show the fractions of the total mass $\sum M$ and the total volume $\sum V$ of all particles above the threshold, i.e. the sums of these quantities for all peaks. M_{Max} and V_{Max} are the mass and the volume of the most massive object at a given threshold. These values are accompanied by the masses of the second, third, tenth, hundredth and thousandth most massive object at a given threshold.

2.4.1 Total Mass and Volume

The total mass and volume above ρ_{thresh} have a very uniform behaviour over the whole range of thresholds in both simulations. The total mass is already close to the full mass at $\rho_{thresh} = 100\,\langle\rho\rangle$ although there are slight differences for the values between the simulations.

In the MSI ($\sum M_{\mathrm{MSI}}(100\,\langle\rho\rangle) \approx 0.45$, $\sum M_{\mathrm{MSI}}(\langle\rho\rangle) \approx 0.8$) there is slightly less mass at a given threshold than in the MSII ($\sum M_{\mathrm{MSII}}(100\,\langle\rho\rangle) \approx 0.55$, $\sum M_{\mathrm{MSII}}(\langle\rho\rangle) \approx 0.9$). The total values only change by less than a factor of 2 over two orders of magnitude in ρ_{thresh}. This means that less than half the total mass resides in objects one would commonly call a halo. This is to be understood as a lower bound as an increase in resolution increases the mass noticeably. The higher mass fraction at a given density in the higher resolution MSII is a result of its ability to represent smaller scale density perturbations and to follow their collapse into objects that would be washed out in the MSI.

The total volume above ρ_{thresh} evolves roughly as $\sum V \propto \rho_{thresh}^{-1}$ as is to be expected from the slow change in $\sum M$. The slope of $\sum V$ is slightly shallower for the MSII. This is again related to the higher resolution of the MSII. Not only is the mean overdensity higher, but also there are lower resolved underdensities. Therefore the range over which the same total volume as in the MSI is attained is wider leading to a shallower slope. In both simulations we find just below 0.1% of the total volume above $100\,\langle\rho\rangle$ while $\sim 8\%$ in the MSI and $\sim 6\%$ in the MSII lie above mean density.

2.4.2 Percolation

While the total quantities $\sum M$ and $\sum V$ show a very smooth and steady behaviour in the covered range, their counterparts M_{max} and V_{max} for the most massive object do not. They exhibit three distinctive behavioural regimes connected with two phases and an intermittent phase change. These phases are sub- and super-critical percolation of the isodensity surface with the phase change at the percolation density ρ_{perc}. This percolation process is the result of reaching a common isodensity envelope of the cosmic web. One can therefore say that ρ_{perc} is the bounding density of the cosmic web.

We first turn our attention to Figure 2.6 as the larger MSI captures the behaviour in a more pristine way. In the first phase at densities above ρ_{perc}, i.e. before the emergence of the infinite cluster, the highly ranked peaks grow exponentially in mass with decreasing ρ_{thresh} as surrounding peaks are linked and as is typical for sub-critical percolation. The merging of two or more very massive systems is visible as sudden jumps in mass. The exponential growth in mass is even visible for the 1000^{th} most massive object.

The phase transition from sub- to super-critical percolation happens in the range $6 \lesssim \rho_{thresh}/\langle\rho\rangle \lesssim 7$ and is visible in both M_{max} and V_{max}. The mass content of the largest connected isodensity surface increases from 1% to over 20% of the total mass in the simulation as the enclosed volume increases by the same factor from less than 0.02% to 0.4% of the box volume. Yet, at this point far less than half of the total mass and volume above ρ_{thresh} is contained in the infinite cluster. As one decreases the threshold more and more of the occupied volume along with the occupying particles are bound into the cluster and M_{max} and V_{max} approach $\sum M$ and $\sum V$, respectively. By the end of the depicted range at $\rho_{thresh} = \langle\rho\rangle$ virtually all mass is part of the largest object as almost no independent overdensities remain. The volume above ρ_{thresh} is not as completely bound in the infinite cluster, with still a noticeable difference between V_{max} and $\sum V$ at $\rho_{thresh} = \langle\rho\rangle$.

In the MSII the process of percolation depicted in Figure 2.7 is much less pronounced.

Up to $\rho_{thresh} \approx 8 \langle\rho\rangle$ there are two objects of almost equal mass of almost 10% of the total mass in the simulation. Subsequently there is a jump in mass of the largest object in the range $7 \lesssim \rho_{thresh}/\langle\rho\rangle \lesssim 8$ which nevertheless only increases it by a factor of 3. This is in stark contrast to the jump of more than an order of magnitude in the MSI. The contrast is even starker in the V_{max} development over a decrease in the threshold: In the MSII this quantity is increasing much more slowly and less smoothly than in the MSI. We interpret this as an expression of the limited box size and subsequent insufficient sampling of the high-mass cut-off of the halo mass function. The few high-mass objects that can be found in the MSII dominate its mass budget so much that the percolation does not produce such pronounced mass and volume increases as in the MSI. Just as in the MSI virtually all mass and close to all volume above $\rho_{thresh} = \langle\rho\rangle$ is part of the infinite cluster.

Turning to the n^{th} most massive objects as shown in Figure 2.6 and Figure 2.7, we find that all traced ranks in the MSI and all up to rank 100 in the MSII show the same sub-critical exponential mass increase as the most massive object discussed before. The third ranked object in the MSII actually undergoes a mass decrease just above $\rho_{thresh} = 10\langle\rho\rangle$ when the original object becomes part of the second or first ranked one. This is yet another effect only visible in the context of the limited box size in the MSII. The larger spread of masses for the same range of mass ranks has the same origin.

In the super-critical phase all but the most massive object become smaller and smaller as the previous holder of a given rank larger than 1 are merged with it. We see intermittent increases in mass for the very highest mass objects as they are growing to encompass part of their environment just before a link to the infinite cluster is formed.

There is a peculiar difference in the behaviour around the percolation threshold for different mass ranks. As noted before all ranks undergo a an initially exponential mass increase with decreasing ρ_{thresh}. With increasing rank this growth transforms into a mass decrease at offsets before the percolation threshold that increase with rank. This leads to the vanishing of a pronounced peak for the highest depicted mass ranks in Figure 2.7. From this we can estimate that the actual percolation happens by the merging of 10 to 100 patches per $(100h^{-1}\,\mathrm{Mpc})^3$.

2.4.3 The Geometry of the Cosmic Web

In this subsection we illustrate how to measure characteristic scales for the cosmic web, defined as the percolating structure identified by our TLT. The variation of its mass and volume fractions with ρ_{thresh} have been discussed above. For the MS case which gives the best statistics, the former increases from 24% to 80% and the latter from 0.4% to 7% as $\rho_{thresh}/\langle\rho\rangle$ drops from 6 to 1 (see Figure 2.6). In Figure 2.8 we show one layer of a 1024^3 Cartesian grid spanning the full MS volume. We colour black every cell that contains at least one particle belonging to the percolating object for $\rho_{thresh} = 5.25\langle\rho\rangle$. (This object contains mass and volume fractions of 35% and 0.62%, respectively). Clearly, this thin slice intersects the (single) percolating object many times. Indeed, defining two black cells in such a slice to be part of the same intersection if they share a face, we find that, on average, a slice intersects the cosmic web 618 times. Thus the mean distance between

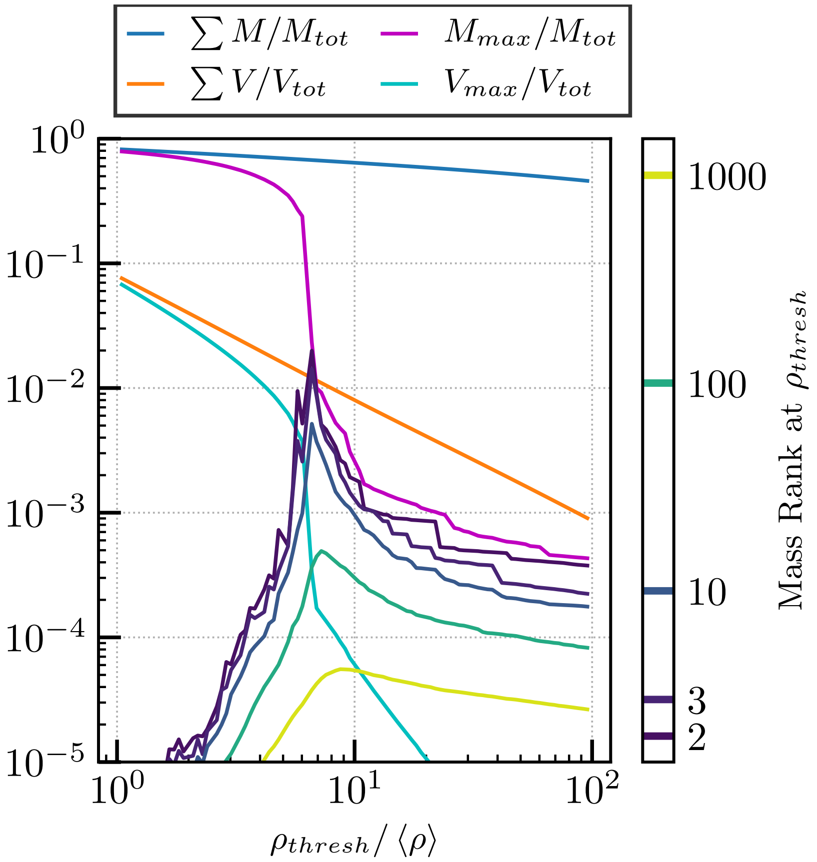

Figure 2.6: Mass and volume above a given threshold ρ_{thresh} in the MSI for all particles above the threshold ($\sum M$ and $\sum V$) and those in the most massive object (M_{max} and V_{max}). Additionally the masses of the second, third, tenth, hundredth and thousandth most massive object at a given threshold are shown (as identified by the colourbar). While the global quantities $\sum M$ and $\sum V$ show a very smooth behaviour there is a clear phase transition connected to a percolation process in the range $6 \lesssim \rho_{thresh}/\langle\rho\rangle \lesssim 7$. We identify the percolation threshold with reaching the density of an isodensity surface enclosing the cosmic web.

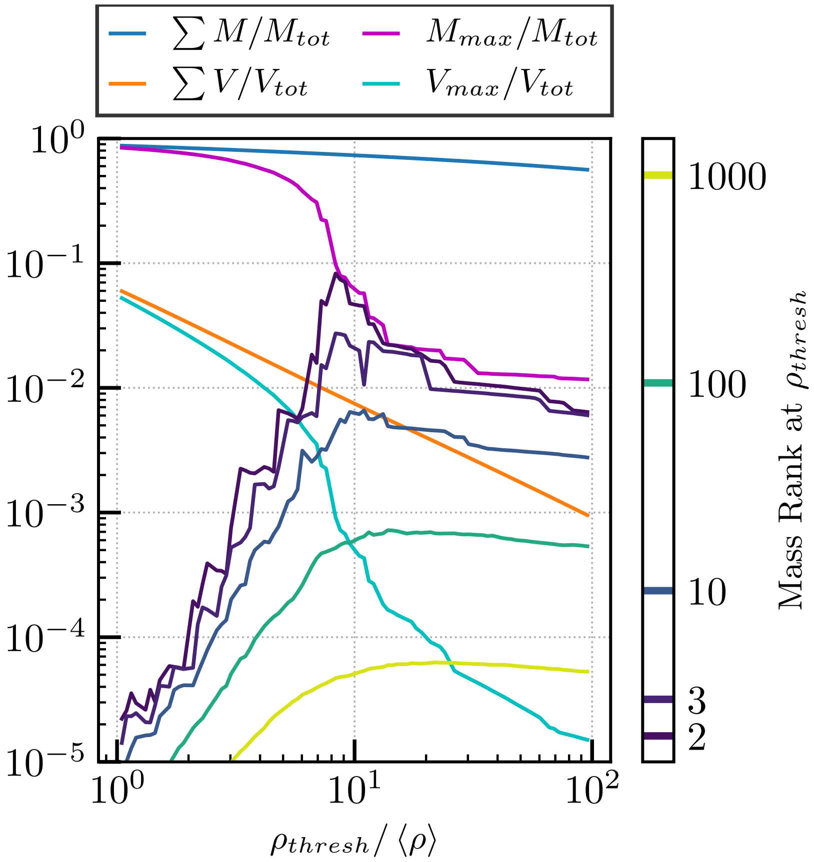

Figure 2.7: Same quantities as in Figure 2.6 but here for the MSII. The smaller box size of the MSII leads to a stronger dominance of the mass and volume budget of the largest objects. This in turn is noticeable as a less smooth and pronounced percolation transition.

such intersections is $20.1h^{-1}\,\mathrm{Mpc}$. This characterises the spacing between filaments of the web. Given that on average black cells occupy 1.71% of the slice area, the average area of an individual intersection is $(2.63h^{-1}\,\mathrm{Mpc})^2$. This length scale characterises the thickness of a filament. Since the web occupies 1.71% of the MS volume, the total length of filaments within this volume is approximately $\pi/2\ 0.171\ (500h^{-1}\,\mathrm{Mpc})^3/(2.63h^{-1}\,\mathrm{Mpc})^2 = 4.85\cdot 10^5 h^{-1}\,\mathrm{Mpc}$, where the factor of $\pi/2$ accounts for the fact that filaments intersect a slice like that of Figure 2.8 at random angles.

An alternative way to characterise length scales is shown by the coloured field outside the percolating object in Figure 2.8. This indicates the 3-D distance from each cell to the nearest cell which is part of the percolating object. This field is called the Euclidean distance transform (EDT) of the percolating cell set, and is defined by

$$E(\boldsymbol{x}) = \min_{\boldsymbol{w}\in\mathcal{W}} \|\boldsymbol{x}-\boldsymbol{w}\|\,, \tag{2.17}$$

where $\mathcal{W}$ is the set of positions of all cells within the percolating object and $\|\cdot\|$ denotes the euclidean norm. The local maxima of this field give the radii of spheres which are entirely outside the cosmic web but touch it at four points. Thus they are locally the largest spherical voids within the web. The particular slice shown in Figure 2.8 was chosen to contain the highest maximum of $E(\boldsymbol{x})$ (and hence the centre of the largest spherical void) in the MS volume, for which $R_{\mathrm{void,max}} \approx 50.5h^{-1}\,\mathrm{Mpc}$. This radius is consistent with the previous results of Cautun et al. (2016).

The properties of $E(\boldsymbol{x})$ can be used in many ways to to quantify the cosmic web. Here, we restrict ourselves to a simple example and to a test of convergence between our two simulations. In Figure 2.9 we show how the distribution of $E(\boldsymbol{x})$ varies with $\rho_{thresh}/\langle\rho\rangle$ in the MS, plotting the 10%, 50% and 90% points of the distribution, as well as its maximum. At the largest threshold shown, $R_{\mathrm{void,max}}$ is already significantly below its maximum possible value $\sqrt{3}\times 250h^{-1}\,\mathrm{Mpc} = 433h^{-1}\,\mathrm{Mpc}$, showing that the most massive object is much larger than an individual halo. As ρ_{thresh} is lowered $R_{\mathrm{void,max}}$ initially decreases slowly, but then drops precipitously to about $100h^{-1}\,\mathrm{Mpc}$ as percolation occurs in two steps over the narrow range $6.3 < \rho_{thresh}/\langle\rho\rangle < 6.8$. As ρ_{thresh} is reduced further $R_{\mathrm{void,max}}$ continues to decrease steeply, reaching a value of about $20h^{-1}\,\mathrm{Mpc}$ for $\rho_{thresh}/\langle\rho\rangle = 2$, the smallest threshold plotted.

The median and the upper and lower decile points of the distance distribution vary with ρ_{thresh} in a qualitatively similar way to its maximum value, but there are some notable systematic differences. The jump across the percolation transition varies substantially, from a factor of 3.5 for $R_{\mathrm{void,max}}$ to factors of 6.5, 9.8 and 12.7 for the 90%, 50% and 10% points, respectively. This reflects a broadening of the distance distribution which continues more slowly as ρ_{thresh} decreases further. The 10% and 90% points differ by factors of 2.8, 10.2, 10.3 and 11.8 for $\rho_{thresh}/\langle\rho\rangle = 7, 6, 5$ and 3, respectively. This change in shape is a consequence of the change in geometry from a single relatively compact object for $\rho_{thresh}/\langle\rho\rangle > 9$ to a volume-filling network of filaments for $\rho_{thresh}/\langle\rho\rangle < 6$. The substantial, quasi-exponential drop (from 12 to $2.3h^{-1}\,\mathrm{Mpc}$) in the median distance as $\rho_{thresh}/\langle\rho\rangle$ decreases from 6 to 2 is due to the growth of low-density filaments which

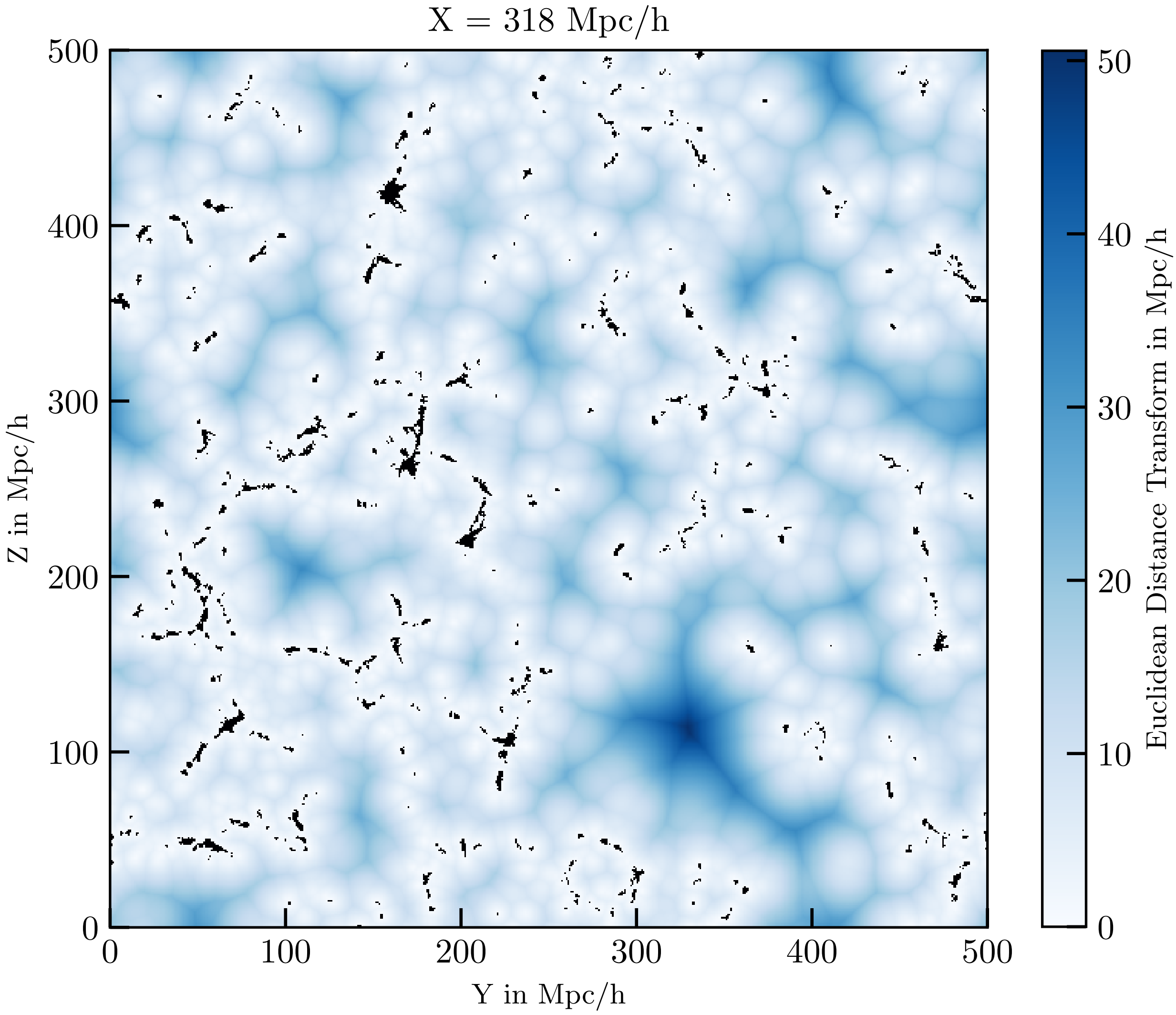

Figure 2.8: A slice through the MS showing (in black) the percolating object for $\rho_{thresh} = 5.25 \langle \rho \rangle$, and (in colour) the Euclidean distance transform (EDT) which gives the minimum 3-D distance from each point to the percolating object. This slice was chosen to contain the global maximum of the EDT, hence the centre of the largest void in the MS for this value of ρ_{thresh}.

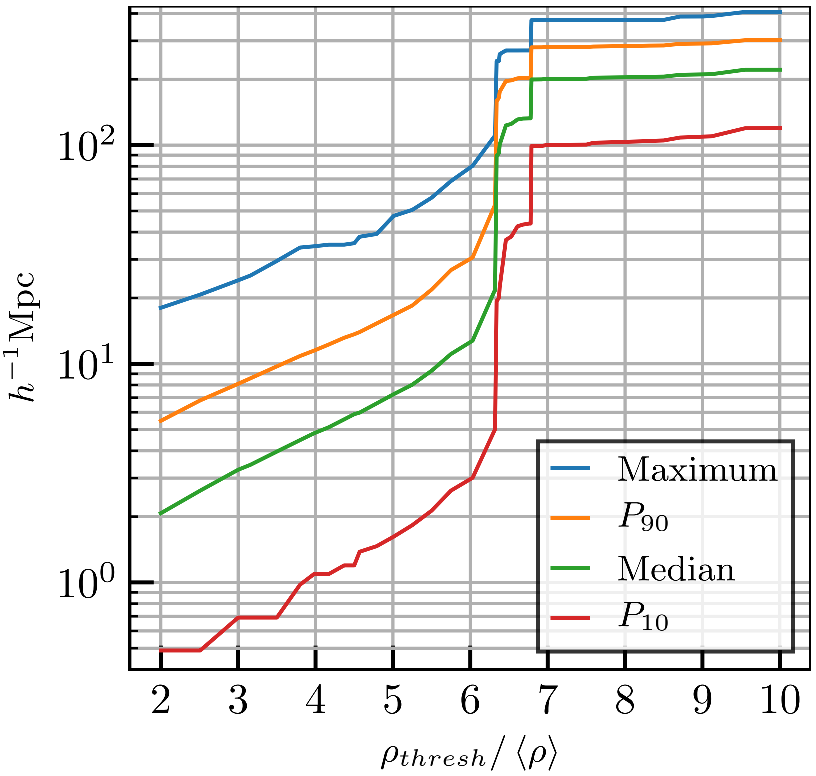

Figure 2.9: Variation of the distance distribution derived from the Euclidean Distance Transform (EDT) of the space external to the largest connected object in the MS as a function of the threshold density ρ_{thresh} of its bounding surface. The blue curve shows the maximum value of the EDT (i.e. the radius of the largest spherical void) while orange, green, and red curves give, respectively, the 90%, the median and the 10% points of the distance distribution at each value of ρ_{thresh}. Percolation is evident in the abrupt jump in these curves at $\rho_{thresh}/\langle\rho\rangle \approx 6.3$. The smaller jump at somewhat higher threshold is due to the merging of two objects of nearly similar size and corresponds to the lowest density at which the first and second most massive objects in Figure 2.6 are of similar mass.

extend from the higher density web into previously empty regions. Over this range, the total length of filaments in the MS (estimated as above) increases from $2.63 \cdot 10^5$ h^{-1} Mpc to $2.65 \cdot 10^6$ h^{-1} Mpc. Even for $\rho_{thresh}/\langle\rho\rangle = 2$ the percolating object fills only about 5% of the total volume, but the remaining 95% of the simulation is much more densely threaded with filaments than in Figure 2.8.

As seen in Figure 2.6 and Figure 2.7 percolation occurs at a higher density threshold in the MSII than in the MS, and the curves showing the mass and volume of the percolating object are both shifted slightly to the right in the MSII case. As a result it is not clear how best to check for convergence of web properties between the two simulations. In Figure 2.10 we compare the EDT distance distributions obtained when the thresholds in the two simulations are matched in such a way that they produce the same mass fraction in the percolating object. Specifically, we choose thresholds $\rho_{thresh}/\langle\rho\rangle \approx \{3, 4, 5\}$ and

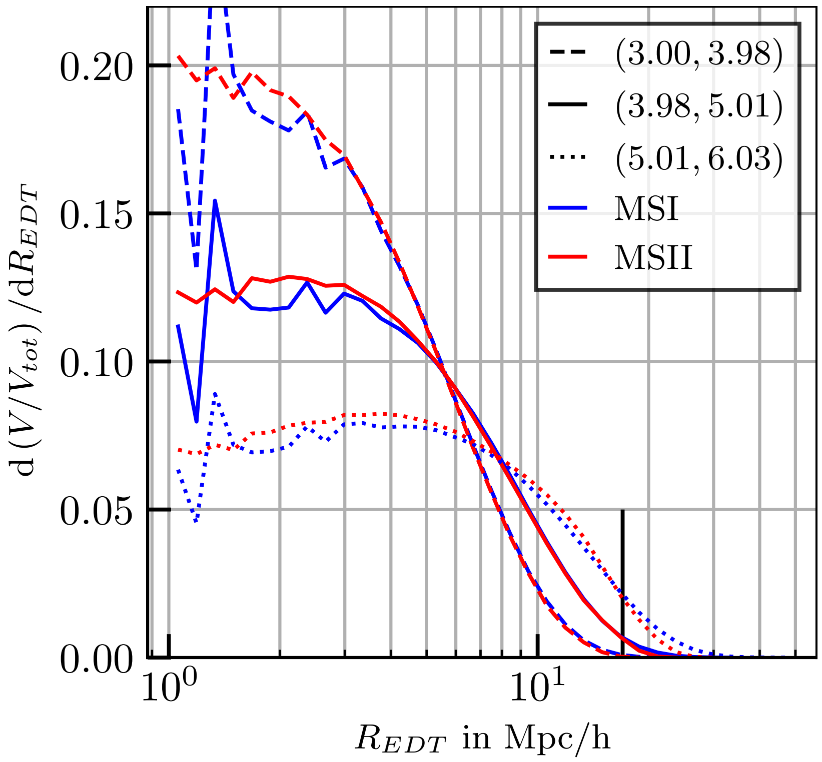

Figure 2.10: The EDT distance distribution relative to the percolating object at threshold densities of $\rho_{thresh} \approx \{3,4,5\}\langle\rho\rangle$ and $\rho_{thresh} \approx \{4,5,6\}\langle\rho\rangle$ in the MS and the MSII, respectively. At these thresholds the cosmic web contains mass fractions of $\{60,50,40\}\%$ in both simulations.

$\{4,5,6\}$ in the MS and MSII respectively, which leads to mass fractions of $\{60\%, 50\%, 40\%\}$ in the percolating object in both simulations. With this choice, the distance distributions agree remarkably well (apart from some small-scale discreteness effects in the MS) despite the difference in mass resolution of a factor of 125 and the change by a factor of two in the median of the EDT distribution over this range of thresholds. The longer tail to large distances in the MS for $\rho_{thresh}/\langle\rho\rangle = 5$ is clearly a reflection of its much larger volume; voids of $100h^{-1}$ Mpc diameter would not fit in the MSII simulation box. Most properties of the cosmic web as characterised by the EDT distance distribution are clearly very well converged in the Millennium Simulations.

2.5 Abundance of Peaks

The abundance of peaks is a simple quantity that allows us to form an understanding of the objects that come out of the process of constructing the Tessellation-Level-Tree. We begin with a comparison of the mass functions of thresholded objects at different thresholds and compare them with those of FOF-haloes in both simulations. We then further differentiate the set of thresholded objects closest to FOF-haloes by their ρ_{lim} values to understand in which local environments they live. We also conduct this investigations for unthresholded objects. We will give the abundances as number densities in units of $h^{-3}\,\mathrm{Mpc}^{-3}$. For the MSI a single object corresponds to an abundance of $n = 8 \cdot 10^{-9} h^{-3}\,\mathrm{Mpc}^{-3}$. For the MSII the minimum number density is $n = 10^{-6} h^{-3}\,\mathrm{Mpc}^{-3}$.

2.5.1 Thresholded Peaks

We begin our discussion of object abundances with thresholded peaks as these are closer to the usual notion of haloes which allows direct comparison to previous work.

2.5.1.1 Mass Function

We compare the mass function of thresholded peaks to that of FOF haloes with $b = 0.2$ in Figure 2.11 where we show the ratios of the cumulative mass functions for a range of threshold densities. We use thresholds from the set $\rho_{thresh}/\langle\rho\rangle \in \{60, 80, 100, 125\}$ (more precisely $10^{\{1.8,1.9,2,2.1\}}$) which are motivated by the analysis in More et al. (2011). There the authors also predict a number and resolution dependence of the bounding density of FOF objects which is nicely reproduced.

The ratios exhibit a large spread at large masses that decreases with mass until all ratios start to coincide as they drop together. This drop is most likely caused by the persistence filter (see Section 2.2.2). The large spread at large masses is a manifestation of the exponential cutoff in the mass function which is shifted to larger masses for smaller thresholds. As the MSI encompasses a larger volume than the MSII the effect is more visible here. The asymmetry in the deviations above/below the original MF for higher/lower thresholds is a result of the larger number of smaller objects than can be shifted to larger masses compared to the number of large objects that can loose sufficient mass by cutting away their peripheries.

While the halo content of the MSII does not extend to the same high masses as that of the MSI we still find a good convergence in the ratios up to the point where the decay in the MSI sets in. Its range $12 \leq \log_{10}(M_{halo}/\mathrm{M}_\odot) \leq 14.5$ covers an important observational range where one finds a large number of groups and clusters. The MSII ratios show a slower decline as objects of similar mass are better resolved.

We find that $\rho_{thresh} = 80$ to be the best choice for the threshold as it is the on average closest over the interesting range of well resolved objects.

2.5.1.2 Distribution in M-ρ_{lim}-Space

The thresholded objects in the TLT have an additional property to objects identified by other halo-finders: their limiting density ρ_{lim}. As it is a measure of the immediate environment of the object it is interesting to how objects of a given mass are distributed over these environments. For this we show the distribution of the thresholded objects at $\rho_{thresh} = 80\,\langle\rho\rangle$ in Figure 2.12. Here we compare the number density of peaks in the MSI

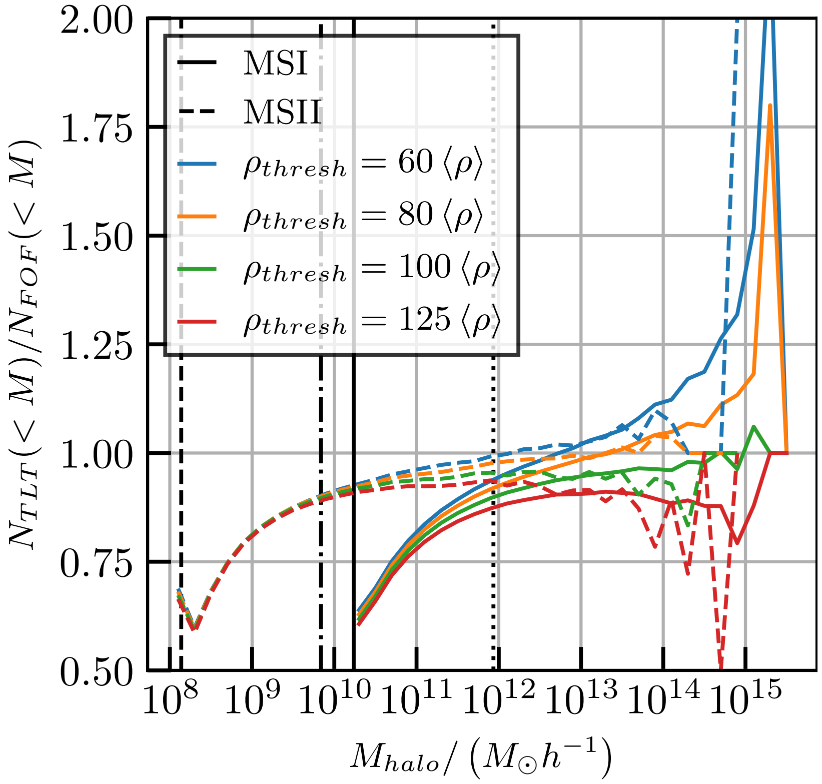

Figure 2.11: Ratios of the cumulative mass functions of TLT-peaks $N_{TLT}(< M)$ and FOF-objects $N_{FOF}(< M)$. The black vertical lines indicate the limits of 1000 and 20 particles in the MSI (solid 20, dotted 1000) and MSII (dashed 20 and dot-dashed 1000).

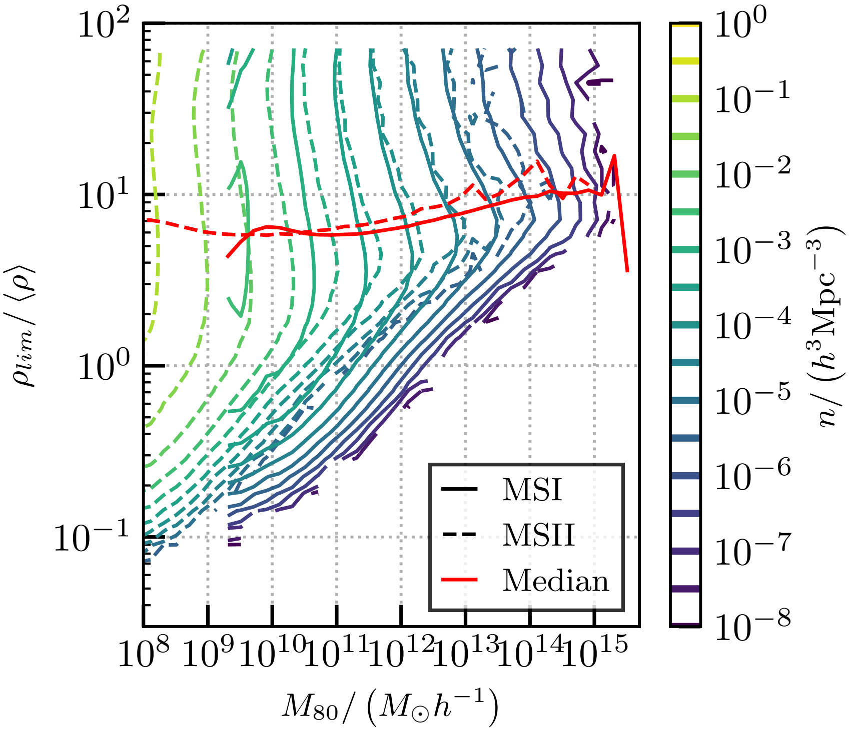

Figure 2.12: Abundance of objects of a given mass in the MSI (solid) and MSII (dashed) in dependence of their ρ_{lim} value. There is good agreement among the simulations for well resolved ($N_{part} > 1000$, $\sim 10^{12}h^{-1}\,\mathrm{M}_\odot$ for the MSI) objects inside the cosmic web.

and MSII by superimposing their sets of logarithmically spaced contours on a regular grid in log M-log ρ_{lim}-space.

We find the distributions for both simulations to be of qualitatively similar shape with some important differences in their detailed structure. Both the solid lines of the MSI-distribution and the dashed ones of the MSII exhibit an almost vertical alignment at ρ_{lim} values greater than $20\,\langle\rho\rangle$ before showing a bulge to higher masses just below the median ρ_{lim}. Below this bulge the contours slant towards low masses and roughly follow a family of power laws of slope ~ 2.5. The median ρ_{lim} values at a given mass agree well with each other and both decrease slightly with mass before going into an up-turn at the resolution limit (see lines in Figure 2.11).

While both simulations agree well with each other for massive objects at densities at or above the percolation transition (see Section 2.4.2) they do not agree for small ρ_{lim} values. The contours in the MSII turn to low masses earlier and are in general shallower. This could either be a consequence of more resolved channels of higher density that connect these objects to the web or less large voids compared to MSI due to the smaller size of the MSII. In this case there would less environments available in which objects could develop sufficiently removed from the web.

2.5.2 Unthresholded Haloes

Unlike the thresholded peaks from the previous section unthresholded peaks do not have a direct correspondence in the usual halo picture. We therefore only look at their abundance in $\log M$-$\log \rho_{lim}$-space. We use this distribution to investigate the effect of persistence filtering by comparing it to the distribution of the unfiltered peak populations in the same space.

2.5.2.1 Distribution in M_{tot}-ρ_{lim}-Space

To compare the abundance distribution of unthresholded objects in MSI and MSII we compare the contours of constant abundance in M_{tot}-ρ_{lim} space in Figure 2.13. Just as in Section 2.5.1 there is a regime of increasing ρ_{lim} with mass above ρ_{perc} in which the abundances are converged between MSI and MSII but here we see that indeed extends to much larger masses. Above this region we find a common deviation of the contours (see Section 2.5.2.2 for a more detailed discussion). Below ρ_{perc} the contours diverge again as the objects in the MSII do not extend to as low ρ_{lim} values as in the MSI at the same mass.

We find similar but slightly lower median ρ_{lim} values in the mass bins. The contours are similar to the thresholded case in the previous section but show a clearer spike to large masses at ρ_{perc}. This is simply a consequence of the additional mass gain between $\rho_{thresh} = 80 \langle\rho\rangle$ and ρ_{perc} by large objects in the cosmic web.

For high ρ_{lim} and small masses of a few particles, the two simulations diverge at $\rho_{lim} \approx 10^{4.5} \langle\rho\rangle$. This causes a noticeable excursion to high ρ_{lim} values to the left of the maximum formed by the previously mentioned turnover in the contours. We will further discuss this problematic region in the next section.

For the lowest ρ_{lim} values we find a flattening in the abundance contours for objects with less than ~ 100 particles.

2.5.2.2 Convergence Without Persistence Filter

As the unthresholded haloes represent the direct result of the TLT we can use their distribution to investigate the effect of the persistence filtering as laid out in Section 2.2.2. For this we compare the distribution after application of a persistence threshold of $r \geq 10$ from the previous section (Figure 2.13) to the abundances of the unfiltered output in Figure 2.14.

We see that there are no differences in the well-resolved medium to low density regime at higher masses. The contours do not change after application of the filter and hence still agree well in this regime.

We do find a strong increase in high ρ_{lim} objects with small masses. These objects form a characteristic 'hump' in the same region where the peculiar concavity with its two bracketing protrusions exists in Figure 2.13.

This pattern of divergence between the two simulations lets us conclude that although a simple persistence filter might help with the filtering of spurious peaks, it is not able to mitigate the problem completely. For the intents and purposes of this work it performs

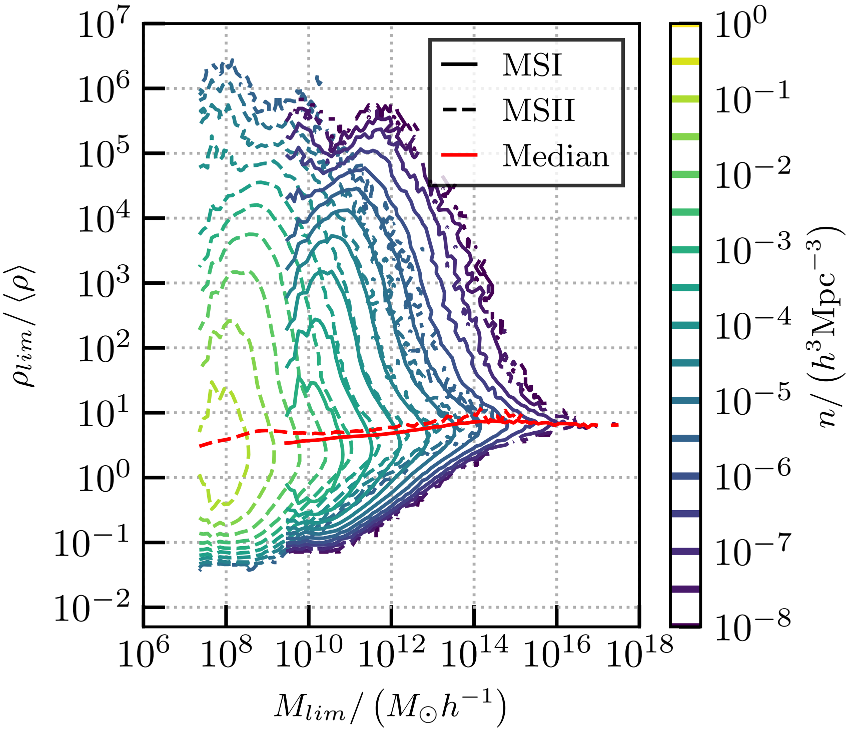

Figure 2.13: The distribution of unthresholded objects over mass and limiting density. Due to volume constraints the MSII is restricted to abundances above $10^{-6}(h^{-1}\,\mathrm{Mpc})^{-3}$

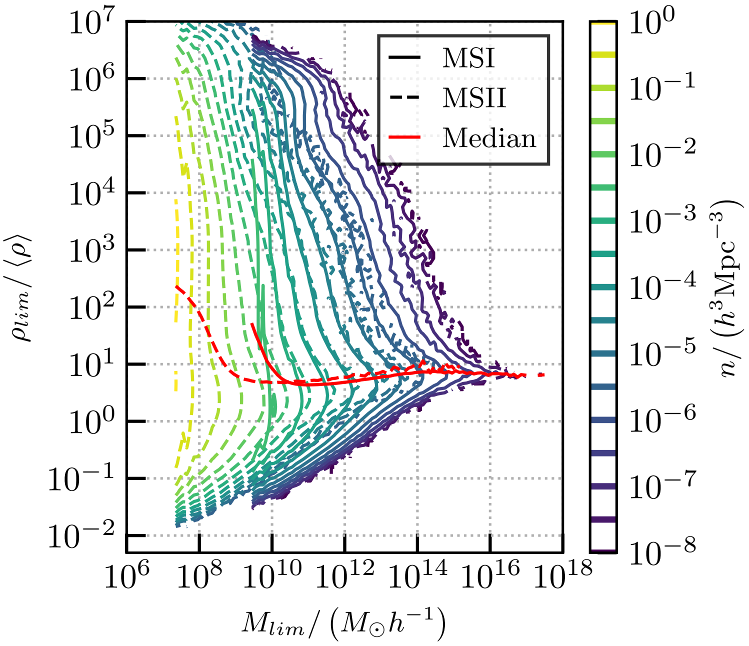

Figure 2.14: The distribution of the unfiltered unthresholded object population over mass and limiting density. Note the clear "hump" at high densities below $\sim$ 1000 particles ($10^{12}h^{-1}\,\mathrm{M}_\odot$ for the MSI, $10^{10}h^{-1}\,\mathrm{M}_\odot$ in the MSII).

reasonably well. As long as we ensure convergence with the well resolved region in the MSII or we deal with the high mass regime of the MSI we can trust our results.

2.6 Density-Mass Profiles of Peaks

The first application of the Tessellation Level Tree in the characterisation of density peaks are the mass density profiles we motivated and described in Section 2.2.4.1. We present them in two different ways, once with and once without substructure, binned by mass M_{80}, in the left and right panels of Figure 2.15, respectively.

Here we depict the median of the mass $M(\rho)$ in the given mass range as obtained at a given density. We decided to show the profiles with the mass on the abscissa as it is a more traditional indicator, used for example in stellar astrophysics. The circles indicate the median values as obtained from the simulation and the solid lines show the fit using a reformulation of the NFW profile as a function of mass instead of radius. The bars at the lowest mass bin indicate the the range of the quartiles around the median. The grey dotted lines indicate the range of the fitting.

The fitting was restricted to the points where the given median profile had at least 50 particles and to densities outside of the previously established convergence radius of 3 times the softening radius. The lower bound in density is formed by the density threshold.

We used (2.11) to fit the profiles using the original NFW-parameters ρ_0 and r_s instead of the equivalent expression in (2.21) which uses M_V and c.

We find that both cases are well fit by NFW-profiles within the fitting range. We see a systematic departure towards lower densities at small masses (small radii) as the gravitational softening limits the attainable central densities.

Interestingly for massive objects the profile including substructures is smoother and better fit by a single NFW profile. The one without substructures indicates a slight deficit in mass at intermediate densities and an excess at higher densities. This might indicate that the densest point in a peak does not necessarily lie in the most massive, but most concentrated substructure.

We will extend the discussion of the internal structure of objects in the TLT in Busch and White (2019b).

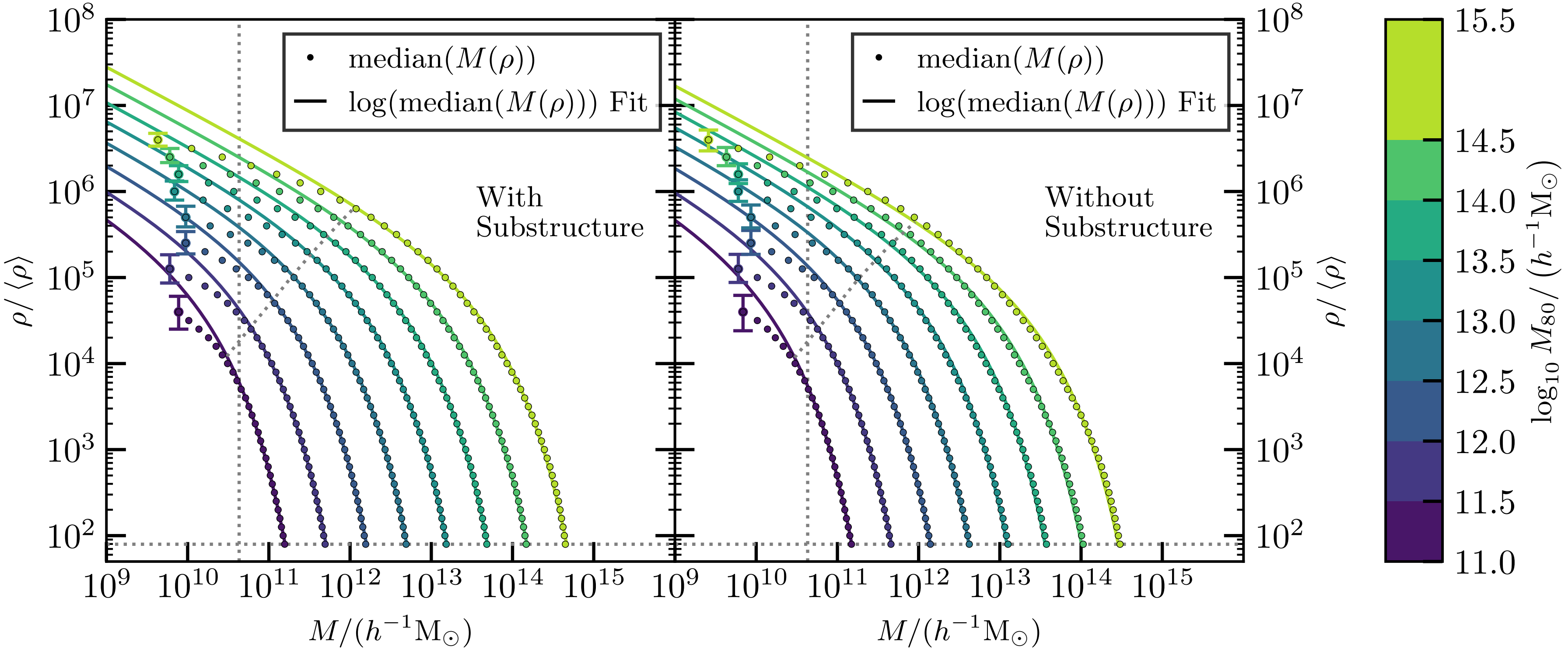

Figure 2.15: Median density profile as function of mass for peaks binned in M_{80} as indicated on the colour bar above. The left panel includes the mass in substructures (TH$^+$) while the right panel only includes mass directly under the peak at a given threshold density (TH$^-$). We present the median masses at a given density threshold down to a limit of 50 particles. We limit the fitting of the profiles to mass above the defining threshold of 80 $\langle\rho\rangle$ and to densities below that given by the the NFW profile at three softening radii $\epsilon = 5h^{-1}$ kpc. We mark the densest point still fitted by indicating the range in between the 25- and 75-percentiles.

2.7 ρ_{lim}-Assembly Bias

The limiting density of peaks is a new local property of dark matter structure. It was previously shown, that selection on local properties can lead to changes in the global clustering statistics resulting in assembly bias (see 1.1.3). In the following we test how ρ_{lim}-selected samples might differ in their clustering.

2.7.1 Assembly Bias in ρ_{lim} Quintiles

In analogy to the procedure in Gao and White (2007), we split the sample of thresholded haloes with $\rho_{thresh} = 80 \langle\rho\rangle$ from MSI and MSII binned in 0.5dex in M_{tot} in quintiles in ρ_{lim} whose boundaries are shown in Figure 2.16. We then calculate the bias value b for each of these sub-samples as introduced in Section 2.2.5. The results in Figure 2.16 show a very clear split between the samples that indicate a stronger assembly bias signature than in any property previously investigated in this way. We additionally present the bias relation of all objects in a given mass bin in comparison to that of FOF haloes. The two agree perfectly over 5 orders of magnitude in mass. We only show bins with at least 100 members to reduce the uncertainties in the halo-matter cross correlations.

The differences between the quintiles are so strong that we find the bottom quintile with masses $M_{tot} \leq 10^{0.5}M^* \approx 10^{13.3}h^{-1}\,\mathrm{M}_\odot$ that are *uncorrelated* with the large scale matter distribution. As we will see in Figure 2.20 in the next section, choosing a lower cut such as the bottom decile would actually lead to a negative bias, i.e. anti-correlation with the density field on large scales.

The spread in bias values increases strongly with Mass starting at $10^{-0.5}M^*$ while it stays rather constant below this mass. This is caused mostly by the delayed departure of the bias of the lowest quintile from its constant value of $b \approx 0$ in the low-mass regime at which it stays up to $10^1 M^*$. Above this mass it rises quickly towards the values that the upper four quintiles already reached after their more steadily rising increase.

To explain this behaviour we have to turn to the more detailed investigation in Figure 2.20 where we will see the lower end of the distribution of limiting densities rise quickly with increasing mass. This rise is quicker than the slope of the zero-bias contour owed to the fact that there are groups but not clusters in large-scale underdense regions, separated from the cosmic web. For the analysis in Figure 2.16 this means the populations in the quintiles shift from structures in voids to objects in parts of the web that are connected via slightly less dense channels.

This behaviour is shown in Figure 2.17 which depicts the four boundaries between the five quintiles and the median ρ_{lim} in the MSI and MSII. We find a constant shift of slightly less than an order of magnitude in mass between the quantiles in the two simulations. As most of the quantiles are rather flat they do not shift much in value apart from the lower ones at high masses and the lowest over the whole mass range where these curves have a stronger slope.

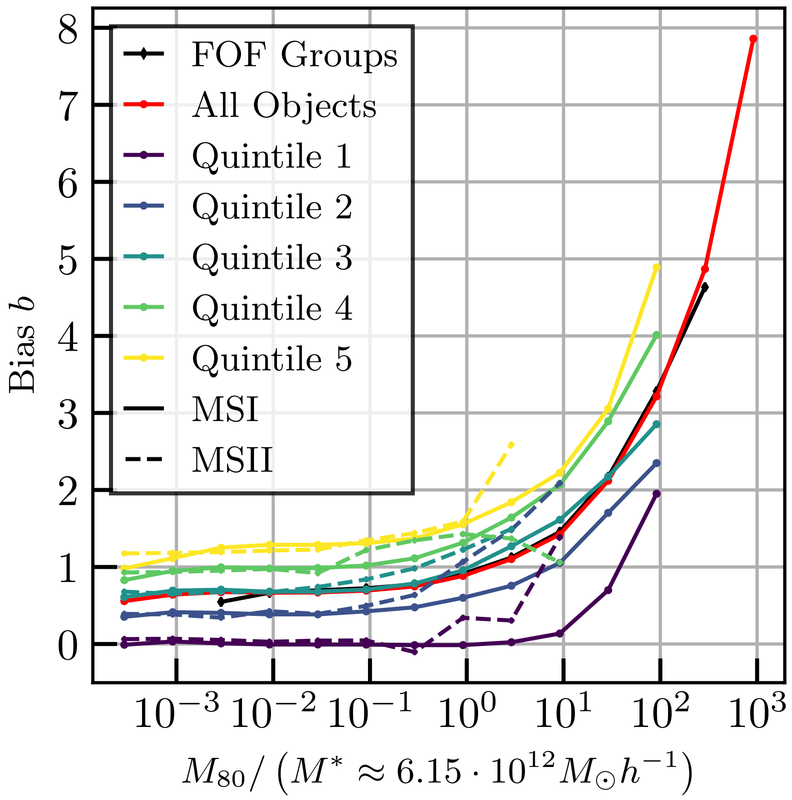

Figure 2.16: Bias for the quintiles in ρ_{lim} in mass-binned subsamples of thresholded haloes with $\rho_{thresh} = 80\,\langle\rho\rangle$ in the MSI and MSII. The red line gives the bias of the mass-binned subsamples not split by ρ_{lim} in the MSI and the black line that of the central subhalo in FOF groups binned by FOF group mass. Only bins with at least 100 members are shown.

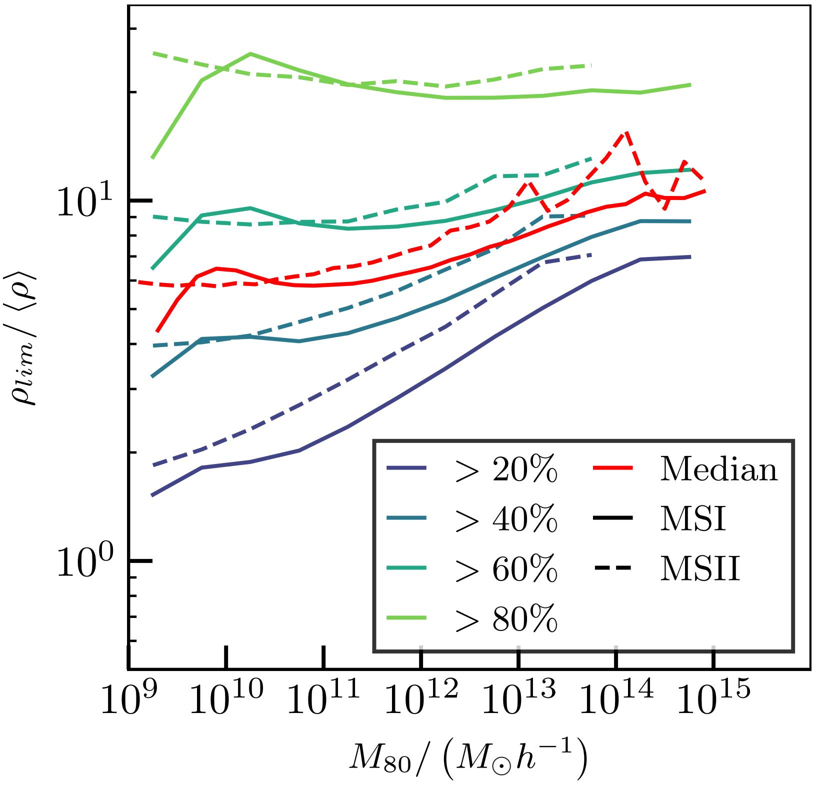

Figure 2.17: Boundary values between the quintiles in ρ_{lim} in evenly spaced logarithmic bins in M_{tot} for MSI and MSII.

2.7.2 Avoidance Between Top and Bottom Quantile

While we established that the top and bottom quintile cluster very differently with respect to matter, it is still unclear how this actually manifests. We try to understand their differences first visually using scatter plots of slices and then using radial cross-correlations.

2.7.2.1 Scatter Plots of Slices

For a visual inspection we show slices of the MSI in Figure 2.18 as inspired by Gao et al. (2005). Each slice is $30h^{-1}\,\mathrm{Mpc}$ thick and covers the whole box. All slices in each row have the same number of points, with the left (black) being a random set of particles as proxy for the matter distribution, and the middle and right panels showing the positions of objects with mass $8.61 \cdot 10^{10} \leq M_{80}/h^{-1}\,\mathrm{M}_\odot \leq 1.72 \cdot 10^{11}$ (100 to 200 particles) in the top (red) and bottom (blue) quantiles in ρ_{lim}, respectively. The upper row depicts quintiles and the lower row deciles.

For the top quantiles the visual impression is quite clear: they simply trace the matter distribution in a more concentrated fashion, hence are more clustered. They avoid underdensities, which especially in the case of the top decile opens up regions dozens of $h^{-1}\,\mathrm{Mpc}$ across without any object in the slice.

The bottom quantiles are more complex and we find clear differences between bottom quintile and decile. The objects in the bottom quintile (with $b \approx 0$) are not evenly distributed, but still exhibit some clear modulation in density. While in some regions one can see an avoidance of overdensities in other regions this is less the case. There is also no visible preference for locations in regions underdense in matter. In contrast, the bottom decile seems to avoid overdensities in matter. In any case the visual impression is much harder to interpret for the two bottom quantiles.

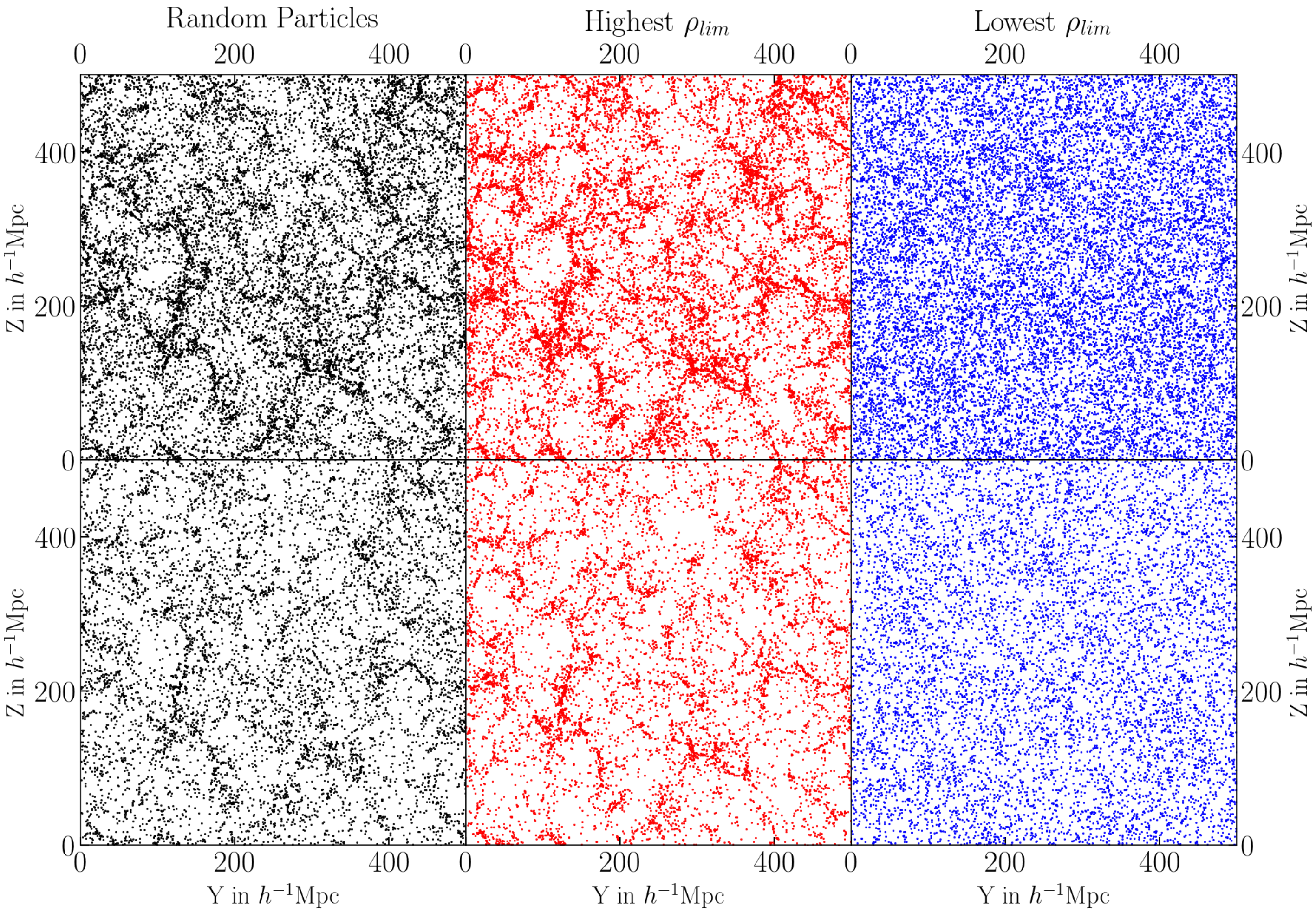

Figure 2.18: Visual impression of the relative clustering of the top and bottom quintiles (upper row) and deciles (lower row) in ρ_{lim} in the MSI in comparison with matter as represented by a random sample of particles with the same size as the quantiles. The plots show the positions in a projected slice of $30h^{-1}$ Mpc thickness along the X-axis. This plot is a reproduction of Figure 4 in Gao et al. (2005) with our data. We therefore chose the same mass range of 100 to 200 particles, corresponding to $8.61 \cdot 10^{10} \leq M_{80}/h^{-1}\,\mathrm{M}_{\odot} \leq 1.72 \cdot 10^{11}$. It is hard to see by eye whether the samples are avoiding each other or not. We therefore show their cross-correlations in Figure 2.19

2.7.2.2 Radial Cross-Correlations

To quantify and reassure ourselves of the results of the visual impression, we plot the radial cross-correlations ξ between the top and bottom quantiles in the same mass range as in the previous section with each other and matter in comparison to the matter auto-correlation in Figure 2.19. Here the grey band denotes the range $6 \leq R_p / \left(h^{-1}\,\mathrm{Mpc}\right) \leq 20$ of our bias b calculation.

As expected, in said range we find the top quintile and decile to lie above the matter auto-correlation and the bottom quintile around zero and the bottom decile slightly below. More interesting are the results for smaller radii than the lower end of the bias range.

The matter cross-correlations show that the objects with the lowest ρ_{lim} live in underdense regions, out to $\sim 5h^{-1}\,\mathrm{Mpc}$ for the lowest 20% and $\sim 10h^{-1}\,\mathrm{Mpc}$ for the lowest 10%. Their density profiles and as such the cross-correlation within a few hundred $h^{-1}\,\mathrm{kpc}$ do not differ.

As the bias of the top quantiles is above unity, their matter cross-correlations also rise steeply with the matter auto-correlation down to $2h^{-1}\,\mathrm{Mpc}$. At this point the cross-correlation stabilizes at 20 (top quintile) to 30 (top decile) before it rises at radii below $200h^{-1}\,\mathrm{kpc}$ due to the one halo term. Inside that radius the profiles agree with each other within the uncertainty so that no internal differences of the peaks can be seen here.

We also show the cross correlations between the top and bottom quantiles. Here we see a very peculiar behaviour. As was to be expected the two extremes in ρ_{lim} avoid each other on large scales. They do so even more than the bottom quintile avoids matter on large scale below a few $h^{-1}\,\mathrm{Mpc}$. However, below $600(500)h^{-1}\,\mathrm{kpc}$ the quintiles (deciles) start to be correlated and their correlation function steeply rises to a maximum in the range of $300h^{-1}\,\mathrm{kpc}$ to $400h^{-1}\,\mathrm{kpc}$ before it again seems to decline and we run out of pairs at $200h^{-1}\,\mathrm{kpc}$.

The avoidance between the samples would have been expected given their large scale clustering with, or avoidance of matter. Nevertheless, we must not forget the characteristic of the correlation function as a summary statistic. While not many of the pairs from opposite quantiles lie close to each other, those that do, do so with much higher frequency than random chance would dictate. The only conceivable configuration, given the opposite extremes in ρ_{lim}, is the one in which the peak with the high ρ_{lim} is grouped under that with low ρ_{lim}. This also explains the large radius at which the quintiles become correlated as here the density between the peaks can fall off more than for the deciles. So while it is indeed true that peaks of opposite extreme quantiles do avoid each other, from time to time they come in close groups. These pairs can also not lie too close to each other as in these cases the ρ_{lim} value would exceed the $\rho_{lim} < \rho_{thresh} = 80\,\langle\rho\rangle$ criterion.

2.7.3 Bias in M_{tot}-ρ_{lim}-Space

To understand how the selection of objects by ρ_{lim} selects on large scale environment we bin the $\rho_{thresh} = 80\,\langle\rho\rangle$ thresholded objects into bins of 0.25dex over wide range in ρ_{lim} and M_{tot}. We then calculate the bias as in Section 2.2.5 for every bin with more than 100

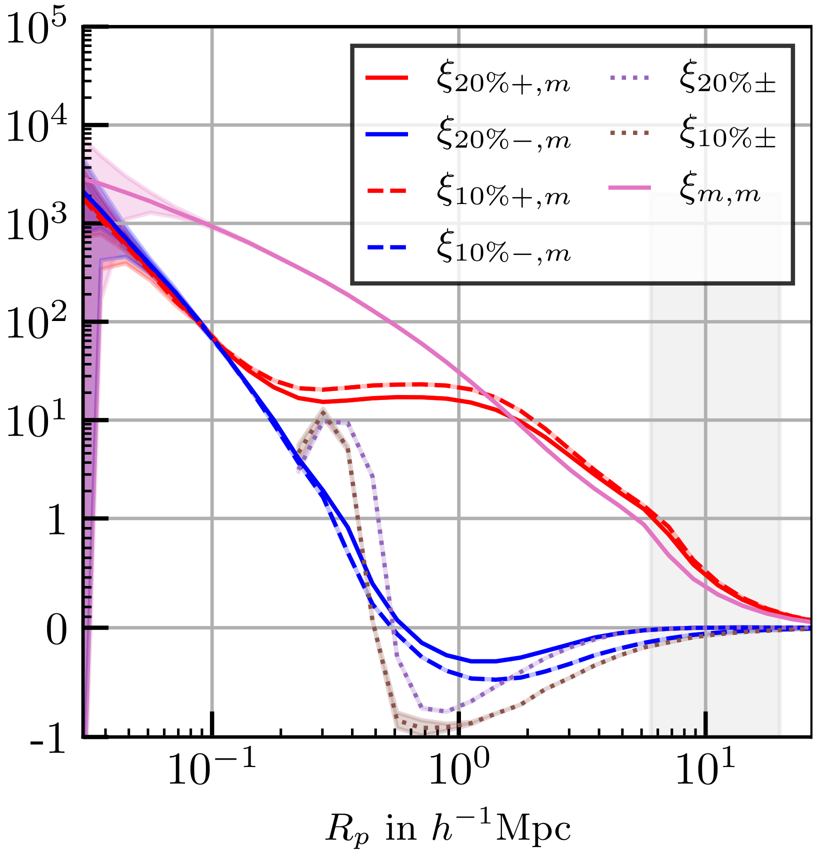

Figure 2.19: Cross correlations between the samples shown in Figure 2.18 out to $30h^{-1}$ Mpc (purple and brown dashed lines). Red and blue lines show the cross correlations between the samples and matter. Bands of colour around the lines depict the uncertainty. The range over which the bias is measured marked in light grey. Note the change from logarithmic to linear scale at 1, leading to the visible break in the slopes.

members. The results are shown in Figure 2.20 for the MSI and in Figure 2.21 for the MSII. As the former contains sufficient numbers of objects for a clustering analysis up to high masses we will concentrate on these results. We use the MSII mostly for the study of convergence for small masses as it is too small to have meaningful numbers of objects above $10^{13}h^{-1}\,\mathrm{M}_{\odot}$.

2.7.3.1 ρ_{lim}-Assembly Bias in the MSI

We find an unprecedented range in bias values $-1 \lesssim b \lesssim 3.5$ in the well covered range with more than 100 members per bin. In those with at least 10 members the range further broadens to $-2 \lesssim b \lesssim 5.5$. At fixed mass the bias value increases with increasing ρ_{lim} from negative values for all but the most massive objects with $M_{tot} \geq 10^{13.5}h^{-1}\,\mathrm{M}_{\odot}$ to positive values. The zero-crossing increases in ρ_{lim} in the form of a shallow power-law $\propto M_{tot}^{\frac{1}{6}}$. We also find a sudden turn in the $b = 1$ contour for small masses and high densities, the area previously shown to be strongly afflicted with spurious peak (c.f. Section 2.5.2.2).

At fixed ρ_{lim} the magnitude of b increases with M_{tot}. The sign of the large mass limit value is dependent on ρ_{lim}, especially on the relative position to the percolation threshold $\rho_{perc} \approx 6.3\,\langle\rho\rangle$ (see Section 2.4.2). Above this value objects are members of the cosmic web while below they only connect to it at densities where the percolating structure is already in place. For smaller values the largest objects at these limiting densities first become increasingly uncorrelated and then start to be anti-correlated with the matter distribution on large scales.

We can understand this distribution when we consider the constraints placed upon the peaks by the construction of the peak-hierarchy in connection with the transition from a local to a global peak-density comparison at ρ_{perc}. For a high mass object to be subordinated at densities above ρ_{perc} we need an even higher peak close by which is much more likely in dense large scale environments. Conversely, a high mass-object can only evade having a high-density channel to the cosmic web if it is in an underdense large scale environment. These restrictions are less strict for small mass objects, as their peak densities will be lower (see 2.6) and therefore the correlation between ρ_{lim} and large scale environment as measured by b is much less pronounced.

2.7.3.2 Connection to the Cosmic Web

Overall the bias contour distribution is fan-shaped with the contours travelling radially outwards from the high mass tip of the distribution in M_{tot}-ρ_{lim} space. This implies that there is a bias contour b_i which is parallel to the mass axis, or differently put, there exists a ρ_{lim} value for which the bias becomes independent of mass. We approximately identified these contours on our grid and added them to the regular contour sets in Figure 2.20 and Figure 2.21. For the MSI we find a bias value of $b_{i,MSI} \approx 0.7$ at $\rho_{lim,i,MSI} \approx 6.5$ and for the MSII $b_{i,MSII} \approx 0.8$ at $\rho_{lim,i,MSII} \approx 9$. These are the densities of percolation ρ_{perc} of the cosmic web in both simulations as discussed in Section 2.4.2 (c.f. Figure 2.6, Figure 2.7).

We do not fully understand the reason for the constant bias value at the cosmic web percolation density and will dedicate further research to this question.

2.7.3.3 Assembly Bias Convergence

To further understand the meaning and robustness of the results discussed in the previous section we have compare the sets of contours derived from MSI and MSII in Figure 2.22. Here we underlay the contours in black if they are inside the region with 100 members per bin. Solid contours come from MSI and dashed from MSII.

The direct comparison shows that just as the abundance of objects of a given mass does extend to lower ρ_{lim} values in the MSI, a given bias contour lies at higher ρ_{lim} in the MSII than in the MSI. Just as the abundance contours converge at higher ρ_{lim} above ρ_{perc} so do the bias contours. Despite the more erratic shape of the contours in the MSII compared to MSI, we do not find any systematic differences for higher ρ_{lim} or b values. Only the sudden upturn in of the $b = 1$ contour at $M_{tot} = 10^{9.5}h^{-1}\,\mathrm{M}_{\odot}$ is clearly only present in MSI. As mentioned in the previous section this is a sign of the less correlated spurious peaks that start to dominate in this region.

2.8 Conclusions

We presented a new method called "Tessellation-Level-Tree" (TLT) to decompose the density field estimated using a Voronoi tessellation of particles in a cosmological N-body simulation into a hierarchy of density peaks. We define a tree structure on the peaks and discuss a number of interesting quantities of its constituents. In addition to their mass and volume as inherited from their particle content these peaks have a novel property called limiting density ρ_{lim}. This is the density at which a peak has a saddle point to a denser peak. It therefore measures the maximum density in the immediate surroundings of the peak. To remove unwanted spurious peaks resulting from discreteness noise we apply a persistence filter of $r \geq 10$. While this does not solve the problem entirely it is a first step on the way to a cleaned peak population.

We additionally introduce the concept of thresholded peaks which are the peaks exceeding a certain density threshold ρ_{thresh} and whose properties are that of their mass content above the threshold.

As first analysis of the generated peak population we looked at the mass of the most massive object above a density threshold in MSI and MSII and found a percolation transition at $6 \lesssim \rho_{perc}/\langle\rho\rangle \lesssim 7$ in the MSI and $7 \lesssim \rho_{perc}/\langle\rho\rangle \lesssim 8$ in the MSII. In the MSII the percolation is less pronounced as the smaller volume leads to a mass function that is dominated by a few objects in the exponential cutoff. For both simulations the percolation transitions sets in at a volume fraction $V(\rho \geq \rho_{perc})/V_{tot} \approx 0.01$ and a mass fraction of $M(\rho \geq \rho_{perc})/M_{tot} \gtrsim 0.7$.

In Section 2.4.3 we found the percolating object to indeed be of filamentary nature. For a threshold density of $\rho_{thresh} = 5.25\,\langle\rho\rangle$ these filaments are usually separated by $20h^{-1}\,\mathrm{Mpc}$

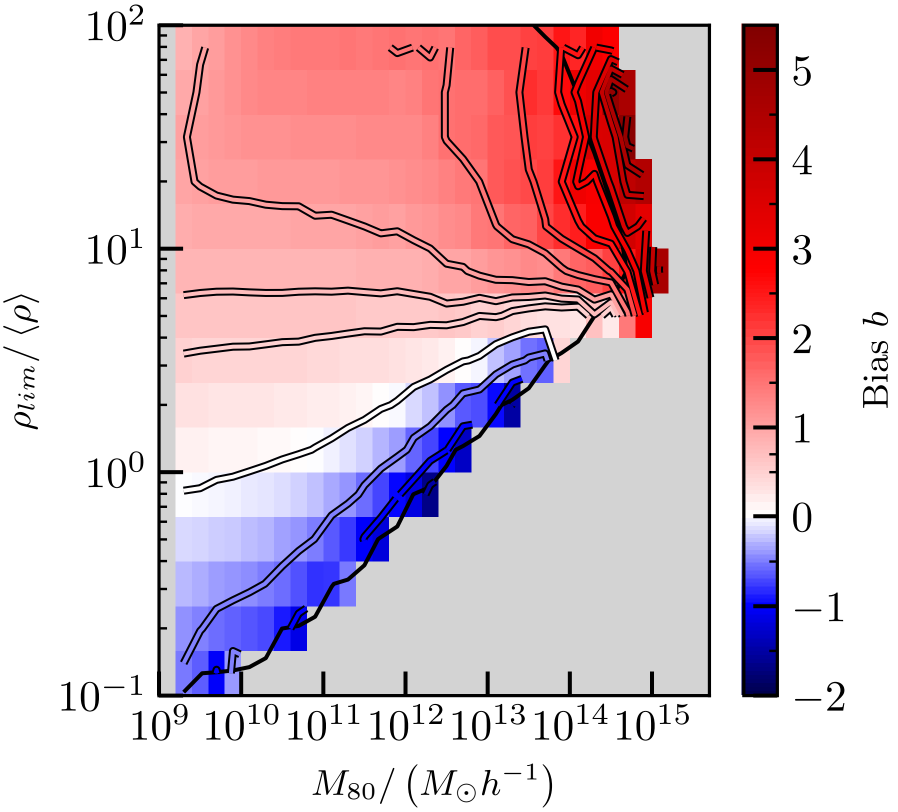

Figure 2.20: Bias distribution of thresholded haloes (with $\rho_{thresh} = 80\,\langle\rho\rangle$) in the MSI. Each cell in the colourmap represents a sub-sample with a given bias indicated by the color. The black outline is the boundary of the region where subsamples have at least 100 members, those with less than 100 but at least 10 lie outside it. The bias contours are evenly spaced with a step of 0.5 as indicated by the ticks in the colourbar. An extra contour at $b = 0.7$ was added as this is seems to be roughly the bias value at which bias becomes mass-independent. The corresponding ρ_{lim} level is very close to the percolation threshold of the cosmic web.

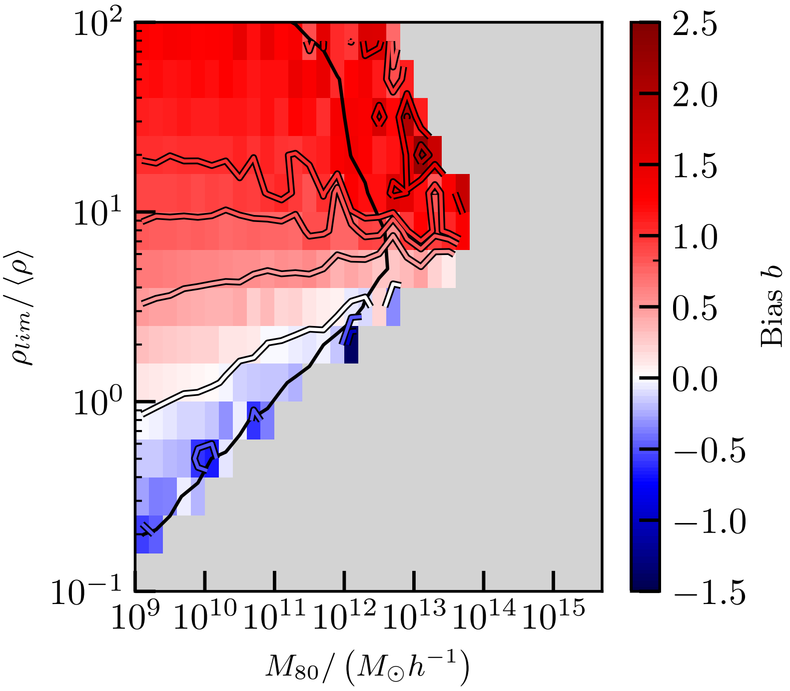

Figure 2.21: Same as Figure 2.20 but here for the MSII. Note the changed colour scale due to the reduced bias range. We again find a contour of constant bias at a ρ_{lim} value corresponding to the cosmic web percolation density. Here the constant bias value is slightly higher at $b \approx 0.8$.

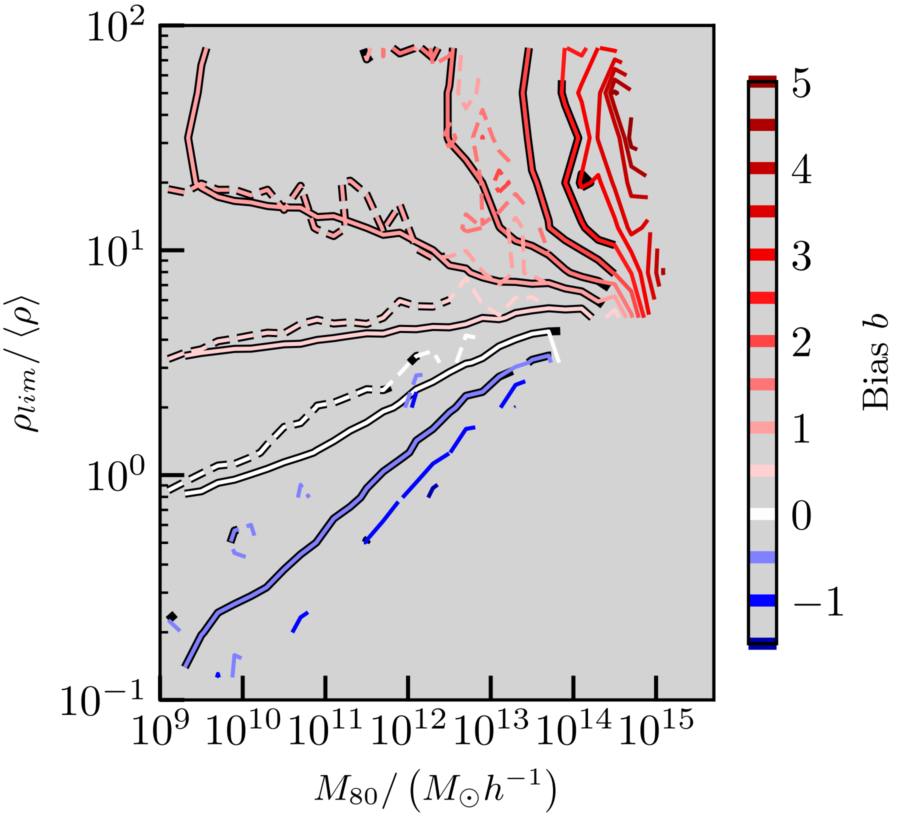

Figure 2.22: Comparison of the bias contours of MSI (solid) and MSII (dashed) as taken from Figure 2.20 and Figure 2.21. Black outlined contours indicate the range where they are calculated using subsamples of at least 100 members and otherwise subsamples of at least 10 members are used.

but the largest void we find in the MSI has a radius over $50h^{-1}$ Mpc, consistent with prior investigations. We also showed that these numbers are highly dependent on the chosen threshold. As the threshold is decreased, more branches of the web extend deeper into what was previously considered to be a void. Despite this, the point of percolation gives us a clear reference point relative to which one can evaluate these quantities. While on larger scales there are restrictions due to the limited box size, we can confirm very good convergence of the distance field on scales out to beyond $10h^{-1}$ Mpc between the MSI and MSII, albeit only when adjusting for the difference in onset of percolation between the two simulations.

We compared the abundances of thresholded objects in the Millennium and Millennium-II simulations with that of FOF haloes and found them to be similar but to have certain peculiar differences in line with previous discussions on FOF haloes. For these and peaks without a threshold we find that objects at large ρ_{lim} values and small masses show signs of being predominantly spurious.

We also introduced the usage of mass-density profiles as a new tool to investigate the structure of density peaks and haloes in N-body simulations. These profiles circumvent some problems of radial density profiles in describing the density structure of haloes by directly connecting mass and density. By avoiding the spherical or spheroidal averages needed for the usual radius-density profiles we are able to fully capture the high densities in sub-haloes.

Using the new limiting density property we investigate the bias for samples of constant mass split by ρ_{lim} to investigate the assembly bias effect in this quantity for objects that are close to the typical FOF halo definition. We find the strongest assembly bias effect to date as the bottom quintile in ρ_{lim} at a given mass is essentially uncorrelated with mass below masses of $M_{halo} = 10^{13.5}h^{-1}\,\mathrm{M}_{\odot}$. We find that halo with very different ρ_{lim} tend to avoid each other and that a low ρ_{lim} locally indicates an avoidance of matter on large scale. We also show for the first times samples of group-sized dark matter haloes that avoid matter on large scales, leading to a negative bias. The fact that we can identify a range of ρ_{lim} values at which the bias becomes constant with mass and that this happens to occur around the point of percolation indicates interesting connections between the two phenomena which warrants further investigation.

Future work around the TLT will focus on the three applications in this chapter, the bias of the peaks, their internal density distribution and the percolating isodensity surface. Regarding the limiting density assembly bias and density structure of peaks it will be interesting to see how these imprint themselves on the properties of galaxies that reside in these peaks and how this can help us to further understand the connection between galaxy and structure formation. The percolating isodensity surface introduces a network that can be used to locate features in the cosmic density distribution. We hope to achieve this and also further characterise this network in a future publication.

2.A Connection to the NFW-Concentration

Another measure of concentration is the well known concentration parameter of the NFW-profile. The bijective nature of the profile function allows us to invert it and obtain the function $r(\rho, c, R_V)$. For the following we are interested in a formulation of this radius which does not depend on the choice of δ_V or the physical scale of the halo. Due to the nature of the profile we only need the ratio of the density ρ and the reference overdensity

$$\beta = \frac{\rho}{\delta_V} \tag{2.18}$$

which gives us

$$\begin{aligned} r\left(\beta, c, R_V\right) &= \frac{R_V}{3c}\left(2^{\frac{1}{3}} t(\beta) + 2^{-\frac{1}{3}} t(\beta)^{-1} - 2\right) \\ &= \frac{R_V}{c} f(\beta), \end{aligned} \tag{2.19}$$

with

$$t(\beta) = \sqrt[3]{\frac{\beta}{2\beta + 3\left(\sqrt{12\beta + 81} + 9\right)}}. \tag{2.20}$$

Using this radius expression we can find the mass inside a bounding density $\rho_b = \beta\delta_V$:

$$\begin{aligned} M(\rho_b, c, M_V) &= 4\pi \int\limits_0^{r(\beta,c,R_V)} \rho_{NFW}(r) r^2 \mathrm{d}r \\ &= \frac{4\pi}{3} R_V^3 \delta_V \frac{1+c}{(1+c)\ln(1+c) - c} \\ &\quad \left[\ln\left(\frac{R_S + R_V c^{-1} f(\beta)}{R_S}\right) - \frac{R_S}{R_S + R_V c^{-1} f(\beta)}\right] \\ &= M_V \frac{1+c}{(1+c)\ln(1+c) - c} \\ &\quad \left[\ln\left(1 + f(\beta)\right) - \frac{1}{1 + f(\beta)}\right] \\ &= M_V g(\beta, c), \end{aligned} \tag{2.21}$$

where we used the relation $c = R_S/R_V$ for the NFW concentration parameter and

$$M_V = \frac{4\pi}{3} R_V^3 \delta_V \tag{2.22}$$

for the characteristic mass M_V. The system of reference is set by a choice of δ_V, frequent choices are δ_{200c} and δ_{200m}, 200 times the critical and mean density, respectively.

To be able to translate the values found by the mass ratios to concentrations in the NFW framework we only need to take the ratio of the masses for two bounding densities with a given η:

$$c_a = \frac{M(\eta\rho, c, M_V)}{M(\rho, c, M_V)} = \frac{g(\eta\beta\delta_V, c)}{g(\beta\delta_V, c)} \tag{2.23}$$

Using the relation (2.23) we can now connect a given mass ratio c_a at a bounding density ratio η as discussed in 2.2.4.4 to a NFW-concentration. This is shown in Figure 2.23 for different values of η. We find that different choices of η give different locations of the region of steepest ascent for c_a. Depending on the objects of interest certain η values should give more precise connections. In the case of unthresholded haloes we want to compare their concentrations defined in relation to a common overdensity, despite them existing over very different mass ranges. By fitting such an NFW-profile in density/mass-space, we can find an equivalent concentration value for each object at a given density.

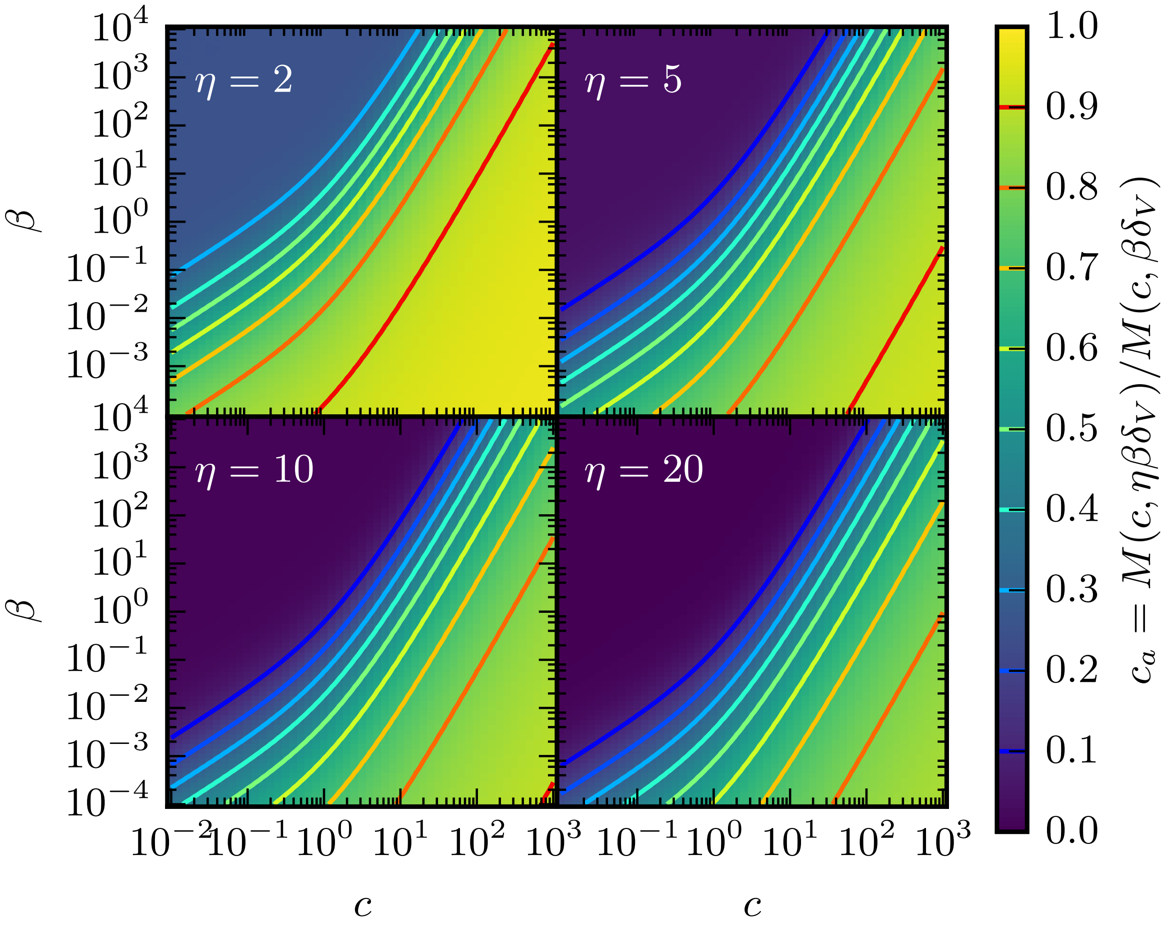

Figure 2.23: The relation between the NFW-concentration c and the mass ratio bounded by two isodensity surfaces with a density ratio η.

Chapter 3

Morphology of Reionisation

This work will be published in the upcoming paper Busch et al. (2019).

3.1 Introduction

As mentioned in Section 1.2 it is as of yet unclear to what degree different types of sources, early stars, stellar remnants, or even quasars (Madau and Haardt, 2015), are responsible for the emission of ionising radiation. One therefore tests the predictions of various models of sources and cosmology and compares different observables from mock and real observations as they become available from ever improving instruments.

As the process of reionisation is highly non-linear and inhomogeneous, it is usually investigated with semi-analytic (Mesinger et al., 2011; Xu et al., 2017; Greig and Mesinger, 2018) or fully numerical simulations (Ciardi et al., 2003, 2012; Kakiichi et al., 2017a; Eide et al., 2018a), sometimes even coupled to hydrodynamics (Gnedin, 2014; Ocvirk et al., 2016; Pawlik et al., 2017; Rosdahl et al., 2018). While semi-analytic methods have to compromise on accuracy, they are fast enough for parameter space explorations that are very expensive using full radiative transfer simulations and prohibitive in the case of coupled simulations.

The characterisation of the morphology of reionisation has been developed along with the increasing capabilities to simulate the process. Sometimes this morphology traced the 21cm signal (Wang et al., 2015) and sometimes ionisation fractions directly (Friedrich et al., 2011). An early method that still sees considerable use today is based upon Minkowski functionals (Gleser et al., 2006; Kapahtia et al., 2018; Chen et al., 2018; Bag et al., 2018). While these are mathematically very well developed, they have the disadvantage of being integral quantities of a given surface and are therefore inherently non-local. While this does not pose a problem in the case of isolated, monolithic bubbles, it does very much so once percolation sets in. There are a number of previous approaches to separate connected ionised regions into separate bubbles to carry the notion of bubbles over to the overlap phase (Lin et al., 2016a).

A problem from an observational point of view is the presence of noise that can hamper morphology detection (Kakiichi et al., 2017b). One possible way is smoothing the image another the introduction of super-pixels (Giri et al., 2018). We neglect a special treatment of this problem for now, but prefer super-pixels for a real mock observation application.

In the following we will introduce our new methodology for describing the morphology of ionised regions, both locally and globally, which extends granulometry as introduced by Kakiichi et al. (2017b) in Section 3.3. We then quickly give a short overview of the background of the different simulations first presented in (Eide et al., 2018a) in Section 3.2. The results of our analysis are split in three sections: global properties of the bubble population (Section 3.4), the percolation transition (Section 3.5) and finally the connection to the density fields in Section 3.6.

3.2 The Simulations

The simulations of reionisation used for this investigations are the product of a radiative transfer post-processing first presented in Eide et al. (2018a, hereafter E18) of the MassiveBlack-II cosmological hydrodynamics-simulation (Khandai et al., 2015, MBII) using the Monte-Carlo ray tracing code `CRASH` (Ciardi et al., 2002; Maselli et al., 2003; Graziani et al., 2013)

3.2.1 Cosmological Simulation

The MBII simulates a $(100h^{-1}\,\mathrm{cMpc})^3$ periodic volume with a WMAP7 cosmology and samples the matter and gas distribution with 1792^3 particles each. The gas and dark matter particles have a mass of $m_{\mathrm{gas}} = 2.2 \cdot 10^6 h^{-1}\,\mathrm{M}_\odot$ and $m_{\mathrm{DM}} = 1.1 \cdot 10^7 h^{-1}\,\mathrm{M}_\odot$ while the Plummer-equivalent softening length is $\epsilon = 1.85h^{-1}\,\mathrm{ckpc}$. The simulation was performed using the unpublished code P-GADGET based upon GADGET2 (Springel, 2005). This code uses smoothed particle hydrodynamics to follow the gas dynamics and implements a number of feedback mechanisms.

The halo identification uses the standard SUBFIND algorithm by Springel et al. (2001). While in the original analysis of the MBII the authors defined a galaxy as a bound group of at least 64 stellar particles within a subhalo, E18a dropped this requirement and instead model each stellar particle individually as described in Section 3.2.2.

For the results in this work we exclusively use grids of $N_C = 256^3$ cells and therefore a comoving side length of $l_C = 0.391h^{-1}\,\mathrm{cMpc}$. For this the gas and star particles of the MBII are deposited on the grid. Cells with at least one stellar particle represent sources from which photon packets will be emitted. As the number of relatively very faint sources increases rapidly beyond $z = 10$ we use a spatial clustering approach to group sources in a flux conserving manner. This is described in detail in (Eide et al., 2018b, hereafter E19). While the parent-simulation has periodic boundary conditions our simulations do not due to technical difficulties connected to X-ray radiative transfer as the large number of crossings of X-ray photons exceeds runtime requirements.

Table 3.1: The source scenarios and their labels.

Lable	Source types
Stars	Stars only
SBH	Stars and super-massive black holes
SXRB	Stars and X-ray binaries
SISM	Stars and thermal interstellar medium emission
SXBI	All sources

3.2.2 Source Scenarios

The five different source scenarios as listed in Table 3.1 were first presented in E18a. They are based on stellar, X-ray binary (XRB), hot interstellar medium (ISM) and nuclear black hole emission (BH), and combinations of these. We use this opportunity to quickly remind the reader of their key aspects and refer to E18a for further details. In the analysis in Section 3.4 and onward we will denote the scenarios with the labels as given in Table 3.1. For a visual comparison of the average spectral energy density (SED) contributions of the different source types we refer the reader to Figure 2 in E18a.

The basic source of ionising photons in all simulations are stars. Each stellar particle from the parent simulation is emitting a single stellar population SED without binary stars according to its age and metallicity. All these SEDs are averaged over the whole simulation at a given snapshot to obtain a mean SED. This mean SED is then applied to all source cells and rescaled by the luminosity in a given cell. This procedure is also followed for the other emission of essentially stellar origin (XRBs, ISM). The stellar emission is the dominating component in the global mean galactic SED at energies up to the ionisation energy of HeII (54.4 eV).

The hot ISM in galaxies is emitting thermal X-rays via bremsstrahlung. This is modelled assuming a 10^6K ISM and a constant SED up to 240 eV and a cubic power law decline thereafter. This shape and the model luminosity make the ISM the dominant radiation source in the range from the HeII ionisation energy at 54.4 eV up to 600 eV.

The luminosity of the XRB population is modelled via a metallicity and star formation rate dependent high-mass XRB sub-population and a time and mass dependent low-mass XRB sub-population. Their SED dominates the galactic emission above 600 eV.

The given ranges of dominance of the different source types are based upon their relative contributions in the case of a combination of all types. As the stellar emission plummets rapidly at energies above 54.4 eV the more energetic sources are always dominant above 100 eV, the ISM immediately above HeII ionisation.

As SMBHs are rare, they will not be found in all source cells and one cannot find a combined mean SED that could be redistributed to all sources as for the other source types. Therefore their emission is treated separately. While the mean SED of the SMBHs is formally dominant over the whole energy range used in theses simulations, they are globally sub-dominant due to their mentioned sparsity. Nevertheless they are locally dominating the radiation field, especially at energies beyond the ionisation threshold of HeII (54.4 eV).

3.3 Methodology

The aim of this investigation is the characterisation of the morphology of ionised bubbles in the five presented source scenarios. For this we first need a definition of a bubble and furthermore a framework to describe its shape in a quantitative manner. We express the bubbles as objects in sets of binary fields (Section 3.3.1) derived from the ionisation fields, and describe their outer and inner structure with the opening field (Section 3.3.2) and the euclidean distance transform (Section 3.3.3). We also present a tree structure to describe the morphology of single bubbles, and use this to describe the growth and merging process of the bubbles as they progress to fill the volume.

3.3.1 Binary Fields

At the core of all tools used in the following to characterise the ionised bubbles and other fields in the EoR lies the *binary field* (BF). The BF is the result of a classification of volume as either filled (1) or empty (0).

We obtain these classifications by testing a binary choice on another field. In the following this will be mostly the comparison of the fraction $x_{\mathrm{I}}(\boldsymbol{r})$ of the ion species I, a field of real values, with a threshold t_{I}:

$$X_{\mathrm{I}}(\boldsymbol{r}) = \begin{cases} 1 & \text{if } x_{\mathrm{I}}(\boldsymbol{r}) \geq t_{\mathrm{I}}, \\ 0 & \text{if } x_{\mathrm{I}}(\boldsymbol{r}) < t_{\mathrm{I}}. \end{cases} \tag{3.1}$$

The choice of filled and empty is decided such that the ionised volume around the sources is filled and its complement is empty. In the following, for the sake of clarity, we will refer to filled/empty cells if they lie above or below the ionisation threshold. It is intuitively clear that for a given field X_{I} different choices of threshold can result in very different binary fields.

The discretisation of a continuous field on only two values comes with a dramatic loss of information in most instances. In cases where the field is almost binary to begin with, such as the ionisation fractions in many circumstances, this merely removes ambiguity. For fields with a large range of values one would need to take sufficiently close thresholds sampling the whole range of values to grasp the possibly changing morphology in different regimes.

3.3.2 Opening Field

Now that we established and motivated the choice of a binary field as a useful abstraction we need a way to morphologically classify this field. As regions in the binary field are featureless within their boundaries the only difference between regions is their size and the shape of their surfaces. A simple measure for the size of a region is its diameter. Unfortunately the value of the diameter will heavily depend on the choice of the axis along which this diameter is measured. To solve this problem we use a set of overlapping

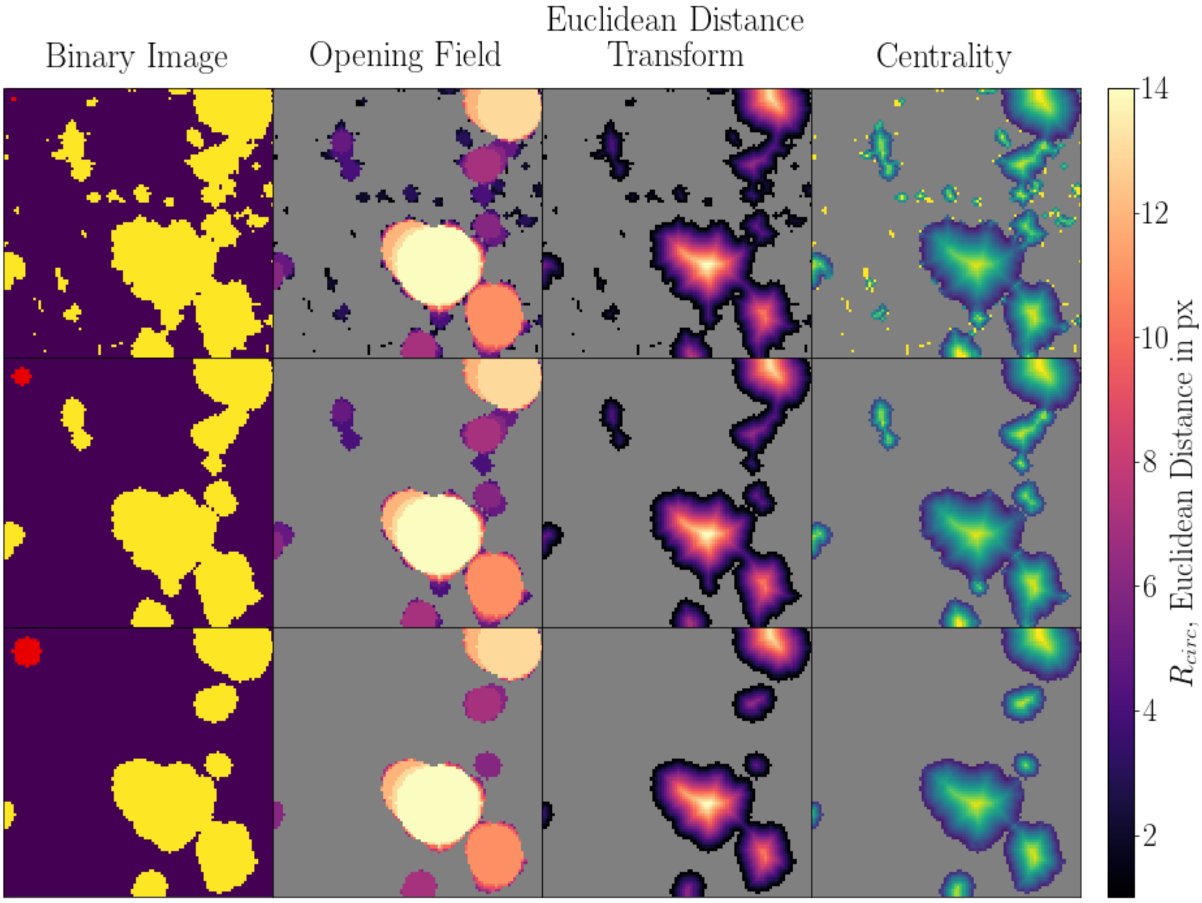

Figure 3.1: Two-dimensional toy examples of an opened binary image (first column) together with the connected opening field (second column), Euclidean distance transform (third column) and centrality (fourth column) for filtrations with opening radii of 0.5, 3.5 and 5.5 cells (rows from top to bottom). The red region in the upper left corner in the panels of the leftmost column depicts these structuring elements. In the three right columns grey covers the undefined background regions. The colourbar gives both the maximum opening radius for the opening field and the Euclidean distance. The centrality values vary linearly from 1 (yellow) in the most central regions to 0 (blue) at the very edge.

spheres of varying diameter to simultaneously measure diameters in all directions via the morphological opening of the binary regions as described in Section 3.3.2.1. To speed up this costly process we employ a technique based on fast Fourier transforms laid out in Section 3.A before connecting this approach to the previously used technique of granulometry in Section 3.3.2.2.

3.3.2.1 Opening Classification

The morphological opening can be intuitively understood as the filling of a volume $V \subseteq \mathbb{X}^d$ with overlapping replications of a certain shape, the structuring element $S \subseteq \mathbb{X}^d$. While the replications can overlap with each other they must never overlap the boundary. Here $\mathbb{X}$ are usually either the reals $\mathbb{R}$ or, as is the case for our simulation grids, the integers $\mathbb{Z}$.

More formally, the morphological opening $V \circ S$ is the result of a successive morphological erosion and dilation of V with S. The morphological erosion (written as $\ominus$) is the set of all possible centres $\boldsymbol{x} \in \mathbb{X}^d$ of translated replications of S, here given as $S_{\boldsymbol{x}} = \{\boldsymbol{s}+\boldsymbol{x} | \boldsymbol{s} \in S\}$, that are fully contained within V:

$$V \ominus S = \{\boldsymbol{x} \in \mathbb{X}^d | S_{\boldsymbol{x}} \subseteq V\}. \tag{3.2}$$

For the present application this translates into a removal of a layer of width r around every surface pixel when we erode a given BF with a sphere of this radius.

The morphological dilation (written as $\oplus$) is closely related to the erosion (see also Section 3.A) and is the union of all $S_{\boldsymbol{x}}$ that are translated within V:

$$V \oplus S = \bigcup_{\boldsymbol{x} \in V} S_{\boldsymbol{x}}. \tag{3.3}$$

For symmetric S as used in our methodology, concatenation of the two operations leads to a filtered version V' of V that is the union of all translations of S that are fully contained within V:

$$V' = (V \ominus S) \oplus S = \bigcup \{S_{\boldsymbol{x}} | S_{\boldsymbol{x}} \subseteq V\}. \tag{3.4}$$

For an alternative description based on Minkowski addition and subtraction see Kakiichi et al. (2017b).

We will exclusively use spheres (or rather their cell-approximations) of increasing radius R_{sph} as structuring elements to filter the binary field. These spheres are members of a set of binary fields defined by the inequality

$$||\boldsymbol{x}|| \leq r, \text{ for } r \in \mathbb{N}_0, \tag{3.5}$$

where $\boldsymbol{x}$ is the position vector to the centre of a cell. This implies that the smallest sphere is a single pixel around the origin, the second one is an n-dimensional cross and so forth. While the discretisation errors are substantial for small r they decrease quickly as r increases. To convert r-values into proper spherical radii we use the total volume in a cell-sphere and calculate the radius of equivalent volume R_{sph}.

Our opening process leaves us with regions that could be filled with overlapping spheres of a given radius. Due to the choice of spheres as structuring elements we know that regions obtained in that manner have at least the same diameter in any given direction. This also explains our choice of spheres, as any other shape would have some kind of preferred direction. The opening of a binary image in two dimensions is exemplified in the first column of Figure 3.1. The top panel shows the original image (or equivalently opened with a single pixel disc) while the lower two images have been morphologically opened with subsequently larger pixelated discs.

While in Kakiichi et al. (2017b) the result of the opening operations was only saved in order to track the global volume change of the filled regions, here we are also interested in recording where exactly volume is lost in the opening hierarchy. We call the successive openings a hierarchy, as the openings with larger R_{sph} are always contained within those of smaller R_{sph}. We can exploit this fact and save only the largest opening radius at which a given cell is retained after performing the opening operation with a sphere of this radius. We name the result of this operation the opening field (OF). An example of this is given in the second column of Figure 3.1, where the colour signifies the OF value in a given pixel. As the opening radius increases from top to bottom we see how regions with smaller OF values get removed while those with larger values stay unchanged.

In the simplest case of isolated spheres of varying dimensions, the OF just contains these exact same spheres labelled with their diameter. In a more realistic scenario there will be regions within a given bubble that are removed at earlier stages of the opening hierarchy than others. Locating and measuring these areas helps to understand the ionisation process on the boundaries of ionised regions and the merging of bubbles.

While we concentrate on the ionised component of the universe it is also of interest to measure the sizes of its complement. We can do this in an exactly symmetric way and simply apply the above methodology to the negation of the initial binary field. We can also save this in the same opening field just using a negative sign for the opening field of the complement.

3.3.2.2 Granulometry

Granulometry measures the global volume distribution in regions of a binary field with a given diameter. For this it uses successive openings with spheres as described in the previous section, and records the volume loss after each step. Kakiichi et al. (2017b) introduced this technique to the field of EoR studies in their investigation of the expected 21-cm spot size.

The central quantity in granulometry is the *pattern spectrum*

$$F(< R) = 1 - \frac{V(B \circ S_R)}{V(B)}, \tag{3.6}$$

where B is the binary image, V is the volume function and S_R the sphere of radius R. In the pattern spectrum we record the fraction of the volume that lies in regions of radius smaller than the opening radius and which are therefore lost during the opening procedure.

$F(< R)$ is thus the cumulative probability distribution of residing in a region of radius smaller than R. Therefore the volume weighted probability distribution of being in a region with opening radius R and local diameter $2R$ is then

$$p_V(R) = \frac{\mathrm{d}F}{\mathrm{d}R}. \tag{3.7}$$

We present our granulometric results in 3.4.2.

3.3.3 Euclidean Distance Transform

In addition to the opening field from the previous section we also use the Euclidean distance transform (EDT) to characterize our binary field. This transform of a BF assigns every filled cell (as defined in eq. 3.1) the minimal Euclidean distance d to an empty cell. In the context discussed here, the EDT thus measures the distance from the first layer of cells that are outside (hence we will also refer to it as surface distance) of the ionised regions as we have defined them. The result of this operation for our example images can be seen in the third column of Figure 3.1.

As (filled) spheres in Euclidean space are sets of points with an upper limit on the euclidean distance from their centre, the EDT can also be seen as the distribution of *centres* of maximally large spheres contained fully within the filled volume. This highlights the tight connection to the opening field with spherical structuring elements, to which it forms a kind of dual field. The same is true for our pixelated spheres.

In fact if both fields are needed for a given binary image B it is usually faster to compute the EDT E and then perform a morphological dilation with spheres of a given radius after applying that radius as a threshold on E. In the special case of spherical structure elements S the opening field O can then be found as

$$O = \sum_j B \circ S_j = \sum_j |E > R_j| \oplus S_j, \tag{3.8}$$

where S_j is the structure element of radius R_j and $|E > R_j|$ is the binary field of regions in E that have a value higher than R_j.

The EDT field has many uses, but here its relation to the OF gives us a measure of centrality in highly irregular objects as discussed in Section 3.3.4. We will also see at which value this field percolates at different times (see Section 3.5.2).

Just as for the opening field before we also calculate the EDT of the negation of the binary image to characterise the space between the ionised regions. Just as for the OF we can save this additional EDT with a negative sign in the same field as the original positive one.

3.3.4 Centrality

A very helpful derived quantity using both the OF and the EDT is *centrality* $C(\boldsymbol{x})$, which is defined as the ratio of EDT $E(\boldsymbol{x})$ and OF $O(\boldsymbol{x})$:

$$C(\boldsymbol{x}) = \frac{E(\boldsymbol{x})}{O(\boldsymbol{x})}. \tag{3.9}$$

A two-dimensional example can be found in the rightmost column of Figure 3.1.

As the OF measures the local diameter of the bubble and the EDT the separation from the surface, this ratio gives a dimensionless measure of how locally central a point is, i.e. how much of the locally maximally attainable separation from the surface is realised in this point. This measure helps us to localise features in a dimensionless manner in bubbles and to further separate overlapping bubbles. Regions with $C = 1$ will be identified with the centres of bubbles while on the surface C vanishes.

3.3.5 The Minimal Bubble Structure (MBS)

To represent the binary field and its accompanying opening field only a few of the structure element replications are covering cells that are not contained at a larger opening scale. These form a minimal structure that resembles a skeletonisation as their centres lie on the local medial axis (but which is not fully traced due to finite resolution). As an example we show an this structure in Figure 3.3 for the binary field in Figure 3.1.

This structure extends our description from knowing the region in which a given volume element is located to the centres of the regions. This way we expose the underlying structure that generates the shape of the given bubble. For this structure naturally only these spherical centres are important that are imprinted on the surface shape.

We use the MBS to find the correlation between the bubble centres and the density field in Section 3.6.

To construct the MBS we find all cells of opening level O that lie closest to a cell at position $\boldsymbol{x}$ for which $\lfloor E(\boldsymbol{x}) \rfloor = O$, where $\lfloor \cdot \rfloor$ is the function. If the number $N_c(\boldsymbol{x})$ of these cells is at least one we know that this is the centre of an essential structure element, i.e. an element that is the largest possible one describing a region within the bubble.

We furthermore calculate a weighting factor w_c of

$$w_c(\boldsymbol{x}) = \frac{N_c(\boldsymbol{x})}{V_{sph}(\lfloor E(\boldsymbol{x}) \rfloor)}. \tag{3.10}$$

Here V_{sph} is the volume of the spherical opening element of the appropriate radius. Cells that do not host a centre of an essential bubble have a weight of 0 while all centres will have a weight in the range $0 < w_c \leq 1$.

This algorithm is illustrated in Figure 3.2: The large bubble A has two sub-bubbles of same size, C and B. The centre of C receives a much smaller weight than the centre of B as C lies mostly within the larger bubble A. B has a sub-bubble itself, here denoted D, but as the larger bubble receives all the volume covered by B and D. The single bubble F is of

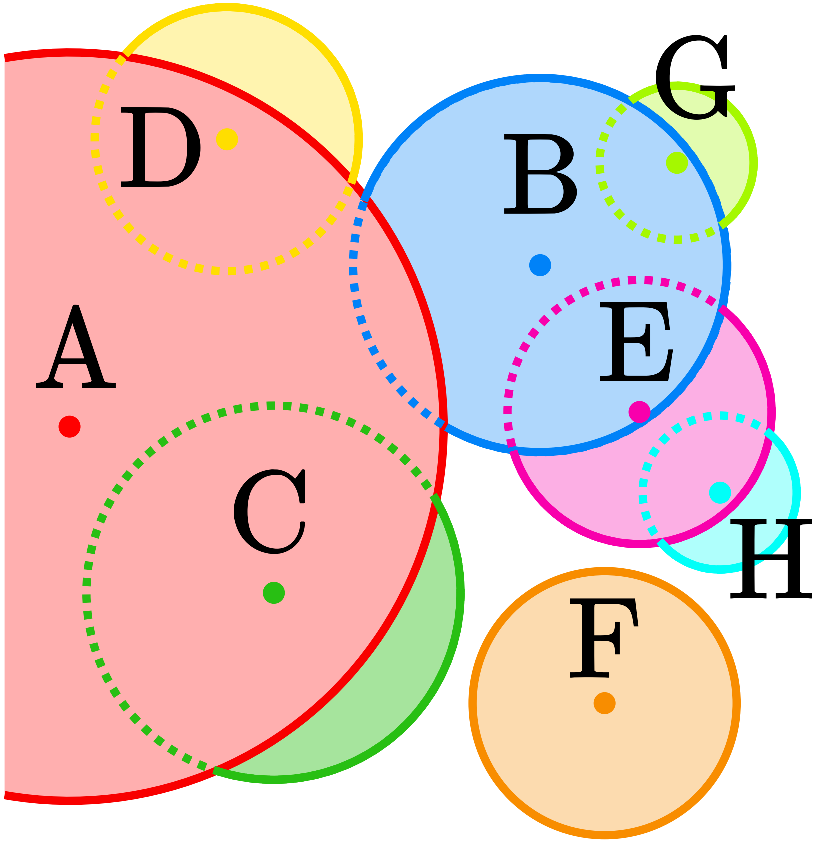

Figure 3.2: Schematic of the basis of the weighting applied to the minimal bubble structure. The weight of a bubble centre (represented by the dots) is calculated as the ratio of the volume of a sphere of the given radius (solid and dashed circles) to the part where it is the centre of the largest bubble in a given point (shaded region). The shaded areas here represent regions of a given opening radius.

the same size as D but as a single bubble without overlap with a larger bubble receives the same unit weight as A.

The above procedure has been implemented in two ways. Either one uses a combination of dilation and erosion operations on thresholded versions of the opening and EDT fields. This can become quite costly, so we will mainly use a second alternative in which we construct a kd-tree on the (necessarily smaller) set of candidate centres (i.e where $\lfloor E(\boldsymbol{x}) \rfloor = O$). We then query the kd-tree for the closest of these position for each cell of opening level O. This leaves us with a complete distribution of all cells with O on centres from the candidate set. In the case that two candidate centres are the same distance from a cell the tie will be broken in a random fashion. We then simply count the occurrences of the centre at $\boldsymbol{x}$ to determine the number $N_c(\boldsymbol{x})$ and the corresponding $w_c(\boldsymbol{x})$.

After repeating the above process for all possible values of O we are left with a field of weights w_c that is only non-zero in the locations of structure element centres necessary for a reconstruction of the shape of the bubble.

As an example of the MBS we give the result of applying the above technique to the example from Figure 3.1 and show the radii and weights of the found centres in the top and middle panel of Figure 3.3, respectively. The distribution of centres and radii reveals a network not unlike a skeleton obtained from a medial axis transform. We find a very large number of centres, most of which have rather small weights. Due to the discreteness of the grid all one-cell structures have a weight of one and we find many centres of small structures with only a few cells radius that have high weights. This is caused by their ability to resolve small scale surface structures that cannot be covered by larger spheres.

To illustrate that the distribution of the important structure centres is rather simple we show the distribution of radii of centres which have at least a weight of $w_c \geq 0.1$ in the bottom panel of Figure 3.3. There is a dramatic reduction in the number of centres. Small bubbles still show the signs of discreteness effects while the centres of large bubbles are often represented by clusters of central pixels with the same radius. This is the result of a combination of slightly elongated shapes of the regions. All these central pixels lie closest to a certain sector of the spheres centred on them and obtain their weight from this sector. Later on we will use these centres to cross-correlate bubbles with other fields (such as density, see Section 3.6) where the weights will represent the uncertainty in establishing a bubble centre.

3.4 Opening Analysis of Ionisation Fields

To learn more about the morphology of the ionisation fields during reionisation we produce binary images using a threshold $0.99 \leq x_i$ at redshifts $z \in \{10, 9, 8, 7.5, 7, 6.5\}$ and analyse them with the techniques as described in Section 3.3. All of these analyses are in one way or the other connected with the opening field, therefore we will summarise them under the term *opening analysis*.

The analysis of the bubble diameters for helium is more complex than for hydrogen as the appearance of a non-negligible HeIII fraction in the centre of ionised regions leads to central holes in the HeII distribution. One would therefore use the combined HeII and HeIII fraction $x_{\mathrm{He(II+III)}} = 1 - x_{\mathrm{HeI}}$ to find the morphology. Incidentally this is virtually the same as the morphology for HII for the chosen threshold in all our scenarios at all redshifts. In the following we therefore give neither the results for HeII alone nor for $\mathrm{He(II+III)}$.

3.4.1 Total Ionised Volume

To understand the following results we first need to know how much volume is ionised according to our classification. We give the share of the total volume that passes the threshold $x_i > 0.99$ in Figure 3.4.

The HeII volume closely follows the HII one and at all times the cells where $x_{\mathrm{He(II+III)}}$ (or $1 - x_{\mathrm{HeI}}$) lies above the threshold outnumber those where this is true for x_{HeII}. This means there is a non-negligible number of cells with sufficient x_{HeIII} at all redshifts. In the regime interesting for percolation studies, i.e. $0.1 \lesssim V(x_i > 0.99)/V_{tot} \lesssim 0.8$, the

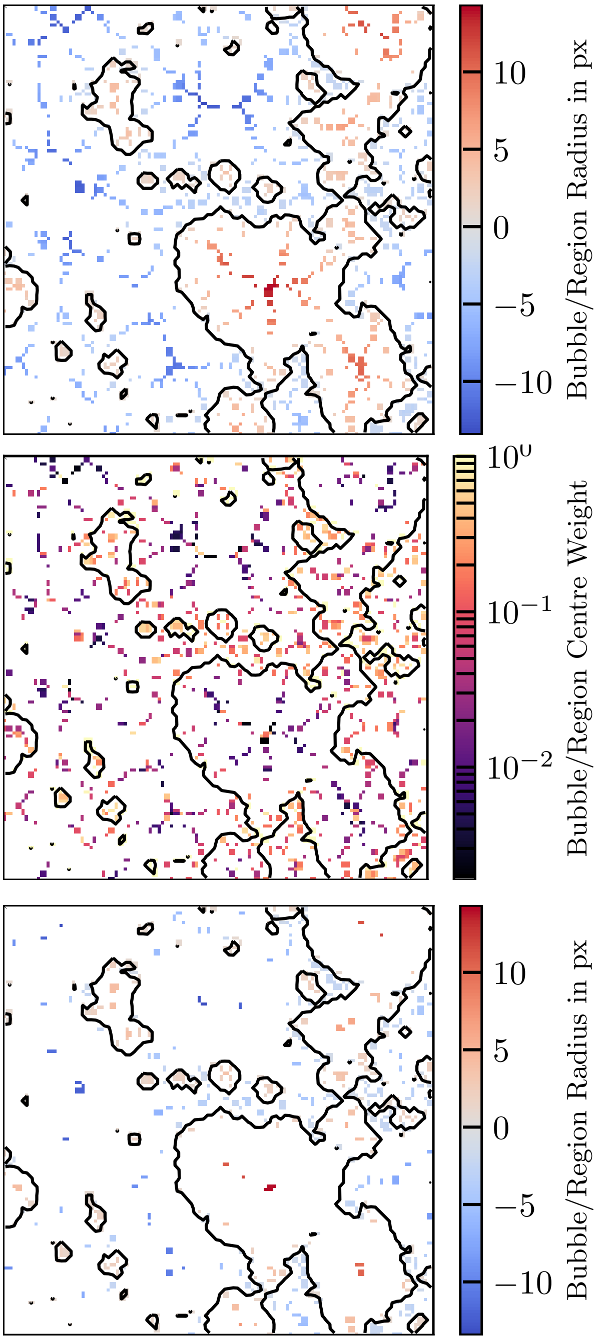

Figure 3.3: Minimal bubble structure (MBS) for the two-dimensional example field as given in Figure 3.1. The black contour shows the boundary between the two levels of the binary image. **Top Panel:** The radii of the ionised and neutral regions centred on a given centre. **Middle Panel:** The weights of the centres as calculated with the technique shown in Figure 3.2. **Bottom Panel:** The radii from the top panel filtered so that only those with a weight of $w_c \geq 0.1$ are retained.

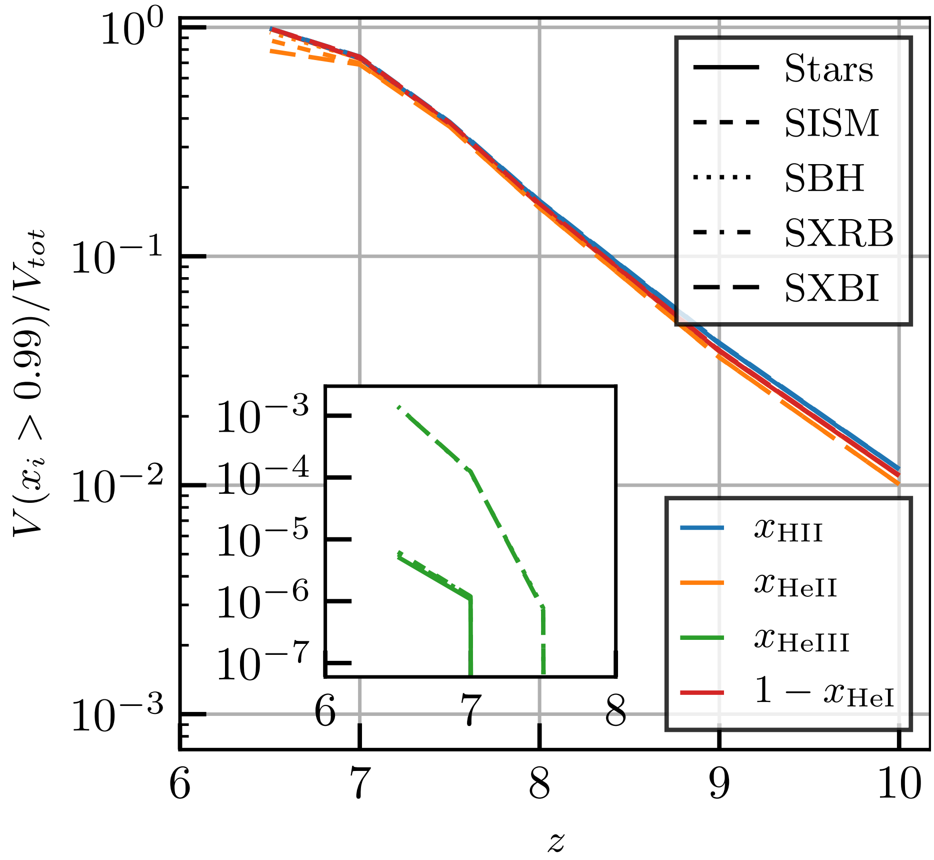

Figure 3.4: Volume fraction of cells with $x_i > 0.99$ for our redshift sample. The increased ionisation of HeII in the scenarios including ISM (and to a lesser degree BH emission) leads to a decrease in $V(x_{\rm HeII} > 0.99)$. This is driven by rather low $x_{\rm HeIII}$ values as can be seen in the inset. We also see that only the BH scenarios are able to produce larger numbers of cells with $x_{\rm HeIII} > 0.99$. This decreases $x_{\rm HeII}$ at these redshifts but not $1 - x_{\rm HeI}$.

HeII ionisation is so widespread that the HeIII component close to the sources does not have a strong impact anymore and HII, HeII and He(II + III) coincide. This only changes below $z = 7$ as the increasing ISM emission and BH population start to produce noticeable amounts of HeIII that decrease the number of cells with $x_{\rm HeII} > 0.99$.

Cells start passing the $x_{\rm HeIII} > 0.99$ threshold only at $z = 7.5$ if BHs are included or $z = 7$ in cases without them. The inset plot in Figure 3.4 shows that BHs are a necessity for a noticeable HeIII production. This might also have considerable impact on the observability of the hyperfine structure transition of HeII (see Section 1.2.3) as nuclear BHs will create cavities in the emitting component by further ionisation. While the ISM emission is not capable of lifting $x_{\rm HeIII}$ above the threshold in many cells it is nevertheless able to reduce the number of cells passing the $x_{\rm HeII}$ thresholds as discussed in the previous paragraph.

3.4.2 Global Ionisation Bubble Statistics

To investigate the typical bubble diameters which globally characterise the ionised regions we present the results of a granulometric analysis (see Section 3.3.2.2) for the five reionisation scenarios in HII and HeIII with the set of binary images described in the introduction to this section.

3.4.2.1 HII **and** He(II + III)

In Figure 3.5 we show the results of the granulometric analysis in terms of the volume weighted probability distribution of being in a region with opening radius R_{sph}. The most striking feature of the size distribution is the almost complete independence from the chosen source scenario. While we see minute differences for the largest radii at a given redshift, it is nonetheless clear that not just the global value, but also the spatial distribution of high $x_{\rm HII}$ values is completely set by stellar sources. Quantitatively we find a strong increase in the typical bubble radius from less than $1h^{-1}$ cMpc at $z = 10$ to about $20h^{-1}$ cMpc at $z = 7$. At the end of reionisation at $z = 6.5$ almost all cells are ionised and therefore most volume resides at the largest possible opening radius of $50h^{-1}$ cMpc in the simulation box.

We interpret this as a signature of the increasing overlap of bubbles into a single huge ionised envelope as we will discuss in 3.5

3.4.2.2 HeIII

As is to be expected, the HeIII bubble radii as depicted in Figure 3.6 are much smaller and are generally dominated by the one-cell-regions. There are indeed no cells passing the threshold at $z > 7$. As discussed in Section 3.4.1 we find a crucial dependence on BH-sources in the formation of HeIII, as the galactic emission (Stars, ISM, XRBs) is not able to carve out bubbles with diameters larger than one cell ($R_{sph} \lesssim 24h^{-1}$ ckpc). The rare, hard and powerful BH sources, on the other hand, are able to form bubbles with local diameters of up to $R_{sph} \approx 1h^{-1}$ Mpc. Interestingly we find a *decrease* in the typical bubble radius at $z = 6.5$ compared to $z = 7$. This is the effect of an increasing number of BHs that increases the total volume above threshold, but at the same time reduces the average bubble size.

3.4.3 Bubble Numbers

Now that we know how much volume is bound at a given opening level we would furthermore like to understand in how many connected regions this volume resides, i.e. how many bubbles, N_B, are present. We define two cells to be connected if they share a face, i.e. we use a 6-connectivity on the cubic lattice. As explained in the previous section we only show the results for HII and HeIII in Figure 3.7.

The bubble number distributions in HII at $z > 7$ follow a truncated power law as is to be expected from the luminosity distribution in a cosmological context. At early times ($z = 10$) the truncation shifts to larger radii with the inclusion of more energetic sources in

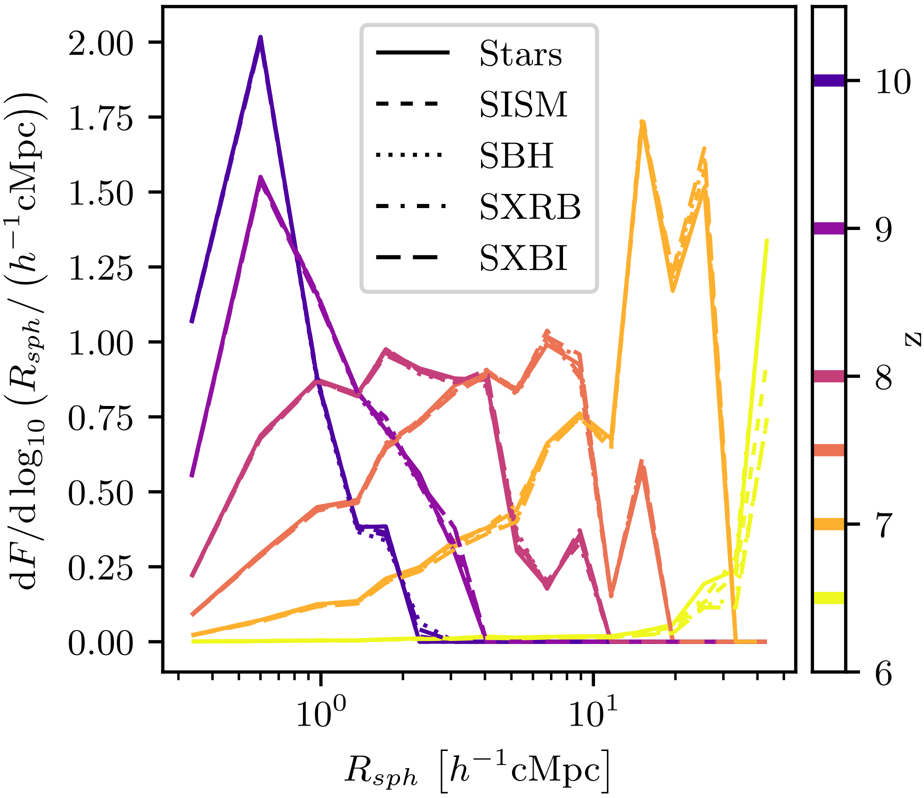

Figure 3.5: Volume weighted probability distribution of being in a fully ionised (i.e. with ionisation fraction > 0.99) region with opening radius R_{sph}. Line style refers to different source types, while color to redshift. We can see a clear shift of the volume weighted HII-region radius from less than $R_{sph} \lesssim 1h^{-1}$ cMpc at $z = 7$ to $R_{sph} \approx 20h^{-1}$ cMpc at $z = 7$.

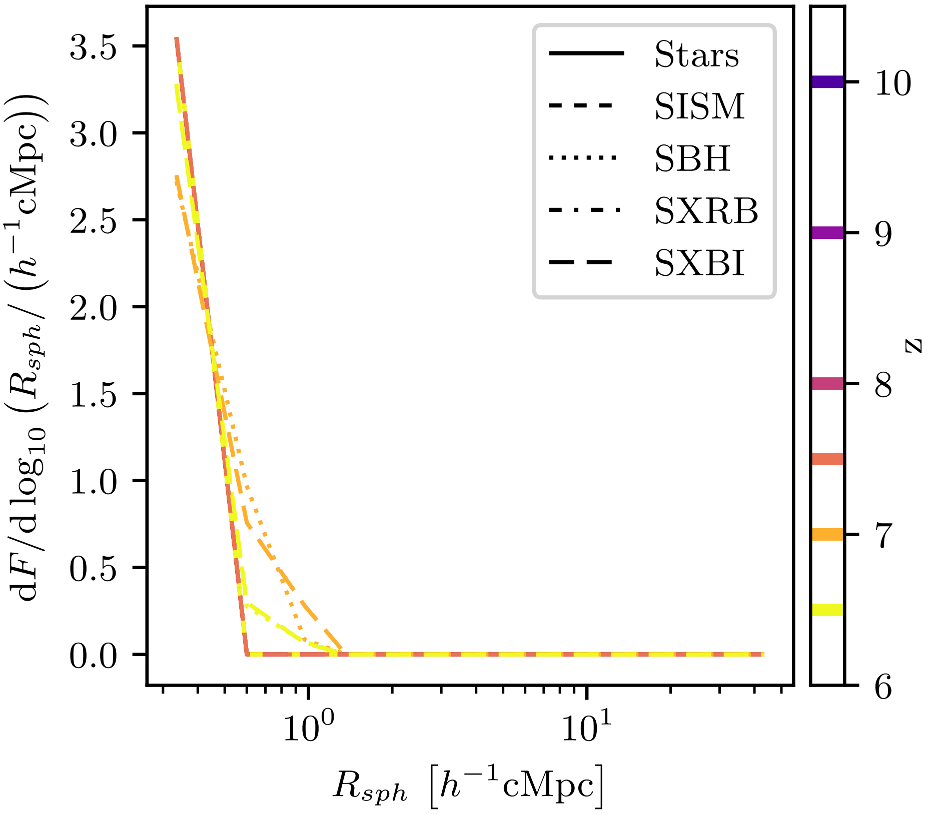

Figure 3.6: Same as Figure 3.5 but for HeIII. The volume weighted probability distribution of HeIII-region radii exhibits clear features of relatively large regions BH radiation. Although ISM-emission also has the

the X-ray enabled scenarios. These slight shifts are the results of a single bubble that exists at a radius of $1.9h^{-1}$ Mpc (and also $2.4h^{-1}$ Mpc for SXBI) around the most luminous source at this redshift. The single bubble is also responsible for the small changes in Figure 3.5. This effect of cosmic variance vanishes as ionisation proceeds. At $z = 8$ we find a smoother behaviour at the large end of the distribution, as the population of grown bubbles is now on average much better resolved. We note that the maximum number of independent bubbles is in place between $z = 9$ and $z = 8$ at $N_B \gtrsim 2 \cdot 10^4$ in our $\left(100h^{-1}\,\mathrm{cMpc}\right)^3$ box.

The $z = 7$ distribution does not exhibit the truncation at large radii, but instead extends all the way down to $N_B = 1$ in a now broken power law. There are also in total less independent bubbles than at previous times. This is an expression of the increasing overlap of bubbles in which smaller ones lose their independence and form superbubbles. The large numbers of bubbles existent at small opening radii are still shaped by the luminosity function, but at the large end we find the shallower number function of the superbubbles. It is also in this regime that we not only see a decrease with increasing radius, but also a momentary increase in N_B as regions of larger diameter may not be connected by smaller diameter connections anymore and are thus counted as disconnected regions.

This process continues at $z = 6.5$ where only a handful of bubbles with $R_{sph} < 5h^{-1}$ Mpc are still independent and almost all ionised volume resides in one large connected component at all scales.

For HeIII the situation is again very simple. Just as in the volume and granulometry, almost exclusively BHs are contributing to full HeII ionisation. Therefore, only in scenarios involving them do we find higher numbers of HeIII-regions starting at $z = 7.5$, to a noticeable extent from $z = 7$ onwards. The scenarios without them only have less than 20 and less than 100 single cell bubbles at redshifts 7.5 and 8, respectively. The bubbles in the BH-scenarios are more numerous and also more extended. We find the first $800h^{-1}$ ckpc bubbles at $z = 7$ at which point there are already over 250 bubbles. At $z = 8$ we already have over 3,000 bubbles.

3.4.4 (In-)Equality of the Bubble Volumes

Above we established the volume in bubbles above a given diameter and the number of these bubbles. However it is still unclear how to connect these two quantities, i.e. how the volume is distributed among the connected regions. To quantify this we employ the Gini coefficient of volume, G_V, which is a volume weighted measure of inequality in bubble volume as defined by

$$G_V = \frac{2\sum_{i=1}^{n} i v_i}{n\sum_{i}^{n} = 1 v_i} - \frac{n+1}{n}, \tag{3.11}$$

where v_i is the volume of the i-th of n bubbles. It is 0 for an absolutely equal distribution, i.e. the volume is distributed among bubbles of the same size, and 1 for an absolutely unequal distribution. The results are shown for HII in Figure 3.8.

At early times we find a decreasing inequality with increasing opening radius. This means that large bubbles are more similar in volume to each other than those obtained

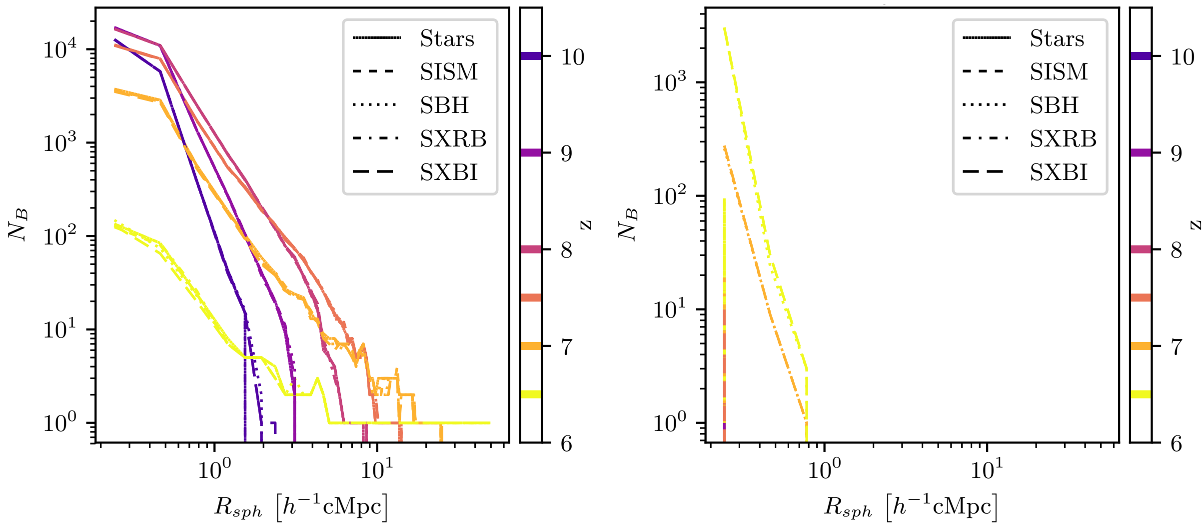

Figure 3.7: Number function of distinct bubbles above a given opening radius throughout reionisation. On the left for a threshold of $x_{\text{HII}} \geq 0.99$ and $x_{\text{HeIII}} \geq 0.99$ on the right. While the hydrogen bubble sizes at early times seem to follow the halo mass function before the number of bubbles decreases and a single large bubble appears. For HeIII there are barely any bubbles which all appear late and stay very small on the scale of the grid resolution.

with a smaller opening radius. This is a result of the fact that as we decrease the opening radius more bubbles come into the comparison and the large bubbles of comparable size now contain much of the total volume compared to a large number of newly included small ones.

In the range $8 \geq z \geq 7$ we observe an upturn in G_V for large radii as some large superbubble complexes start to dominate the volume budget at very large opening radii. As reionisation comes to an end below $z = 7$ we find a very unequal volume distribution at all radii. At this point almost all volume is concentrated in a single structure and thus there is no well-defined inequality in the volumes anymore.

We omit the analysis for HeIII as this ionisation state is not abundant enough in the epoch we consider.

3.5 Percolation Analysis

As the ionisation bubbles grow they start to overlap and form larger structures. The subsequent merging of these larger structures leads to the formation of a connected structure spanning the whole box and thus forming an infinite cluster. The formation of this object is mathematically speaking a percolation process.

We will investigate the percolation from two slightly different angles. First, we answer the question what size of bubbles form the percolating structure. In the second subsection we will investigate what diameter the overlaps have, or less formally, what size of sphere

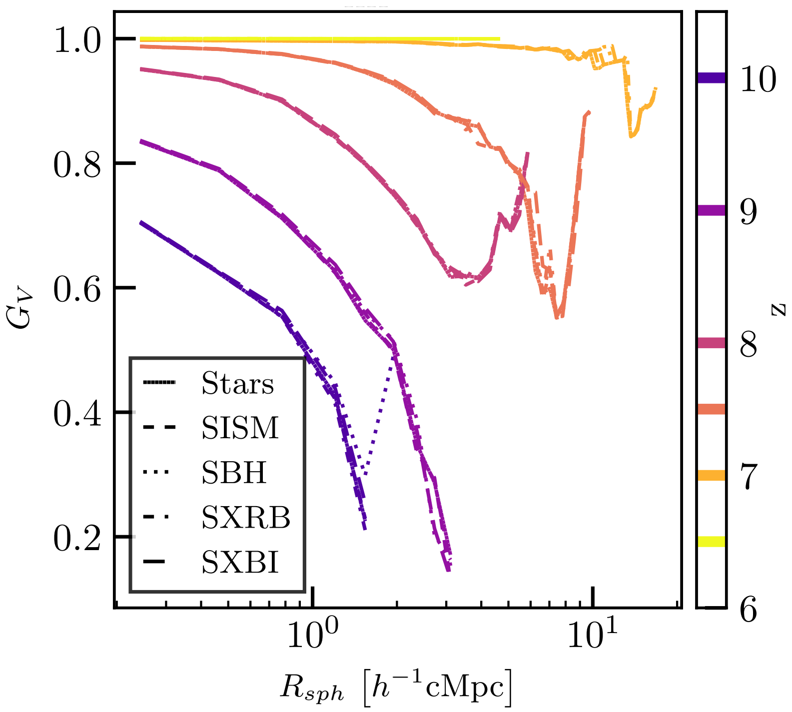

Figure 3.8: Volume Gini coefficient of the HII-bubbles in dependence of opening radius R_{sph}. A value of 0 marks absolute equality and 1 absolute inequality. If only one object is present the coefficient is not defined.

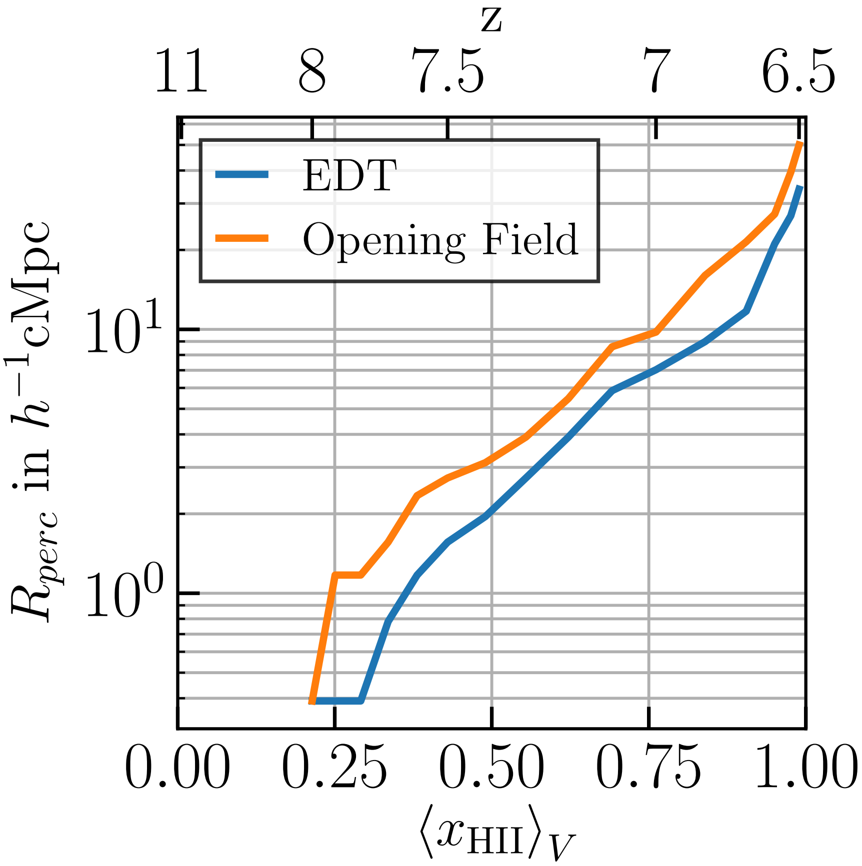

Figure 3.9: Percolation of reionisation with a threshold of $x_{\rm HII} = 0.99$ from two different perspectives. The opening field percolation measures which is the largest radius of bubbles which still form a percolating network. The EDT percolation tells us which distance to the surface will not be undershot. One can also understand this as the size of ball that can be freely moved through the percolating network.

can be pushed through a subset of the network that still percolates.

In the following we will adopt the same definition as above (Section 3.4) and see two cells as connected if they share a face.

We note that we are aware of the percolation results of Furlanetto and Oh (2016) which qualitatively agree but quantitatively differ somewhat from ours. It is hard to compare the results due to the very different approaches and resolutions used.

3.5.1 Percolation as a Function of Opening Radius

The percolation radius of the opening field as shown in Figure 3.9 describes at up to which radius the bubbles (here $x_{\rm HII} \geq 0.99$) percolate. If we restrict ourselves to regions larger than the indicated radius the connections are severed and no percolation occurs. The results in Figure 3.9 are obtained from the stellar sources only simulation but hold for HII in all scenarios. As HeII follows HII in general there is no appreciable difference either.

We see that percolation sets in at $\langle x_{\mathrm{HII}} \rangle \lesssim 0.25$ at $z \gtrsim 8$. As we mentioned in the previous sections on the size and number distributions of bubbles this coincides with a number of drastic changes indicative of percolation. Examples include the growing inequality among the largest radius bubbles and the peaking of the number of bubbles.

The percolation radius grows exponentially with global volume ionisation fraction before it approaches the box-size as a stand in for infinity, as the whole volume becomes ionised and the restricting neutral patches recede and make room for an overlap of the largest bubbles.

3.5.2 EDT Percolation

In Figure 3.9 we also show a curve labelled “EDT”. It shows in a similar fashion as the OF curve up to which radius the a thresholded version of the EDT field percolates. Just as the centrality value (see Section 3.3.4) has an upper bound of 1 because the OF in a particular spot will always be at least as high as the EDT value, so will the EDT percolation curve never exceed its OF counterpart.

3.6 Connecting Matter Density and Bubbles

While we formed an understanding of the population statistics of the ionised bubbles, we have so far neglected their connection to the universe in which they are formed. In this section we will therefore determine how bubbles are connected to the cosmic density field.

As the ionising radiation that creates bubbles originates in galaxies that form in matter overdensities, it is natural to assume that their centres lie on these same densities. Here we will investigate how much truth there is to this assumption. For this we will first take a look at the cross-correlation between the bubble centre distribution and the density field for some representative cases. We will then calculate the bias of the bubbles throughout the history of reionisation. Finally we invert the question and investigate the morphological history of patches of a certain density.

3.6.1 Cross-Correlation Between Bubble Centres and Matter Density

To understand what large scale density environments bubbles of different radius inhabit at different times we discuss the evolution of the cross-correlation, ξ_{bm}, between the bubble centres and the matter density. Here a bubble centre is a centre of the minimal bubble covering as presented in Section 3.3.5. The bubble centres are weighted in the correlation using the weights as described in the same section.

In the left panel of Figure 3.10 we show the correlations of bubble centres with the density field for various bubble radii and redshifts. At all times we find that the centres of bubbles with radii up to $\sim 7h^{-1}$ Mpc exhibit a correlation with the density field which increases with the bubble radius. This is a consequence of the need for a higher luminosity to produce larger bubbles that can only be satisfied in denser regions. This trend is reversed once larger bubbles appear, for which the central density drops below that of smaller bubbles, suggesting that an increasing number of these bubbles is not centred on a density peak but rather surrounding a group of sources. In most cases we also observe a rise in the cross-correlation up to a few h^{-1} Mpc, where the density peaks hosting the sources that formed the superbubble reside. Even in the case of these superbubbles large bubbles are more likely to be found in overdense regions. Therefore we find their cross-correlations exceeding these of smaller superbubbles at radial separations larger than $\sim 10h^{-1}$ Mpc.

The very smallest scale matter-bubble cross-correlation values drop steadily with increasing redshift, not only at a given bubble radius, but even for the currently largest populated bubble radius bin. This could come about via two different routes. Either we see the formation of bubbles around haloes in lower density local environments that are only possible at later times due to the diluted IGM and the increased ionising output of these haloes, or large bubbles increasingly form relatively small scale features on their surface in an increasingly structured IGM that inhibits uniform growth.

3.6.2 Bias of Bubble Centres

An important summary statistic of the cross-correlations between bubble centres and matter density field is the linear bias factor $b = \xi_{bm}/\xi_{mm}$ of the bubbles, where ξ_{mm} is the matter auto-correlation and ξ_{bm} is the bubble-matter cross-correlation as presented in the previous section. The bias of bubbles has previously been considered in the literature (e.g. in Meerburg et al., 2013), although only as a simple transformation from the bias of the sources, neglecting overlap and the detailed structure of the IGM and bubbles. We are unaware of any previous measurement in simulations.

We evaluate b (following Gao and White, 2007) in four logarithmic radial bins from $6h^{-1}$ Mpc to $20h^{-1}$ Mpc and find their best common value via least squares. We only show the bias for bins that include at least 100 points for the cross-correlation.

The distribution in Figure 3.11 shows that the bubbles are extremely biased tracers of the density field. In general, bubbles of all sizes appear first in overdense regions, and hence are highly biased. They subsequently grow and merge until a particular size of bubble can only be found in large scale underdense regions, becoming negatively biased. Here they are out of reach of the radiation from early forming galaxies, while they are ionised by galaxies forming at later times in these dynamically younger patches of the universe.

More specifically, during the initial phases of the reionisation process, small bubbles (with radii smaller than $\sim 5h^{-1}$ Mpc) appear that can essentially be associated with single haloes and have a relatively mild positive bias just as the sources that power them. As reionisation proceeds, these bubbles are the first to enter the anti-biased regime at $z \sim 8$, reaching values below -15 in an almost completely reionised universe. This is a consequence of the progressive merging of bubbles, that only leaves small bubbles in the most isolated regions of the universe. Larger bubbles, instead, reach much higher bias values, also exceeding 30 during the intermediate stages of reionisation. These stand at the be-

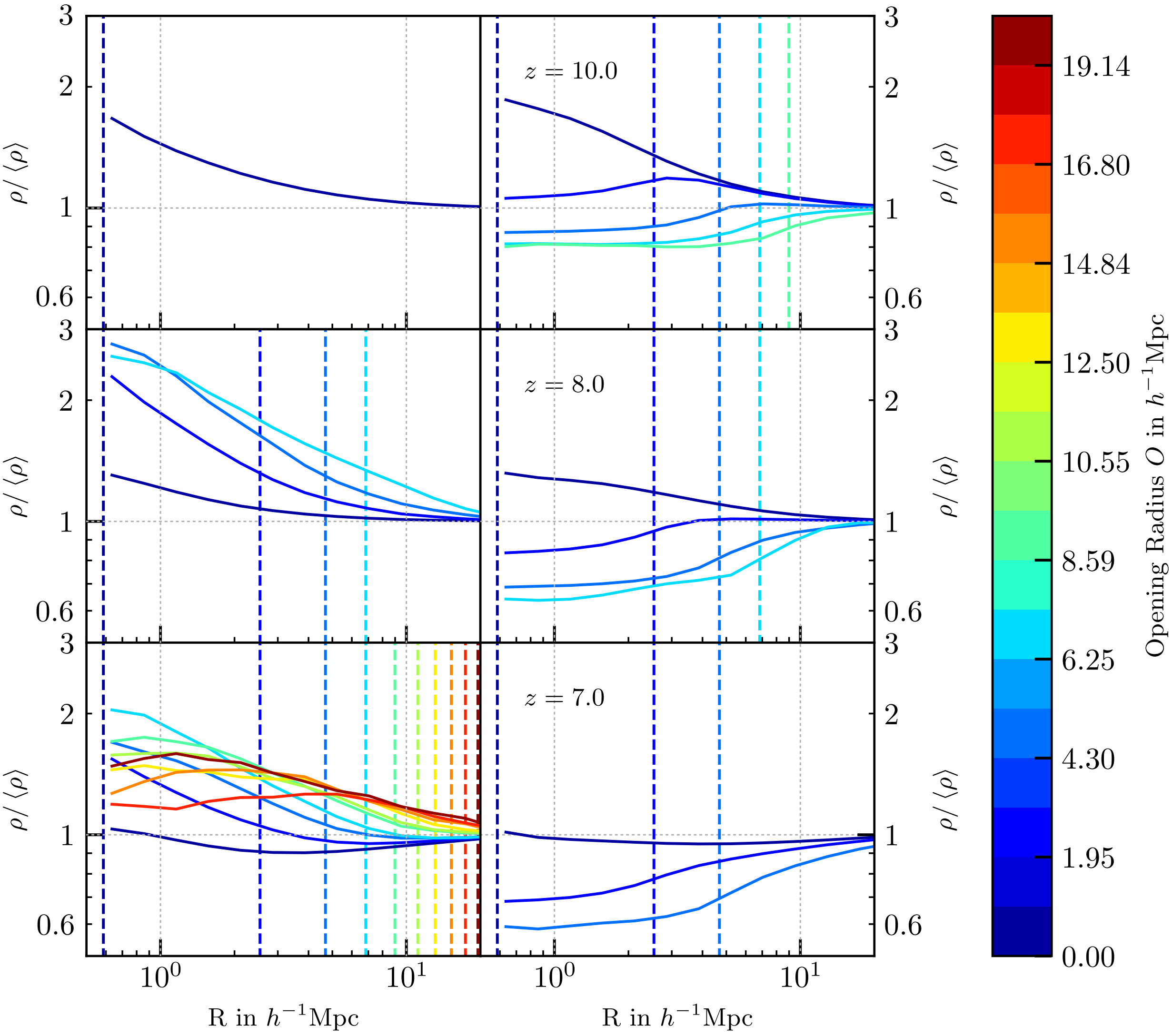

Figure 3.10: Cross-correlation ξ_{bm} between the matter density and the centres of ionised bubbles (left panels) and of neutral regions (right) at $z = 10$ (upper panels), 8 (middle) and 7 (lower). The bins of bubble radius is given by the color bar to the right. The vertical dashed lines of different colors indicate the radius of the bubble in relation to the correlation function.

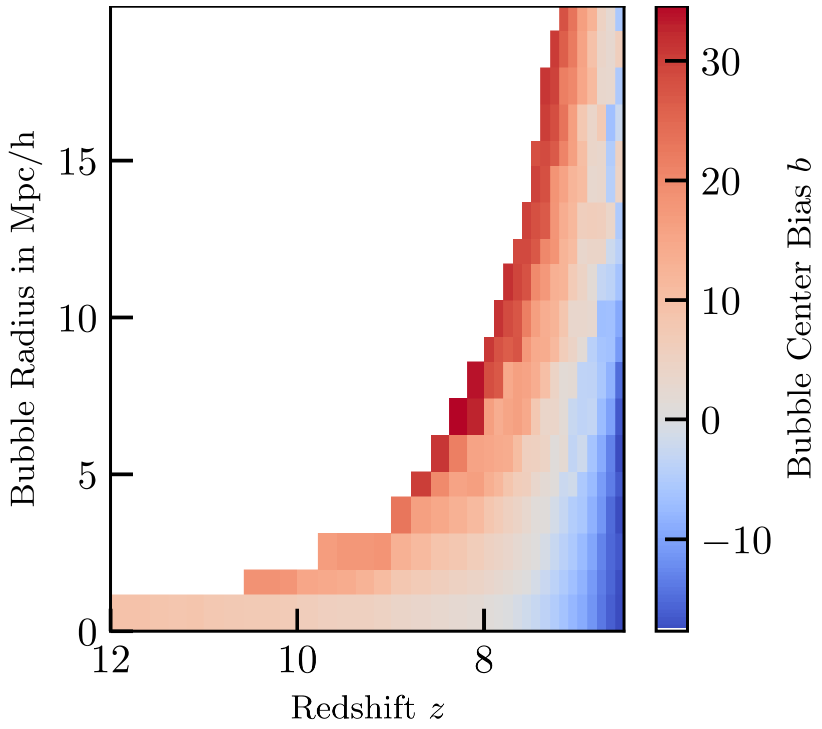

Figure 3.11: Bias of bubble centres, b, as a function of bubble radius and redshift. The largest bubbles at a given time are always the most strongly biased ones. In the later stages of reionisation we see a transition from uniformly positively bias to completely negative bias. This is the result of the overlapping process that only allows smaller bubbles with a sufficient separation to neighbouring bubbles.

ginning of the percolation process that first connects the high-density regions in an ionised network before filling in the underdense regions in between.

While we recognize that a simulation of $100h^{-1}$ Mpc side length is not an ideal testbed for this kind of studies we can nonetheless obtain first results in this setting. We plan to extend these investigations on larger boxes in the future.

3.6.3 Density Distribution Within Bubbles of Varying Size

In this section we want to complete our discussion of the relationship between bubbles and the density field by looking at the one-point statistics in bubbles. First we will investigate the density distribution within bubbles of a given size and then in which bubbles cells of a given density reside.

In Figure 3.12 we show the density distribution in ionised bubbles (red) and neutral regions (blue) of a given opening radius in comparison to the global density distribution (black). For this we bin the cells in the simulation in logarithmic density bins with N_{bin} cells in them and normalise that number to N_{tot}, the total number of cells with the given opening radius as indicated by the segments in the colour bar. We show the results at the same three redshifts that were used for the cross-correlations in the section above. Note that negative radii refer to distances into the neutral ($x_{\mathrm{HII}} < 0.99$) regions.

During the early stage of reionisation we find ionised bubbles and small neutral regions to be clearly overdense, while the larger neutral regions are mildly underdense. In fact, the smallest neutral regions at $z = 10$ are more overdense than the ionised volume. As the reionisation process proceeds and ionised bubbles become bigger this picture persists, but the ionised regions become less overdense, while the larger neutral regions become more underdense. Finally, towards the end of reionisation at $z = 7$, most of the volume is ionised and therefore the ionised and global density distribution coincide, while the (small) persisting neutral regions are all underdense. At all times the peak of the global and neutral or ionised density distributions are never separated by more than one FWHM.

The most remarkable finding from Figure 3.12 is that while the neutral regions of varying size show a clear order in their density distributions, the density distributions of the ionised bubbles of various sizes is very similar. This is insofar remarkable as the cross-correlations between bubble centres and density field in Figure 3.10 show clear differences even within the bubble radii.

3.6.4 Opening Radius Distribution of Cells of a Given Density

We now want to take the opposite perspective and investigate the opening radius distribution of cells of a given density. For this we plot the distribution of cells within 0.25dex bins in density as a function of the opening radius O (or bubble radius in the case of the ionised component) in Figure 3.13. As before, the black line signifies the marginal distribution over all densities and thus its crossing of the zero opening radius denotes the global fraction of neutral cells.

The three panels nicely illustrate the initial ionisation in the most overdense regions of the universe. This transitions into a bi-partition of neutral underdense and ionised overdense regions when 80% of the volume of the universe is neutral at $z = 8$ (note the zero-crossing of the black line in Figure 3.13). At this redshift, every cell denser than a few times the mean is completely ionised, while more than 90% of the underdense cells are still neutral. At $z = 7$ roughly 70% of the volume is ionised and at this point the underdense regions have also mostly followed and are less than 50% neutral.

The slope of the cumulative distribution shows the value of the differential distribution function in a given range. From the slope we learn that at $z = 10$ almost all the underdense volume resides in neutral regions with an opening radius between $5h^{-1}$ cMpc and $10h^{-1}$ cMpc. The most overdense cells are usually ionised but exclusively reside in bubbles of less than $2h^{-1}$ cMpc radius. Only $\sim 15\%$ of them are neutral and they heavily favour the smallest possible neutral regions. The less dense a cell is the more likely it resides in larger neutral patches and the less likely it will be in small neutral regions. This results in the prominent flattening just left of the $O = 0$ vertical.

At $z = 8$ this flattening has almost disappeared for overdense cells but it is still very clear for underdense cells and therefore the total distribution of cells. At this time the neutral regions have shrunk and now usually measure $3h^{-1}$ cMpc to $7h^{-1}$ cMpc. The most

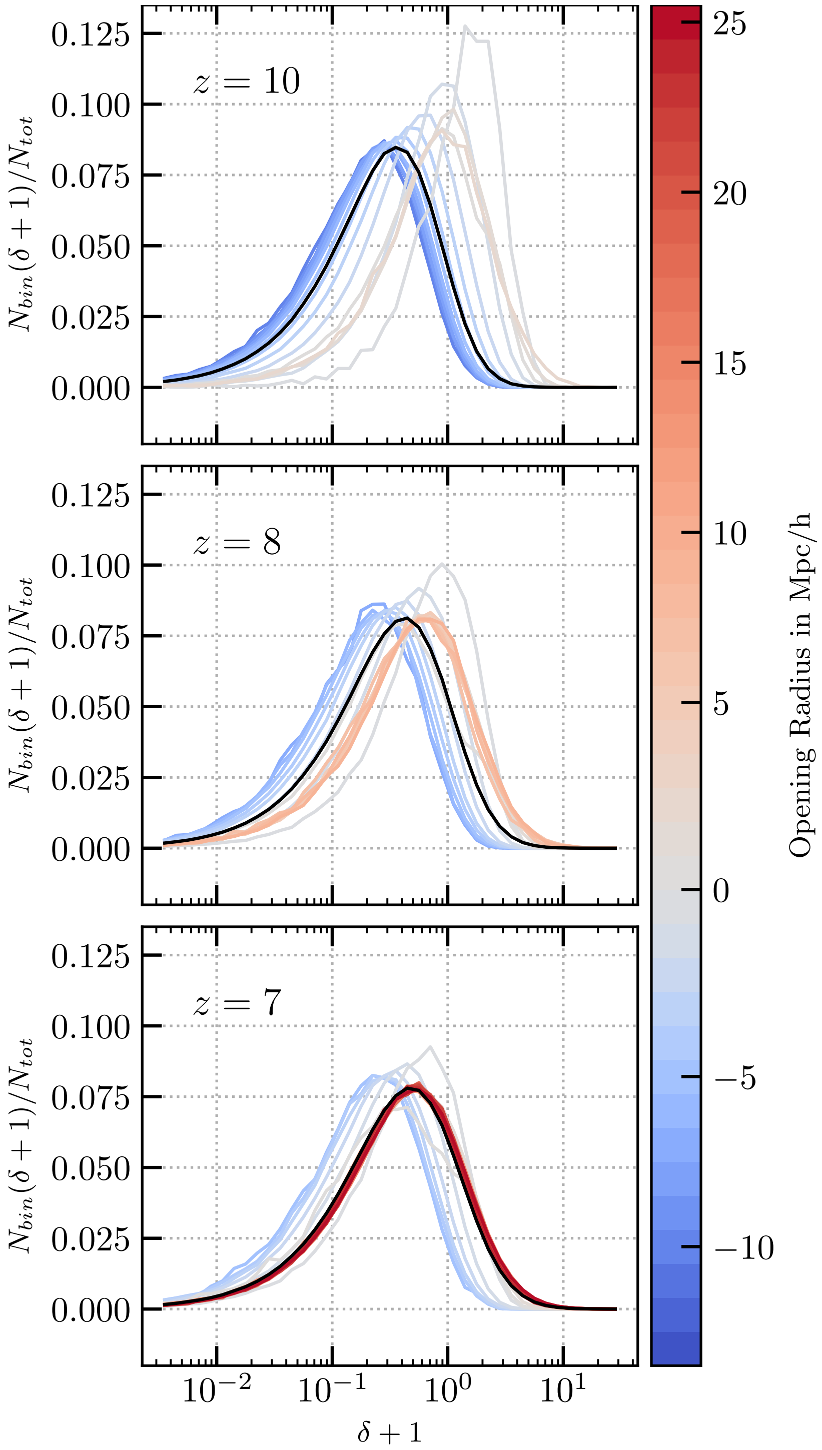

Figure 3.12: Density distribution in ionised bubbles (red lines) and neutral regions (blue) of various opening radius at $z = 10$ (upper panel), $z = 8$ (middle) and $z = 7$ (bottom). N_{bin} is the number of cells in the given logarithmic density bin and N_{tot} the total number of cells in the colour-indicated radius bin. The global density distribution is given in black.

overdense cells are now almost completely ionised and the regions they reside in have expanded to roughly $4h^{-1}$ cMpc. Indeed we find a few percent of the cells in the higher overdensity bins in bubbles up to $8h^{-1}$ cMpc radius.

In the bottom panel depicting the distributions at $z = 7$ we find that the underdense cells are still residing in $2h^{-1}$ cMpc to $7h^{-1}$ cMpc cells in 25 to 50% of cases, increasing with underdensity. They still do not inhabit small neutral regions. Even the least overdense cells are at least 73% ionised. The bubble sizes have increased dramatically and in fact more than half of the cells with more than 5 times the mean density reside in bubbles of more than $10h^{-1}$ cMpc radius. Underdense ionised cells on the other hand preferentially occupy smaller bubbles. There is a curious flattening just above $10h^{-1}$ cMpc and a subsequent steepening beyond $13h^{-1}$ cMpc. This is a sign of the increased importance of superbubbles and and the slight deficit in bubbles of just smaller radius than them (see Section 3.4.2.1).

These distributions show a clear "inside-out" progress of reionisation in which regions in the universe are generally ionised in order of decreasing density. The fact that neutral overdense cells are in narrower regions than their underdense counterparts supports this, as they show the residual neutral patches in the denser regions of the IGM are tightly wedged in between ionised regions.

3.7 Conclusions

We used a binary representation of the ionisation fraction fields of hydrogen and helium obtained by thresholding to investigate the morphology of the ionised bubbles. For this we transformed the binary images with the Euclidean distance transform and the morphological opening transform with a series of spherical structure elements of increasing radius. The combination of these two transforms allows us to deconstruct the bubbles into a minimal set of overlapping, maximally large spherical regions we call the minimal bubble structure. Measuring the volume of the regions, density distributions and their connectivity at different radius thresholds allows us to understand more about the bubble sizes and arrangements. Furthermore we can use the centers of the bubbles to find cross correlations with the density field to estimate typical density profiles and bias values of bubbles.

A very general finding of the present study is that in the chosen model, there are no morphological indicators of the hydrogen reionisation ionisation fraction that let us reliably distinguish between the different considered scenarios. This might change when looking at the 21cm-images obtained from the different scenarios as the harder radiation can change the thermal state of the neutral IGM between bubbles.

We find no difference in either totally ionised volume or number of bubbles between the different scenarios apart from an increase in HeIII-regions at late times in scenarios including hard sources. However, these regions are still so small that we cannot say much about their morphology given our resolution.

For most of the time range, $6.5 \leq z \leq 10$, considered here the volume fraction of totally ionised cells increases exponentially with decreasing redshift with a slight flattening when approaching unity between $z = 7$ and $z = 6$. At the same time we find that the connected

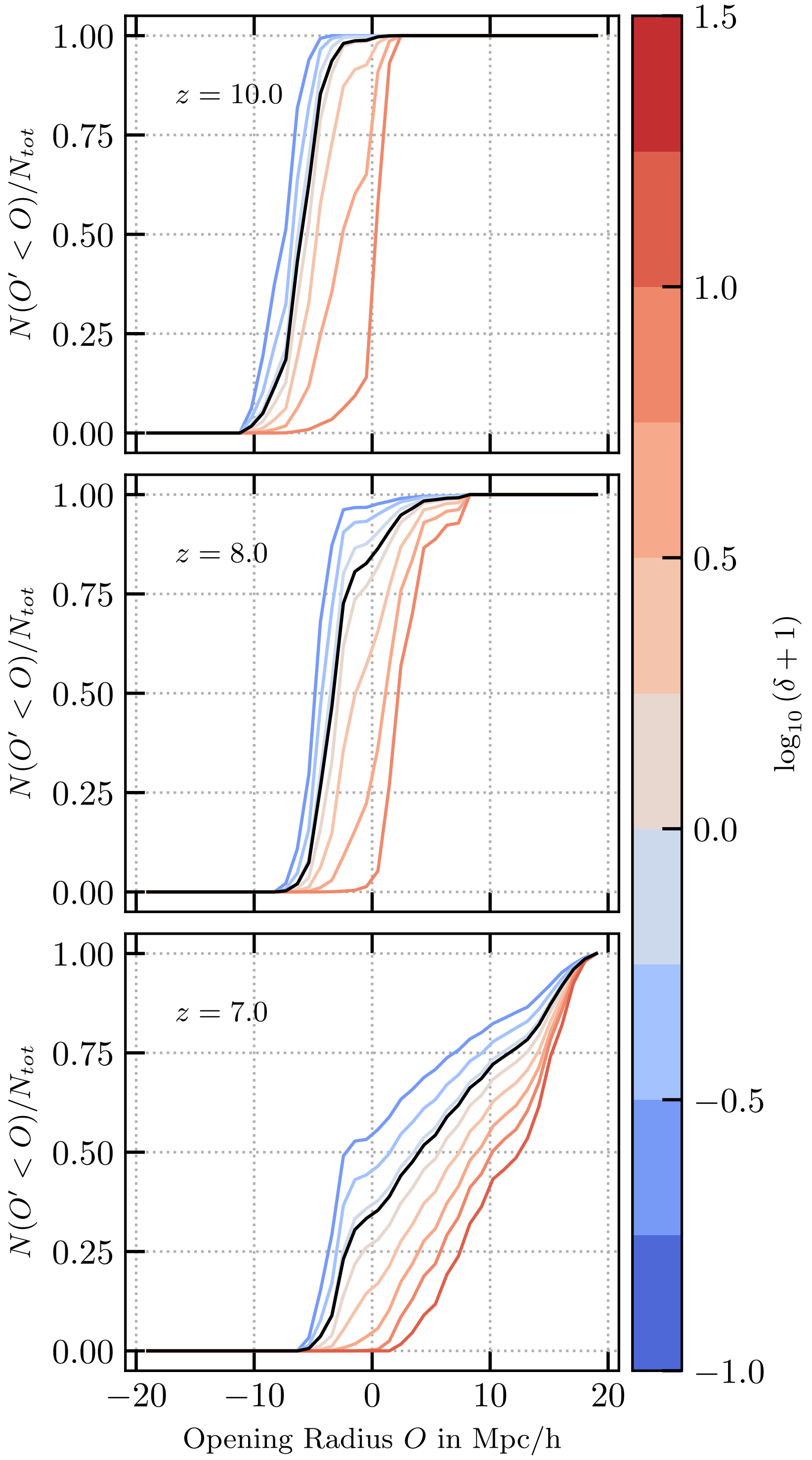

Figure 3.13: Cumulative opening radius distribution of cells of a given density as a function of bubble radius (with negative radii for neutral regions) at z=10 (upper panel), 8 (middle) and 7 (lower). The overdense (underdense) regions are represented by red (blue) lines. The black line shows the distribution marginalised over all densities. This means that its zero-crossing gives us the volume share in cells below the ionisation threshold $x_{\rm HII} = 0.99$.

ionised regions shift in their size distribution from a few cells and typical radii below $1h^{-1}$ cMpc to a few h^{-1} cMpc just before percolation sets during the height of reionisation ($7 \leq z \leq 8$) where most of the volume of the universe gets ionised. At the end of this process the volume is dominated by ionised regions with tens of h^{-1} cMpc radius. It is also at this point that the first resolved HeIII-regions appear in the presence of thermal ISM emission and AGNs, although they do not outgrow $1h^{-1}$ cMpc.

The volume in bubbles of different scales is distributed over a varying number of bubbles. Their size function changes from a shape reminiscent of the mass function of haloes that form them (a power-law with exponential cut-off at the upper end) at early times to a single connected region on the largest scales and just a few disconnected smaller regions, well after percolation. In the investigated time range HeIII stays in the first stage as the bubbles are still very much connected with the local drivers.

As one can already suspect from the number functions, the volume is distributed in a with time increasingly uneven fashion among the ionised regions. Quantifying this uneveness using the Gini-coefficient, we find that large bubbles tend to be of more similar volume than those found with smaller opening radii. This changes for the very largest bubble radii around the time of percolation as a dominating connected region appears, even for the largest bubble radii. Nevertheless, there are still separated large regions that lead to a Gini coefficient below unity. As the percolation radius progresses it asymptotically approaches this value. For regions below these extreme values, who's numbers are apparently dictated by occurrence of the driving haloes, we still find an increasingly uneven volume distribution with decreasing bubble radius. At the very end of reionisation, when only a few neutral islands are left, we find that $G_V \approx 1$ for all radii.

To find the time of the percolation process resulting from ionisation bubble overlap and the scales of the resulting object, we search for connected regions above a given radius. We find that percolation first occurs in our simulations at $z = 8/\langle x_{\mathrm{HII}} \rangle \gtrsim 0.2$. The percolating object rapidly expands its radius so that just after $z = 7.5$ at $\langle x_{\mathrm{HII}} \rangle = 0.5$ we already find a percolating object that allows a ball of $2h^{-1}$ cMpc to pass through it. Knowing the timing of the percolation process allows us to interpret the findings above.

Turning to the connection between bubbles and the cosmic density field, we show that the cross-correlations between the centres of ionised bubbles and neutral regions can be surprisingly counter-intuitive. At early times ($z = 10$) we find that the smallest neutral regions are actually in more overdense regions than their ionised counterparts. This can be explained by the fact that in order to restrict a neutral region to such a small size we need a number of ionised regions restricting it, which only happens in very special overdense regions. Larger neutral regions tend to be centred on underdensities of likewise increasing radius. As the neutral regions shrink with time, so does their central density decrease. In parallel there is an increase in central overdensity for ionised regions as they grow, while for a given size this quantity decreases. This regularity is broken during the late stages of reionisation around $z = 7$ as we find a few more peculiarities. Not only can small ionised regions only be found in large-scale underdense regions, but there is also a decrease in density towards the center of the largest ionised bubbles. We interpret this as the result of the merger of formerly separated smaller bubbles centred on high overdensities that then appear in the outskirts of the superbubbles, while smaller bubbles can only evade merging if they are far enough removed from the bulk of the ionised volume.

Using these bubble-matter cross correlation functions we also, for the first time to our knowledge but certainly using our bubble size estimate, calculate linear bias values for the bubble centers. We find the largest bubbles at a given time to be extremely biased with respect to matter with values of $10 \lesssim b \lesssim 35$. This is a result of the exceptional circumstances that are required to form these largest ionised regions. Only multiple regions of strong sources whose bubbles merge are able to produce the largest bubbles first. This is supported by the fact that the maximum bias value at a given time increases towards the point of percolation and subsequently decreases again, when the effect of radiation is increasingly unlocalised due to the dramatically increased mean free path. Just as we already saw in the correlation functions for the central values, we also find on large scales that small bubbles are avoiding matter, which leads to negative bias values. It is only in very underdense regions that they are not merged into large bubble complexes.

In a more detailed look at the density distribution within ionised bubbles of varying radius throughout reionisation we find that ionised bubbles start off much denser than the average volume subsequently approach the global density distribution. However, we find that the density distribution is independent of the size of the ionised bubble. The opposite is the case for neutral regions. Here we find that small neutral regions below $R \approx 5h^{-1}$ cMpc are very dense, initially even denser (at the grid scale) than ionised regions in the case of the smallest radius bin. They decrease in density with time but nevertheless retain their relative position to the ionised regions. The shrinking large neutral regions stay similar in their density distribution throughout reionisation.

To investigate the question of an "inside-out" or "outside-in" reionisation we compare the distribution over opening radii for cells of various density. Here we find that the most overdense cells almost immediately ionised by $z = 10$ while the deepest voids are still largely neutral at a stage when the universe as a whole is 70% ionised. This means that we agree with the common picture of an inside-out reionisation that first ionises overdense regions and then spreads into the voids.

3.A Implementation of Morphological Operations with the Fast Fourier Transform

The scale and resolution of our simulations dictates the use of rather large structuring elements. A simple direct-comparison approach for the morphological erosion and dilation operations scales as $\mathcal{O}(N \cdot M)$, where N is the number of simulation cells and M the number of cells in the structuring elements. M scales as R_{sph}^3 with the spherical opening radius R_{sph} and therefore becomes large very quickly. To speed up the operation, we reimplement the opening procedure and use a fast Fourier transform based approach following Kosheleva et al. (1997), which therefore has a scaling of $\mathcal{O}(N \log N)$. This approach uses two facts:

1. Morphological dilation $\oplus$ can be implemented as a convolution operation between the structuring element S and the BF X, which in turn can be implemented by a multiplication in Fourier space.

2. The erosion of a BF is just the negation of the dilation of the negation of the same BF: $X \ominus S = \neg(\neg X \oplus S)$.

As we only implement a new dilation operation, the opening now becomes

$$X \circ S = (\neg(\neg X \oplus S)) \oplus S. \tag{3.12}$$

While the new approach slows the calculation for small structuring elements by orders of magnitude, it also decreases the run time by similar factors for large structuring elements. While we solely rely on the Fourier approach, a scenario with many openings using small structure elements would warrant either an adaptive choice of the dilation procedure or complete return to the direct approach.

We point the reader to Section 3.3.3 for an approach to calculating the opening field in an even more time-saving manner if the Euclidean distance transform as described in the referred to section is also to be calculated.

Chapter 4

Assembly Bias and Splashback in Optically Selected Galaxy Clusters

The contents of this chapter were published as Busch and White (2017).

4.1 Introduction

Until recently, attempts to detect an observational signal of assembly bias (see Section 1.1.3) were inconclusive (e.g Yang et al., 2006; Tinker et al., 2012; Wang et al., 2013; Hearin et al., 2014) and controversial (e.g. Lin et al., 2016b). A strong indication of assembly bias as a function of halo concentration was identified by Miyatake et al. (2016) in their study of weak gravitational lensing by a large sample of clusters identified in the SDSS/DR-8 photometric data. Their result was confirmed at much higher signal-to-noise by More et al. (2016), who cross-correlated this same cluster sample with individual SDSS galaxies. In both studies, the mean projected distance of individual cluster members from cluster centre was adopted as a measure of concentration and used to split the sample into equal high- and low-concentration subsamples. Differences at large radius in the mean projected mass and galaxy number density profiles of these two subsamples then provided the evidence for surprisingly strong assembly bias, $b_{lo}/b_{hi} \sim 1.5$.

More et al. also used their stacked galaxy number density profiles to search for splashback signals. From Diemer and Kravtsov (2014) we expect an increase in the splashback radius with an increase of the concentration (see Section 1.1.5). When More et al. (2016) examined the profiles of their low- and high-concentration subsamples, however, they found the opposite ordering both in the minimum slope value and in the radius where it is attained. In addition, these radii were smaller than they expected given their estimates of cluster mass, particularly for the high-concentration subsample. Assuming that cluster galaxies trace the dark matter density profile of their host halo at these outer radii this is in conflict with the simulation results.

The cluster sample analysed by Miyatake et al. (2016) and More et al. (2016) was based on application of the redMaPPer algorithm (Rykoff et al., 2014) to the SDSS/DR8 photometric galaxy catalogues. As its name implies, this cluster finder uses only the non-star-forming 'red' galaxies in the catalogue. Clusters are assumed to be centred on their brightest red galaxy, and every red galaxy is assigned a probability of belonging to any particular cluster which depends on its projected distance and maximal possible redshift offset (based on the SDSS photometry) from the cluster central galaxy. This necessarily introduces a non-negligible uncertainty in the true redshift spread among cluster members. The effect of this uncertainty on cluster properties is one of the main focuses of this chapter. Another important element of redMaPPer is the introduction of an outer cluster radius that increases slowly with the number of cluster members and is used by the algorithm to define the cluster richness and to limit the projected region over which membership probabilities are non-zero. As we shall show below, this radius, in part because of its important role in the definition of cluster concentration used by Miyatake et al. (2016) and More et al. (2016), has a significant influence on the apparent assembly bias and splashback signals identified by these authors.

This chapter is organized in seven sections. Following this introduction, Section 4.2 describes the publicly available simulation data we use, the simplified versions of the redMaPPer and concentration estimation procedures that we apply to them, and the global properties of the resulting cluster samples. Section 4.3 begins by demonstrating that our simulated cluster samples reproduce quite well the projected mean mass and galaxy number density profiles obtained by Miyatake et al. (2016) and More et al. (2016), including the strong apparent assembly bias signal and the surprising concentration-dependence of the apparent splashback signal. We then investigate how this apparent success is affected by the maximum offset in depth allowed for potential cluster members, our simplified representation of the effect of photometric redshift uncertainties. In Section 4.4, we study how well the assembly bias and splashback features measured in projection correspond to their analogues inferred from the full three-dimensional mass and galaxy distributions. Section 4.5 then looks in more detail at our stacked profiles to clarify the distribution in depth of the galaxies which give rise to the differences in mean projected galaxy number profile between low- and high-concentration clusters, while Section 4.6 examines how profile shapes are influenced by the radius used by redMaPPer as the effective limit of clusters. Finally, Section 4.7 gives our principal conclusions.

While we were completing the analysis for this chapter, Zu et al. (2017) published a preprint in which they repeat the lensing analysis of Miyatake et al. (2016) but with the cluster sample split according to a modified definition of concentration which, as they demonstrate, is significantly less sensitive to projection effects. With this new definition, low- and high-concentration clusters show no detectable large-scale assembly bias. Zu et al. (2017) conclude, as we do below, that the strong signal in the original analysis is a result of projection effects. Our own analysis (in Section 4.5) shows explicitly how this contamination of the low-concentration clusters is distributed in depth and explains why it produces an apparently constant assembly bias signal at large projected separations.

4.2 Methodology

Our goal in this chapter is to see whether the assembly bias and splashback signals detected by Miyatake et al. (2016) and More et al. (2016) are consistent with current models for galaxy formation in a ΛCDM universe. In particular, we would like to understand the origin of the strong observed dependence of bias on cluster concentration, of the unexpectedly small scale of the detected splashback signal, and of the fact that this signal varies between high and low concentration clusters in the opposite sense to that expected both in strength and in radius. For this purpose, we need a realistic simulation of the formation and evolution of the galaxy population throughout a sufficiently large volume for our analogue of redMaPPer to identify a large sample of rich galaxy clusters.

4.2.1 Data

4.2.1.1 Dark matter distribution

Our analysis is based on the *Millennium Simulation* described in Springel et al. (2005). This followed structure development within a periodic box of side $500h^{-1}$ Mpc assuming a flat ΛCDM cosmology with parameters from the first-year WMAP results. Although these parameters are not consistent with more recent data, the offsets are relatively small and are actually helpful for this work since they enhance the abundance of rich clusters in the mass range of interest. The dynamical N-body simulation followed the collisionless dark matter only, representing it with $2160^3 \sim 10^{10}$ particles of individual mass $8.6 \times 10^8 h^{-1} M_\odot$ and gravitational softening length $5h^{-1}$ kpc.

Haloes and their self-bound subhaloes were identified in 64 stored outputs of this simulation using the SUBFIND algorithm (Springel et al., 2001), and these were linked across time to build subhalo trees which record the assembly history of every $z = 0$ halo and its subhaloes. These trees are the basis for simulation (in post-processing) of the formation and evolution of the galaxy population. Galaxies are assumed to form as gas cools, condenses and turns into stars at the centre of every dark matter halo and are carried along as halos grow by accretion and merging. Both the subhalo merger trees and the specific galaxy formation simulation used in this work (and discussed next) are publicly available in the Millennium Database[1] (Lemson and the Virgo Consortium, 2006).

4.2.1.2 The galaxies

The particular galaxy population used in this work was created using the semianalytic model described in detail in Guo et al. (2011). These authors implemented their model simultaneously on the Millennium Simulation and on the 125 times higher resolution but smaller volume Millennium-II Simulation (Boylan-Kolchin et al., 2009). This allowed them to tune its parameters in order to reproduce the $z = 0$ galaxy population over a very wide mass range. In this work we will only need to consider relatively bright galaxies, well above

---

[1] `http://www.mpa-garching.mpg.de/Millennium/`

the limit to which results for the two simulations converge. As a result we will only use data from the larger volume simulation. We will analyse the simulation data from a single snapshot at $z = 0.24$. This is the mean redshift of the clusters in the SDSS sample we compare with and is the closest snapshot to its median redshift of 0.25.

For all galaxies, the simulated galaxy catalogue provides positions, velocities and a range of intrinsic properties, including estimated magnitudes in the SDSS photometric bands. We restrict ourselves to galaxies with i-band absolute magnitude, $M_i < -19.43 + 5\log_{10} h$, which, for our adopted value $h = 0.7$, gives $M_i < -20.20$. The chosen magnitude limit is very close to the one corresponding to the redMaPPer luminosity limit of $0.2L_*$ at $z = 0.24$, i.e. $M_i = -20.25$ (see Rykoff et al., 2012). This selection criterion leaves us with 2,239,661 galaxies and matches that adopted by More et al. (2016) for their SDSS galaxies in order to achieve volume completeness over the redshift range, $0.1 \leq z \leq 0.33$.

The next step in mimicking redMaPPer procedures is to define a class of passive or 'red' galaxies. For simplicity, we require the specific star formation rate (SSFR) of model galaxies to lie below $1.5 \times 10^{-11}\, h\,\mathrm{yr}^{-1}$. This avoids using model colour directly which would introduce a dependence on the (uncertain) modelling of dust effects. However, the two methods produce very similar results in practice, so the choice has has no significant effect on our analysis. 897,604 galaxies qualify as red by our criterion.

4.2.2 Cluster Identification and Classification

Given the galaxy data described above, we wish to identify clusters using a simplified version of the scheme applied to the SDSS photometric data to generate the catalogue analysed by Miyatake et al. (2016) and More et al. (2016). We project the simulated galaxy and mass distributions along each of the three principal axes of the Millennium simulation to obtain three 'sky' images, for each of which depth information is available for the galaxies either in real space or in redshift space. In the latter case, the line-of-sight peculiar velocities of galaxies are added to their Hubble velocities to produce redshift space distortions (RSD). These are important when considering how the use of photometric redshifts affects the assignment of galaxies to clusters (see 4.2.2.1). The following describes our cluster identification scheme and explains how we split the clusters into equal high- and low-concentration subsamples.

4.2.2.1 Cluster identification algorithm

Our cluster identification algorithm, inspired by redMaPPer, finds clusters in the projected distribution of red galaxies. Every red galaxy in each of our three projections is considered as the potential centre of a cluster. The algorithm grows clusters by adding new red galaxies (defined as in 4.2.1.2) in order of increasing projected separation until the richness λ and the cluster radius R_c reach the largest values satisfying the relation given by Rykoff et al. (2014),

$$R_c(\lambda) = 1.0 \left(\frac{\lambda}{100}\right)^{0.2} h^{-1}\,\mathrm{Mpc} \tag{4.1}$$

Table 4.1: The size of simulated cluster samples for different maximal depth offsets, Δz_m.

Sample Name	Δz_m h^{-1} Mpc	No. Members
CS60	60	9196
CS120	120	9213
CS250	250	8930

in physical (rather than comoving) units. Initialising with $\lambda = 1$ and $R_c(1)$,

1. we consider as possible members the N_g red galaxies which lie within R_c and have a (redshift space) depth offset below Δz_m ,
2. we calculate $\bar{N}$, the expected number of uncorrelated ('background') galaxies within R_c and Δz_m,
3. we update $\lambda = N_g - \bar{N}$ and $R_c(\lambda)$,
4. we check whether the current central galaxy still has a higher stellar mass than any other cluster member, otherwise we delete it as a potential central and move to the next one,
5. we start the next iteration at (i) if λ has increased, otherwise we stop.

This process usually converges quickly and only in a few cases is it unsuccessful in finding a cluster. Note that we choose to require that the central galaxy should be the one with the highest stellar mass. Only in ~ 5 per cent of the cases is it not simultaneously the brightest in the i-band, and we have checked that choosing to require instead that it should the most luminous has a negligible effect on our results. In the following we will only consider clusters with $20 \leq \lambda \leq 100$, again in accordance with More et al. (2016).

We will consider three different values for the maximal redshift-space offset allowed for cluster members, $\Delta z_m = 60h^{-1}$ Mpc, $120h^{-1}$ Mpc and $250h^{-1}$ Mpc; the largest of these is equivalent to projecting through the full Millennium Simulation. For comparison, the 1σ uncertainty in the photometric redshift of a single SDSS red galaxy is estimated by Rykoff et al. (2014) to be about $90h^{-1}$ Mpc at the median redshift of the observed cluster sample. The total number of clusters found (summed over the three projections) is given in Table 4.1.

These numbers are similar to the number of clusters (8,648) in the observed sample we are comparing with. This is a coincidence since the volume of the Millennium Simulation is only about a tenth of that in the SDSS footprint over the redshift range $0.1 \leq z \leq 0.33$, but the abundance of rich clusters is enhanced by a factor of about three in the simulation because it assumes $\sigma_8 = 0.9$, significantly above current estimates[2].

---

[2] We checked the impact of changing the cosmology on our results using the public semianalytic catalogue of Henriques et al. (2015) which is implemented on a version of the Millennium Simulation rescaled to the

Table 4.2: The fractional overlap between different cluster samples.

Base sample	Comparison sample		
	CS60	CS120	CS250
CS60	1.0	0.876	0.736
CS120	0.874	1.0	0.783
CS250	0.758	0.808	1.0

There is, of course, a very substantial overlap between these three cluster samples, but it is not perfect. In Table 4.2 we give the fraction of clusters in a given sample that share their central galaxy (in the same projection) with a cluster in a comparison sample and pass the richness filter in both. We see that most clusters are indeed duplicated. Those that are not, fail because in one of the two samples either a more massive potential member is included or the richness falls outside the allowed range. Such differences are a first indication of sensitivity to projection effects, an issue that is discussed further in subsection 4.2.2.3.

Notice that the algorithm described above allows a given galaxy to be considered a member of more than one cluster. Although the majority of our simulated clusters do not have such overlaps, they are not negligible; the fraction of clusters which share at least one galaxy with another cluster in the same projection is 18.8, 21.8 and 26.7 per cent for CS60, CS120 and CS250, respectively. The average number of galaxies in these overlaps is ~ 14, which should be compared with the mean number of galaxies per cluster which is 37 to 46. In order to check the importance of the overlaps, we have repeated our analysis using only the $\sim$ 80–75 per cent of clusters which have no overlap. These are clearly a biased subset with respect to their surroundings, and as a result the stacked profiles change noticeably. However, the conclusions we draw below are not significantly affected, and for the rest of this chapter we show only results based on the full cluster samples, noting that the redMaPPer algorithm also allows a given red galaxy to be considered part of more than one cluster, albeit by assigning probabilities to each potential membership based on the galaxy's photometric redshift, its projected separation from each cluster centre, and the richness of the clusters. The consistent use of such probabilities is the principal difference between the actual redMaPPer algorithm and the simplified version we use here.

4.2.2.2 Cluster concentrations

At the core of the following analysis is the separation of each cluster sample into two equal subsamples with identical richness distributions, but disjoint distributions of concentration $c_{\rm gal}$ as introduced by Miyatake et al. (2016). This concentration is based on the mean projected distance from cluster centre of red galaxy members, $c_{\rm gal} = R_c/\langle R_{\rm mem}\rangle$ where in

---

Planck 2013 cosmology (Planck Collaboration, 2014). We find far fewer clusters: 2407, 2244 and 2307 for the equivalents of CS250, CS120, and CS60, respectively. This corresponds to 83.1%, 77.5% and 79.6% of the expected number of clusters in three times the (rescaled) volume of the simulation. We decided to stay with the original cosmology since the larger number of clusters provides much better statistics.

our case

$$\langle R_{\rm mem} \rangle = \frac{1}{N_{\rm mem}} \sum_{i}^{N_{\rm mem}} R_{{\rm mem},i}. \tag{4.2}$$

We classify a particular cluster as high or low concentration, depending on whether $c_{\rm gal}$ lies above or below the median for all clusters of the same richness. For richness values with fewer than 200 clusters in a given sample, we bin together neighbouring richness bins to exceed this number before determining the median. For the observed clusters Miyatake et al. (2016) binned clusters by both richness and redshift before determining the median, but redshift binning is not necessary for the simulated samples since they are all taken from the same simulation output.

4.2.2.3 The cluster-halo correspondence

It is not straightforward to connect a galaxy cluster defined in projection with a specific three-dimensional cluster, in our case a specific SUBFIND halo. The idealised model of a spherically symmetric cluster centred on its most massive galaxy and containing all the cluster members identified in projection corresponds poorly to most of the clusters identified either in the simulation or, most likely, in the SDSS. In almost all cases, the galaxies identified as members in 2D reside in multiple 3D objects distributed along the line-of-sight. This makes the cross-identification between 2D and 3D ambiguous.

Here we consider two possibilities for defining the 3D counterpart of each 2D cluster: the dark matter halo that hosts the central galaxy and the one that hosts the largest number of member galaxies. The former definition follows the logic of the cluster centring, while the latter ensures that the richness of the 3D counterpart corresponds most closely to that of the 2D system. It is interesting to see how often these definitions coincide, i.e., how often the central galaxy is actually part of the dominant galaxy aggregation along the line-of-sight. We give in Table 4.3 the fraction of clusters in each of our three samples for which both methods lead to the same FoF halo. These numbers show that that the two definitions are generally in good agreement, and that this is better for smaller maximal depth offsets and for more concentrated clusters. These trends reflect the projection effects discussed in detail in Section 4.5.

It is also interesting to see how many of the potential cluster members identified in 2D are, in fact, part of the same 3D object. For each of our clusters we find the maximal fraction of its members contained in a single 3D FoF halo. The third column of Table 4.3 then gives the average of these fractions. This can be compared with the average fraction of its members contained in the FoF halo of its central galaxy (fourth column) and with the average expected as a result of our background correction, $\langle \lambda / N_g \rangle$, given in the last column.

The values for $\langle F_{\rm biggest} \rangle$, $\langle F_{\rm central} \rangle$ and $\langle \lambda / N_g \rangle$ in Table 4.3 show that we consistently find more 'foreign' galaxies in our clusters than we would expect from contamination by a uniform background. The more concentrated clusters have contamination ratios close to, yet still a few percent below the expected ones. The low-concentration clusters have

Table 4.3: The fraction of clusters where the central galaxy resides in the FoF halo contributing the largest number of potential 2D members; the mean fraction of such members in this halo; the mean fraction of such members in the FoF halo of the central galaxy; the mean membership fraction predicted by 'standard' background subtraction.

Subs.	Sample	F_{centred}	$\langle F_{\text{biggest}} \rangle$	$\langle F_{\text{central}} \rangle$	$\langle \lambda / N_g \rangle$
All	CS60	0.93	0.826	0.803	0.922
	CS120	0.903	0.755	0.726	0.856
	CS250	0.848	0.635	0.595	0.743
high c_{gal}	CS60	0.983	0.880	0.874	0.922
	CS120	0.973	0.819	0.812	0.855
	CS250	0.948	0.709	0.697	0.742
low c_{gal}	CS60	0.876	0.772	0.732	0.923
	CS120	0.833	0.69	0.64	0.857
	CS250	0.749	0.561	0.494	0.744

contamination fractions more than twice the expected values. We therefore conclude that the identified clusters are biased towards arrangements of multiple objects along the LoS, especially in the low c_{gal} case. Again, this is very much in line with our discussion on the preferential selection of aligned systems in Section 4.5.

4.3 Results In Projection

We are now in a position to investigate whether the assembly bias and splashback features identified in SDSS data by Miyatake et al. (2016) and More et al. (2016) are reproduced when our simplified version of the redMaPPer algorithm is applied to the public Millennium Simulation data . We begin by comparing the observed mean galaxy and mass profiles to directly analogous profiles for CS250, finding that both the surprisingly strong assembly bias and the unexpected properties of the apparent splashback signal are reproduced well. Most differences can be ascribed to the finite size of the simulation or to the simplifications of our cluster identification scheme. We then use our three cluster catalogues to investigate the dependence of these successes on Δz_m, the maximal depth offset allowed for potential cluster members, finding that the assembly bias signal is sensitive to this parameter but the splashback features are not. Finally we look in more detail at the radial dependence of the ratio of the profiles of low- and high-concentration clusters. Later sections employ the full 3D information available for the simulation to explore the origin of the observed features, and vary our cluster identification scheme to demonstrate how its imprint on the measured profiles can confuse identification of the splashback signal.

4.3.1 Comparison of profiles for SDSS and CS250

We collect the main profile results for the CS250 sample in Figures 4.1 to 4.3. Here and in the following, unless noted otherwise, the solid line represents the median value from 10000 bootstrap resamplings of the relevant cluster sample. The shaded regions denote the 68 per cent (darker) and 95 per cent (lighter) confidence intervals around this median.

We calculate the mean galaxy surface number density profile for each cluster sample as

$$\Delta\Sigma_g(R) = \Sigma_g(R) - \bar{\Sigma}_g \tag{4.3}$$

where we use all galaxies brighter than $M_i = -19.43 + 5\log_{10} h$, not just the red ones, and we impose no maximal depth offset from the cluster. $\Sigma_g(R)$ is then the mean over all clusters of the surface number density of such galaxies in an annular bin at projected distance R from the central galaxy, and $\bar{\Sigma}_g$ is the mean surface density over the full projected area of the simulation.

Figure 4.1 shows that CS250 reproduces the findings of More et al. (2016) remarkably well. The deviation at large scales ($> 20h^{-1}\,\mathrm{Mpc}$) is expected and reflects a lack of large-scale power in the Millennium Simulation due to its finite size. The offset between the high- and low-concentration subsamples at $R > 3h^{-1}\,\mathrm{Mpc}$ shows that the simulation reproduces the strong assembly bias seen in the SDSS data. On small scales ($< 300h^{-1}\,\mathrm{kpc}$) the number density profile is slightly too steep for the high-concentration clusters, but shows otherwise very good agreement, while there is an offset of 0.1 dex for the low-concentration subsample inside $400h^{-1}\,\mathrm{kpc}$. The most notable differences are on intermediate scales, especially in the range $1h^{-1}\,\mathrm{Mpc} \leq R \leq 3h^{-1}\,\mathrm{Mpc}$ for the low-concentration case. For high-concentration clusters the agreement in this range is excellent and extends out to well beyond $10h^{-1}\,\mathrm{Mpc}$. This is the radial range where splashback features are expected, but is also comparable to the radius, R_c, used operationally to define clusters. These differences are highlighted in the radial variations of the profile slope, which we look at next.

In Figure 4.2 we plot the logarithmic derivative $\mathrm{d}\log\Delta\Sigma_g/\mathrm{d}\log R$ over a restricted radial range for these same two CS250 subsamples, comparing with the same quantity for SDSS clusters as plotted by More et al. (2016). The simulated curves appear noisier than those observed This is at least in part because of the more direct derivative estimator used here. Nevertheless, we reproduce the main features highlighted by More et al. (2016), who identified the position of the minimum of these curves (i.e. the steepest profile slope) as their estimate of the splashback radius. The minima occur at similar radii in the observed and simulated data which, as More et al. (2016) pointed out, are smaller than expected given lensing estimates of cluster mass. Further the minimum is deeper for the high concentration sample and occurs at smaller radius, whereas the opposite is expected from earlier work on the dependence of splashback radius on halo accretion history (and hence concentration, see Diemer and Kravtsov (2014)). In addition, there are clear differences between the observed and simulated curves. In particular, the profiles of simulated low-concentration clusters are clearly shallower than observed in the range $200h^{-1}\,\mathrm{kpc} < R < 1.5h^{-1}\,\mathrm{Mpc}$.

We discuss these features in more detail in Section 4.6, showing them to result from the superposition of effects induced by the cluster selection algorithms on the true splashback

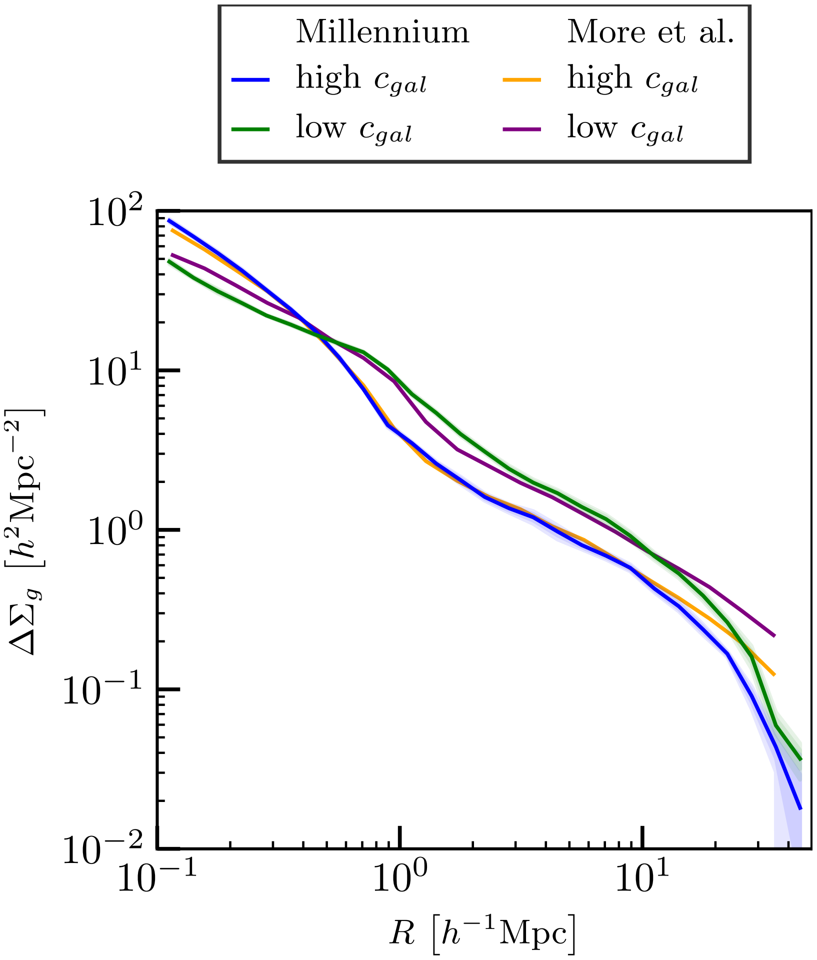

Figure 4.1: Mean surface number density profiles $\Delta\Sigma_g$ for galaxies with $M_i < -20.20$ surrounding clusters in the low- and high-concentration subsamples of CS250 are compared with observational results from More et al. (2016).

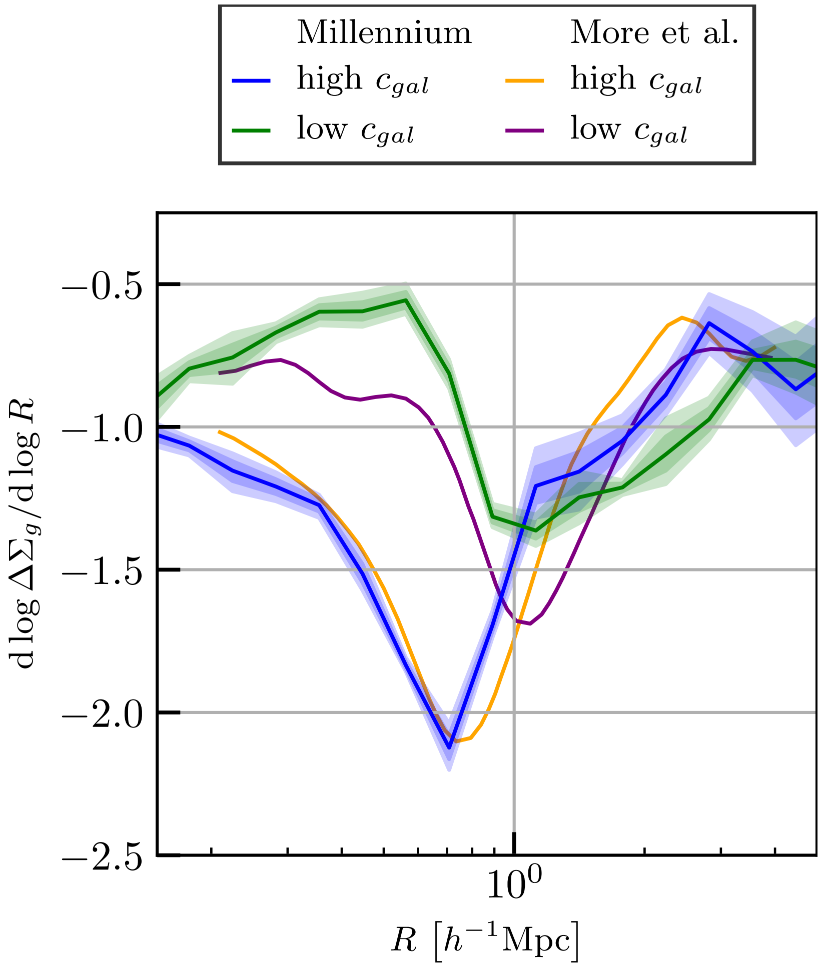

Figure 4.2: The logarithmic derivatives of the $\Delta\Sigma_g$ profiles for CS250 shown in Figure 4.1 are compared with those plotted for their SDSS clusters by More et al. (2016).

signal.

Mean mass density profiles can be computed much more straightforwardly for our simulated cluster samples than is possible observationally, where such profiles are obtained from the correlated orientation of background galaxies induced by gravitational lensing. In order to compare with the lensing results in Miyatake et al. (2016), we bin up the projected mass distribution of the simulation around cluster central galaxies in exact analogy to the above procedure for creating galaxy number density profiles, and we manipulate the resulting profiles to obtain the directly observable quantity,

$$\Delta\Sigma_m(< R) = \Sigma_m(< R) - \Sigma_m(R). \tag{4.4}$$

Here, $\Sigma_m(R)$ is the surface mass density profile analogous to $\Delta\Sigma_g(R)$ above, while $\Sigma_m(< R)$ is the mean of this quantity over projected radii interior to R. Note that despite the similarity in notation (which we have inherited from earlier work) $\Delta\Sigma_m(< R)$ is not directly analogous to $\Delta\Sigma_g(R)$ and will differ from it in shape even if the projected mass and galaxy number density profiles parallel each other exactly.

In Figure 4.3 we compare $\Delta\Sigma_m(< R)$ obtained in this way for the high- and low-concentration subsamples of CS250 to the profiles inferred by Miyatake et al. from their SDSS lensing data. Whereas the observational data show at most small differences between the high- and low-concentration subsamples for $R < 10h^{-1}$ Mpc, our simulated profiles differ significantly in a way which is related to the differences seen in Figure 4.1. Indeed, we have plotted the surface mass density profiles $\Sigma_m(R)$ directly, and find they are very similar in shape and relative amplitude to the simulated galaxy surface density profiles of Figure 4.1. We note that the disagreement between simulation and observation is limited to low-concentration clusters – agreement is very good for the high-concentration systems on all scales below about $15h^{-1}$ Mpc. We have found no explanation for this discrepancy. The uncertainties on the $\Delta\Sigma_m(< R)$ inferred from lensing data are much larger than the purely statistical uncertainty in the simulation results, but below $1h^{-1}$ Mpc the simulation results for low-concentration clusters lie systematically below the observations, while beyond $3h^{-1}$ Mpc they tend to lie above them. (Note that the coloured bands in Figure 4.3 show the estimated 1σ uncertainties in the observations.) This disagreement is in line with the stronger differences between the projected galaxy profiles for the low-concentration subsample. Our findings for the differences in the inner part are close to the findings of Dvornik et al. (2017) who recently investigated the mass profiles of galaxy groups. These less massive objects were identified with a different group finder (based on the FoF algorithm), but the same c_{gal} projected concentration measure was used to divide the sample. While they found a similar split at small scales in the lensing profiles, they did not see a significant signal of assembly bias on large scales. This is expected around the masses of groups when splitting by concentration.

Miyatake et al. (2016) inferred almost equal mean total masses, $M_{200m} \sim 2\times10^{14}h^{-1}M_\odot$, for high- low-concentration clusters from their measured $\Delta\Sigma_m(< R)$ profiles. Processed in the same way, our simulated profile for high-concentration clusters would give a very similar answer, whereas that for low-concentration clusters would give a lower value by a few tens of percent. (For $M_{200m} = 2\times10^{14}h^{-1}M_\odot$, $R_{200m} = 1.5h^{-1}$ Mpc, in the middle of the range where simulated and observed $\Delta\Sigma_m(< R)$ agree best.) Thus the overall mass-scale of the clusters identified in the Guo et al. (2011) galaxy catalogues by our redMaPPer-like algorithm is close to that of the SDSS clusters studied by Miyatake et al. (2016) and More et al. (2016).

4.3.2 The influence of cluster selection depth

The simulation results shown in the last section all referred to CS250 for which any red galaxy projected within R_c is considered a potential cluster member, regardless of its distance in depth ("redshift") from the central galaxy. As noted previously, Rykoff et al. (2014) estimate the 1σ uncertainty in the photometric redshift of an individual red SDSS member galaxy to increase with distance and to be $90h^{-1}$ Mpc at $z = 0.25$, the median redshift of the SDSS cluster sample. Thus many of the observed clusters may be better

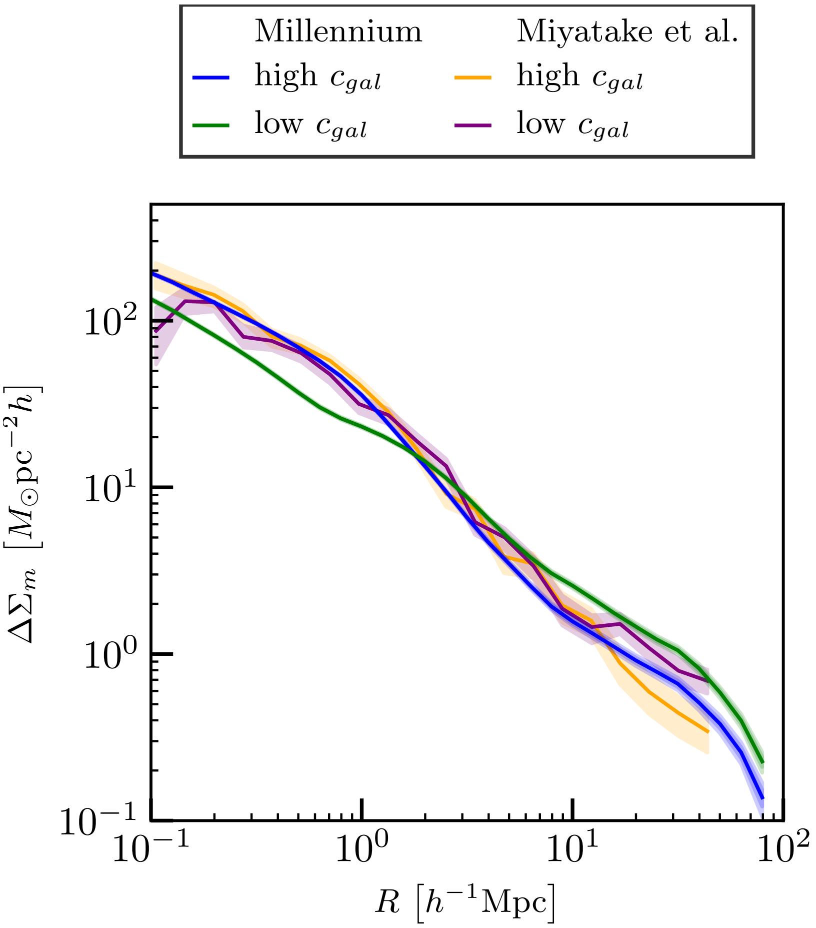

Figure 4.3: The mean lensing observable $\Delta\Sigma_m$for high- and low-concentration clusters in the CS250 sample is compared to observational results for SDSS clusters from Miyatake et al. (2016).

localised in depth than in the CS250 catalogue. In this section we compare galaxy and mass profiles among our three simulated cluster catalogues, CS250, CS120 and CS60, for which the depth selection parameter $\Delta z_m = 250, 120$ and $60h^{-1}\,$Mpc, respectively. This allows us to assess how strongly the effective selection depth of clusters affects their apparent splashback and assembly bias signals. We find that effects are small on the former, but can be substantial on the latter.

Figure 4.4 shows the overall shape of the projected galaxy number density profiles to be very similar in the three cluster catalogues. The high concentration profiles differ from each other by at most 10 per cent within R_c and remain within the same bound out to $\sim 20h^{-1}\,$Mpc. Beyond this point the uncertainties increase drastically and the ratios of the profiles with smaller Δz_m quickly depart from unity but stay within a less than the 68-percentile of the bootstrap distribution of it. The variation is somewhat smaller for low-concentration clusters and is also below 10 per cent within R_c, but also below 25 per cent all the way out $\sim 30h^{-1}\,$Mpc. Beyond R_c the profile amplitude of low-concentration clusters decreases with decreasing Δz_m at all separations where it is reliably determined.

This level of agreement is such that all three catalogues agree almost equally well with observation. In the profiles themselves, systematic differences only start to become noticeable outside R_c and the largest effect is the shift in the large-scale amplitude of the profile for the low-concentration clusters, which, as we will see below (in Section 4.3.3) is enough to affect the apparent level of assembly bias significantly. At the intermediate radii relevant for splashback detection, the profile shapes are sufficiently similar that curves like those of Figure 4.2 show almost no dependence on Δz_m.

The $\Delta\Sigma_m$profiles (shown in Figure 4.5) also vary only slightly as a function of effective cluster depth, Δz_m, with shifts of similar amplitude to those seen in the projected galaxy number density profiles. For high-concentration clusters these are even smaller than for the previous case, while for low-concentration clusters they are larger within R_c and have the effect of increasingly smoothing the sudden changes in slope seen in the CS250 profile as Δz_m decreases. For both cases the amplitude of the profiles on large scales is decreased for smaller Δz_m, though by less than 25 per cent out to $\sim 50h^{-1}\,$Mpc.

4.3.3 Profile ratios and assembly bias

By taking the ratio of the profiles discussed in the previous section we can obtain a measure of the relative bias of high- and low-concentration clusters at fixed cluster richness, hence of *assembly bias*. In Figure 4.6 we show this ratio for the $\Delta\Sigma_g$ profiles as a function of projected separation for our three catalogues of simulated clusters. In order to measure the large-scale bias, More et al. (2016) only plotted this ratio at $R \geq 3h^{-1}\,$Mpc (the orange points with error bars in Figure 4.6). However, since they give the individual profiles for high- and low-concentration clusters, it is straightforward to reconstruct the observed ratio on smaller scale. We show this as a dashed orange line in Figure 4.6.

The observed and the simulated ratios show similar behaviour which separates into three distinct radial regimes. At $R \geq 3h^{-1}\,$Mpc, the relative bias varies little and the observed value of 1.48 ± 0.07 matches very well that for CS250 outside of $R = 8h^{-1}\,$Mpc. CS120 gives

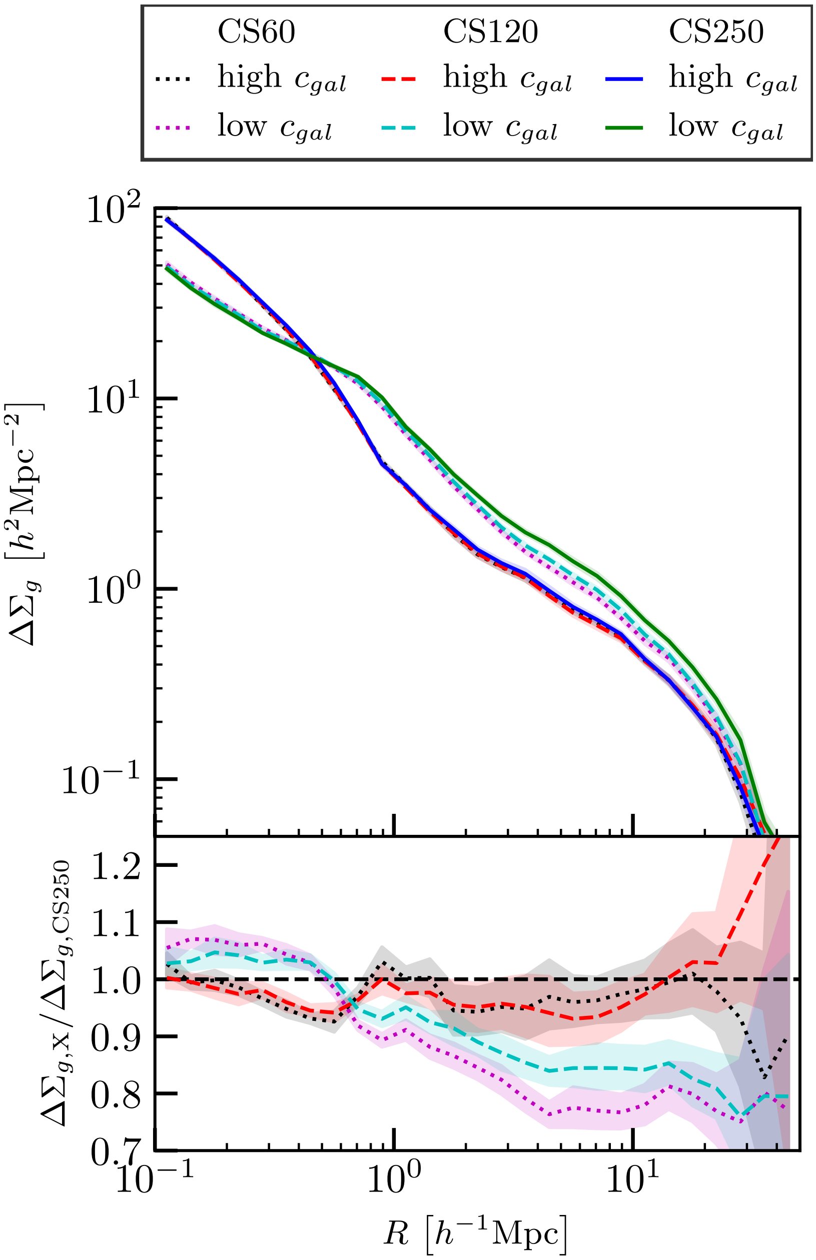

Figure 4.4: Comparison of the $\Delta\Sigma_g$profiles for the high c_{gal} and low c_{gal} subsamples of our three simulated cluster catalogues (upper panel) and ratios of the profile amplitudes for CS120 and CS60 to that for CS250 (lower panel).

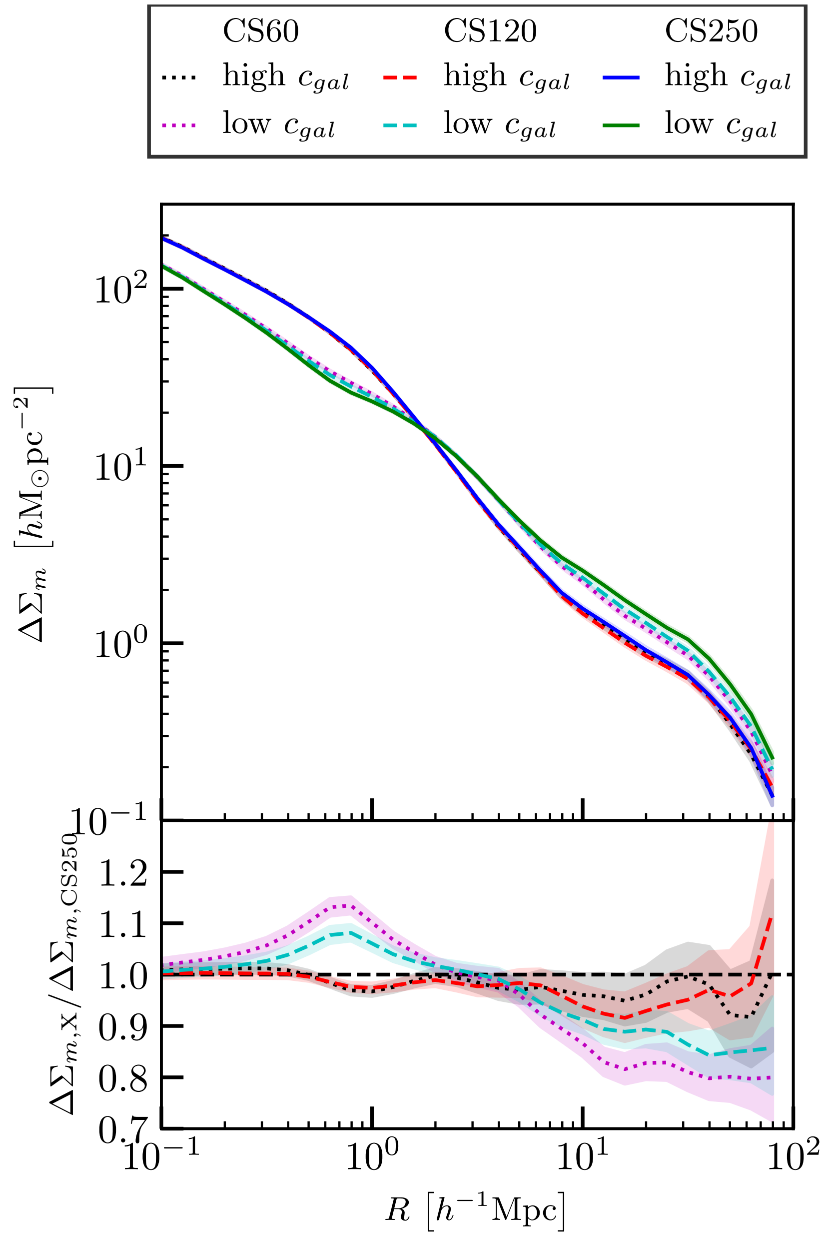

Figure 4.5: Comparison of the $\Delta\Sigma_m$profiles for the high c_{gal} and low c_{gal} subsamples of our three simulated cluster catalogues (upper panel) and ratios of the profile amplitudes for CS120 and CS60 to that for CS250 (lower panel).

a somewhat smaller value fitting the observations well between 3 and $10h^{-1}$ Mpc, while at larger R it is still within about 1σ. CS60 has even weaker relative bias barely within 1σ. Both these signals appear to decline with increasing R. The behaviour at smaller scales differs markedly on either side of a sharp peak which, for the simulated clusters, occurs almost exactly at $\langle R_c \rangle \sim 1h^{-1}$ Mpc, coinciding with that for the observed clusters. At smaller R, the ratio of the profiles increases smoothly and strongly with R, reflecting the requirement that the two cluster subsamples have similar richness but systematically different values of $\langle R_{\text{mem}} \rangle$. This also enforces a ratio substantially above unity at $R = R_c$. At intermediate radii, $R_c < R < 3h^{-1}$ Mpc, the ratio has to decline from the high value at the peak to the more modest value characteristic of the large-scale assembly bias. In all three samples there is a noticeable change in slope just outside $2h^{-1}$ Mpc which appears to reflect true splashback effects (see Section 4.4.2).

These properties demonstrate that the operational definition of clusters has a substantial effect on the ratio of the profiles out to at least $3h^{-1}$ Mpc. These effects must therefore be present also in the individual profiles, and hence must affect their use for identifying splashback features. In addition, the variation of the ratios at large R among our three cluster catalogues shows that the apparent assembly bias signal is significantly affected by projection effects.

The ratio of the $\Delta\Sigma_m$profiles for the high- and low concentration subsamples of each of our three simulated cluster catalogues are shown in Figure 4.7 in exactly analogous format to Figure 4.6. They are compared to observational results taken directly from Miyatake et al. (2016). The difference in shape between the simulation curves in Figures 4.7 and 4.6 is due primarily to the conversion of $\Sigma_m(R)$ to $\Delta\Sigma_m(< R)$. A ratio plot constructed using $\Sigma_m(R)$ directly is quite similar to Figure 4.6, although the peak at $\langle R_c \rangle$ is less sharply defined. The behaviour of the observational points in Figure 4.7 is quite erratic and looks rather implausible when compared with the smooth variation predicted by the simulation. Over the ranges $3h^{-1}$ Mpc $< R < 14h^{-1}$ Mpc and $R > 15h^{-1}$ Mpc the predicted assembly bias signal is almost constant, but over the first range it is much larger than and apparently inconsistent with that observed, whereas over the second it is smaller than and again apparently inconsistent with that observed. It is our impression that the uncertainties of these observational points are too large for secure interpretation to be possible.

The differences in large-scale assembly bias between our three simulated cluster catalogues are similar to those seen for the cluster number density profiles of Figure 4.6, although pushed out to systematically larger radii. Again this is a consequence of the conversion from $\Sigma_m(R)$ to $\Delta\Sigma_m(< R)$. On small scales the simulation curves lie well below the observational points. This is a restatement of the fact that the simulated profiles in Figure 4.3 differ much more at these radii than the observed profiles.

4.4 The 3D Perspective

Miyatake et al. (2016) and More et al. (2016) interpret their SDSS results under the implicit assumption that the features seen in the stacked 2D profiles correspond to similar features in

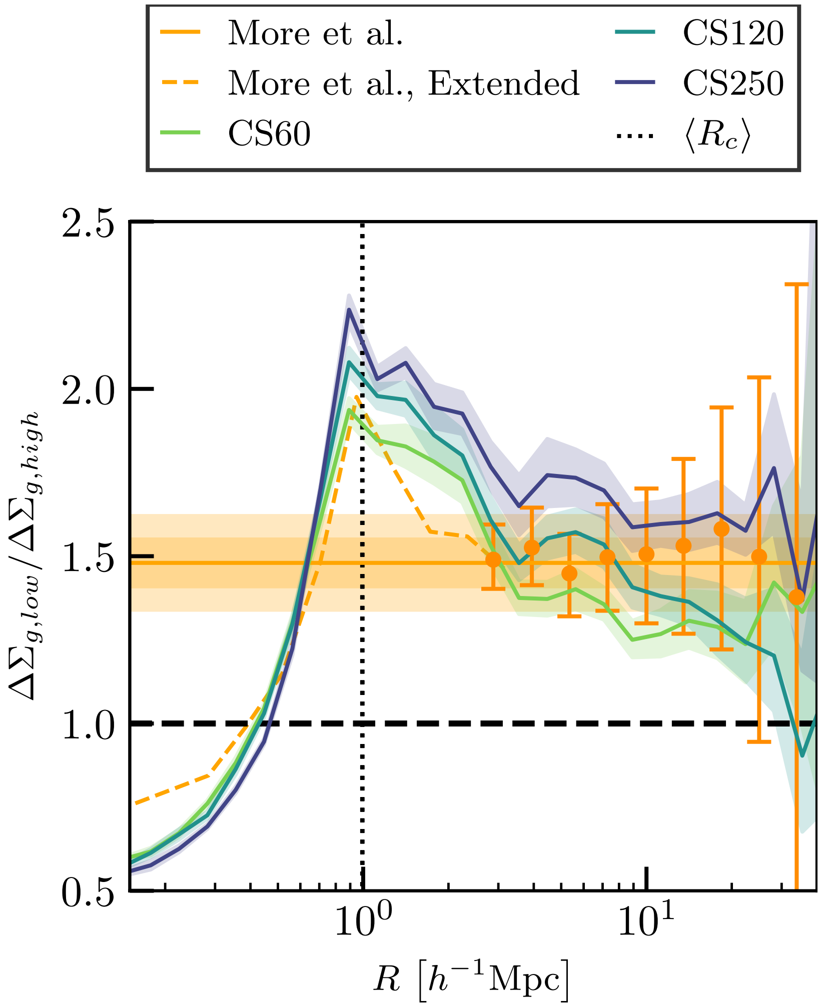

Figure 4.6: The ratio of the projected galaxy number density profiles of the low c_{gal} and high c_{gal} subsamples of our three simulated cluster catalogues (solid lines surrounded by their 68 per cent confidence regions). Points with error bars are observational data taken directly from More et al. (2016), while the continuation of these data to smaller scales (the dashed orange line) was calculated from the individual profiles in their paper. The dotted vertical line indicates $\langle R_c \rangle$ for the simulated clusters. The horizontal orange band is the observed assembly bias signal quoted by More et al. (2016) with its 68 and 95 per cent confidence ranges.

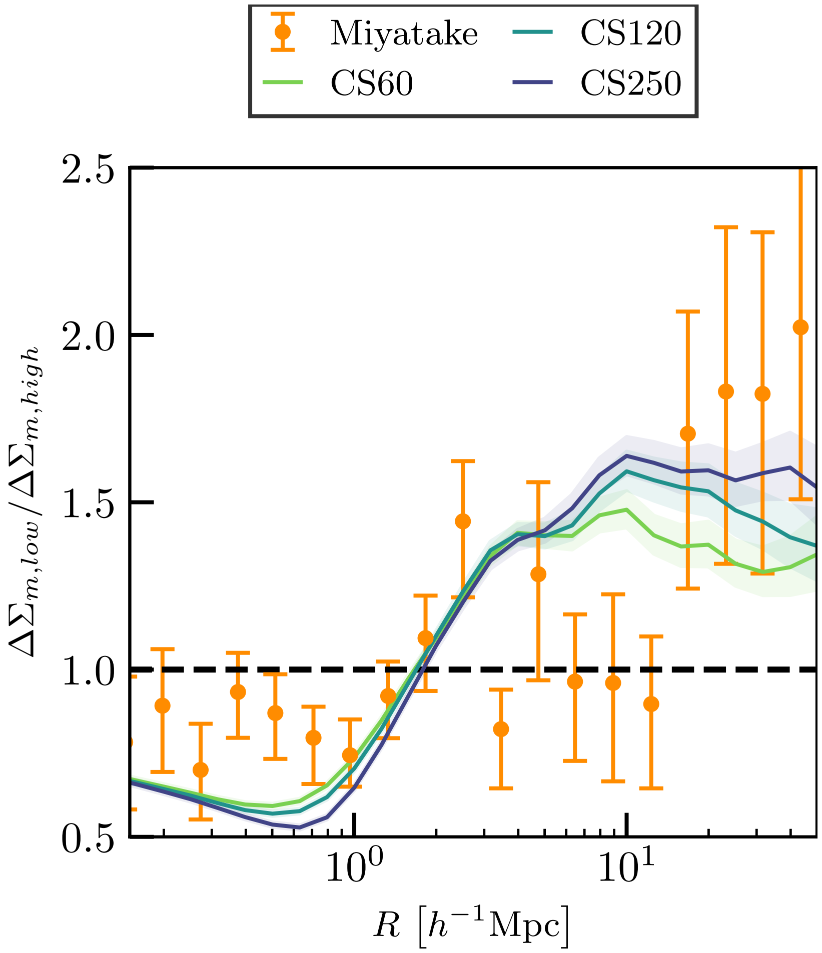

Figure 4.7: Ratios of $\Delta\Sigma_m$for the high- and low-concentration subsamples of our three cluster catalogues (solid lines with their 68 per cent confidence ranges). Points with error bars are results derived from the gravitational lensing signal of SDSS clusters by Miyatake et al. (2016).

the 'true' 3D profiles. In our simulations, it is possible to test the extent to which this is the case, so in this section we compute stacked 3D profiles of mass density and of galaxy number density around the central galaxies of our three cluster catalogues, splitting them into high- and low-concentration subsamples as before using the 2D values of $c_{\rm gal} = R_c(\lambda)/\langle R_{\rm mem}\rangle$. This allows us to make plots directly analogous to those discussed above, and so to check the 2D – 3D correspondence. In this section all profiles are calculated in true position space rather than in redshift space. Note that we here use a standard definition of the spherically averaged mass density profile rather than some 3D analogue of $\Delta\Sigma_m$. Note also that since each central galaxy can appear in one to three different projections, we give it the corresponding weight when constructing the 3D profiles in order to keep as close a correspondence as possible to the 2D results discussed previously.

4.4.1 Splashback Radius

As was the case in 2D, we find that plots of the 3D profile slope, analogous to those of Figure 4.2, are very similar for our three cluster catalogues. In Figures 4.8 and 4.9 we therefore show results for CS250 only. Since recent theoretical work on splashback properties has concentrated on cluster mass profiles (e.g. Diemer and Kravtsov, 2014, hereafter DK14), we start with a discussion of Figure 4.8 which shows logarithmic slope (referred to as γ below) as a function of 3D radius r .

These slope profiles show relatively smooth behaviour with well-defined minima at $r \sim 1.8h^{-1}\,$Mpc. The mean M_{200m} values in the two sub-samples correspond to $R_{200m} \sim 1.45h^{-1}\,$Mpc and $R_{200m} \sim 1.37h^{-1}\,$Mpc, so these minima occur at $1.2R_{200m}$ and $1.3R_{200m}$ for the high- and low-concentration samples, respectively. These values are very close to the expected values given in More et al. (2015) for the expected mass accretion rates at the given masses and redshift. The slopes at minimum are significantly shallower for our stacks ($\gamma \sim -2.8$) than DK14 found for halos of similar mass ($\gamma \sim -3.5$). As shown in the Appendix, this is because such profiles depend both on the definition of the sample to be stacked and on the details of stack construction. In particular, DK14 scale each individual profile to its own R_{200m} and then take the median density at each r/R_{200m}, whereas we take the mean density at each radius directly. The DK14 procedure typically produces deeper and sharper minima, hence better defined splashback radii which occur at slightly smaller radii, but it is not easily implemented on observed samples. For example, the redMaPPer samples are defined to have similar (and known) values of R_c but their individual values of R_{200m} are unknown. In addition, weak lensing reconstructions of the mass distribution naturally produce mean rather than median mass profiles.

The two slope profiles of Figure 4.8 differ significantly in shape. In the inner regions ($r < R_c$) this reflects the fact that the two samples are separated by galaxy concentration (in practice, by $\langle R_{\rm mem}\rangle/R_c$) so that, by definition, the low-concentration clusters have shallower 2D galaxy density profiles within R_c than the high-concentration clusters. Figure 4.9 shows that this requirement carries over to the 3D galaxy profiles, and it is still very visible in Figure 4.8. Similar effects are seen in Figure 14 of DK14 where they split their halo sample by 3D mass concentration. However, our results do not agree with the trend

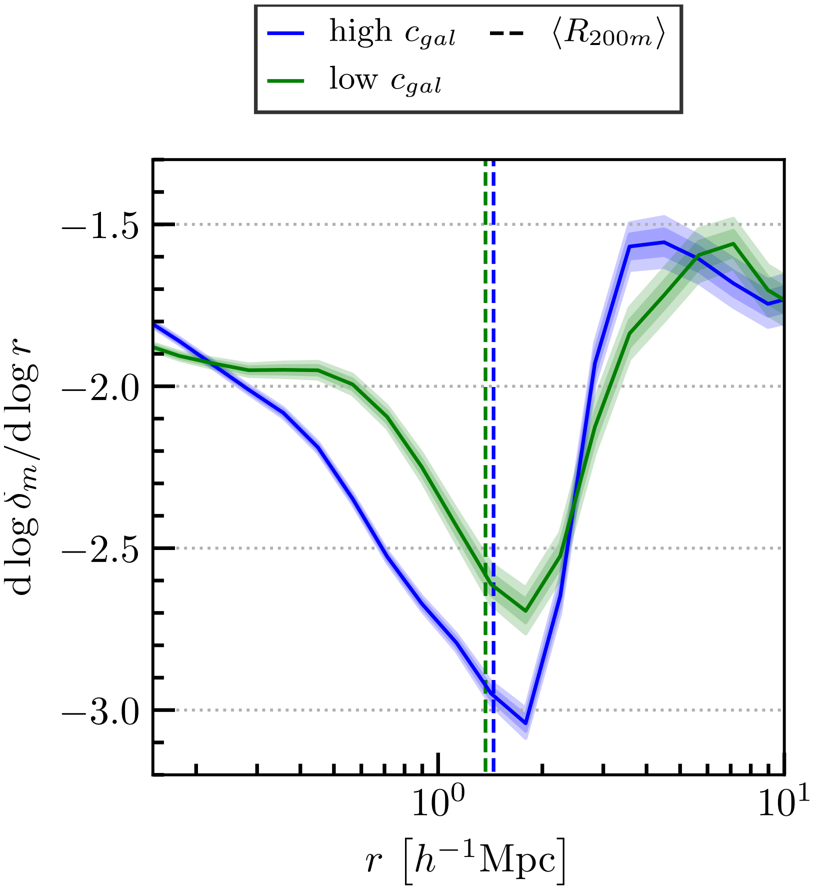

Figure 4.8: Logarithmic derivative profiles of the 3D mass overdensity around the central galaxies of the high- and low-concentration subsamples of CS250. Vertical lines mark the R_{200m} values for the two samples calculated directly from their stacked mass profiles.

they find for more concentrated clusters to have a shallower minimum slope and a larger splashback radius. We have checked that if we follow their scaling and median stacking procedures, our high-concentration clusters still have a steeper minimum slope and the same splashback radius as our low-concentration clusters. The discrepancy must reflect the difference between selecting halos by 3D mass and mass concentration and selecting clusters by 2D richness and galaxy concentration.

The shapes of the 3D slope profiles for the mass (Figure 4.8) and for the galaxies (Figure 4.9) are very similar, in particular, beyond the splashback minimum. At smaller radii the features induced by cluster selection are stronger in the galaxy profile, with a secondary minimum just inside $\langle R_c \rangle$ which is just visible as a slight inflection in the mass profile. Overall, however, the features in the galaxy profile are much less dramatic than in its 2D analogue, Figure 4.2. This just reflects the fact that clusters were selected and their concentrations estimated using the 2D data

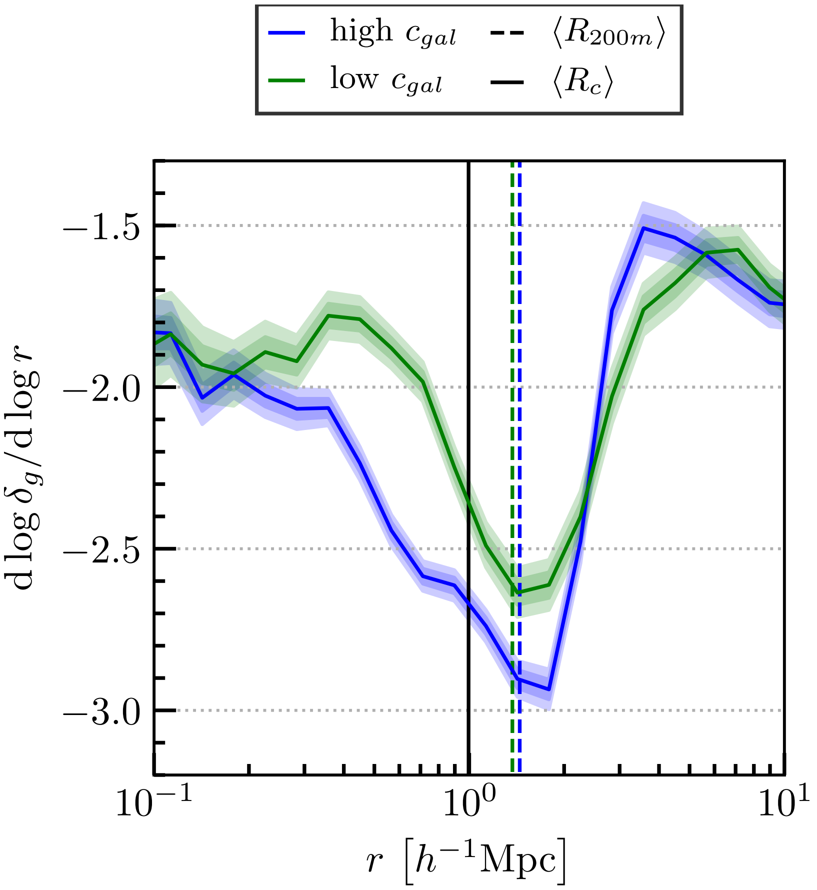

Figure 4.9: Logarithmic derivative profiles of the 3D galaxy number overdensity around the central galaxies of the high- and low-concentration subsamples of CS250 in identical format to Figure 4.8 except that a solid vertical line indicates $\langle R_c \rangle$ for the two samples.

4.4.2 Large-scale environment

We now look at the ratios of stacked 3D mass overdensity profiles for our low- and high-concentration clusters, and at the corresponding ratios of their galaxy number overdensity profiles. These are directly analogous to the ratios of 2D galaxy number overdensity profiles shown in Figure 4.6. As in that figure, we here compare results for the three samples, CS60, CS120 and CS250. Ratios as a function of r are shown for mass overdensities in Figure 4.10 and for galaxy number overdensities in Figure 4.11. The shapes of the curves and their relative positions for the three samples are very similar in these two figures.

In the inner regions, $r < R_c$, all curves are rapidly and smoothly rising, showing that the difference in 2D galaxy profiles resulting from our classification by concentration carries over to the 3D galaxy and mass profiles. In this regime and in both plots the ratio for CS60 is slightly larger than that for CS120 and significantly larger than that for CS250. This behaviour mirrors that of the ratio of the fractions of 2D potential members which are part of the central galaxy's FoF group (see Table 4.3). Interestingly, this ranking of amplitudes for the three samples persists to much larger scales and is opposite to that seen in 2D (Figure 4.6). Clearly, with increasing Δz_m, projection effects contribute more strongly to low- than to high-concentration clusters not only at $R \sim R_c$ but also at much larger projected separation.

In the range $R_c < r < 5h^{-1}$ Mpc, all curves continue to rise to a sharp peak before dropping again to a value which remains approximately constant over the interval $5h^{-1}$ Mpc $< r < 30h^{-1}$ Mpc. The peak corresponds to the crossing of the derivative curves for the low- and high-concentration subsamples in Figures 4.8 and 4.9. It thus reflects differences in the way the splashback feature merges into larger scale structure in the two cases. As noted above, it appears to be visible as a sharp change in slope in the profiles of Figure 4.6 (see also Figure 4.15 below). Between R_c and the peak, effects from sample definition clearly modulate galaxy overdensity profile ratios more strongly than mass overdensity profile ratios but the difference is quite small.

The constant profile ratios seen over the range $5h^{-1}$ Mpc $< r < 30h^{-1}$ Mpc are a direct measurement of the 3D assembly bias for cluster samples split by 2D concentration. These values are significantly smaller than the 2D values inferred from Figure 4.6. In addition, they rank in the opposite sense with Δz_m, they are consistent between Figures 4.8 and 4.9, and they are similar to the values expected from previous work on assembly bias for cluster mass haloes split by concentration (e.g. More et al., 2016). As we will see in the next section, a clue to the origin of this difference between the 2D and 3D estimates of assembly bias comes from the largest r bins in these figures where, although noisy, the ratios of the profiles rise to large values.

4.5 Projection contamination

In the preceding sections we found a number of differences in the apparent splashback and assembly bias signals between the 2D and the 3D profiles of our simulated galaxy clusters.

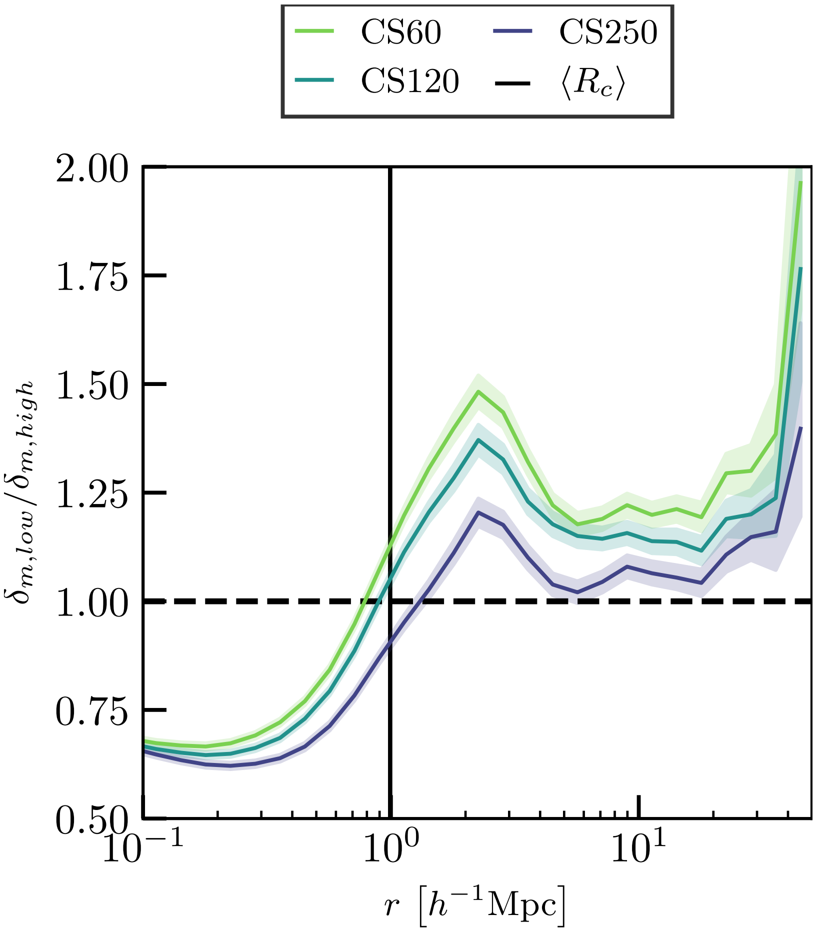

Figure 4.10: Ratios of the 3D mass overdensity profiles of low- and high-concentration clusters for each of our three cluster samples. The vertical line indicates the mean cluster radius $\langle R_c \rangle$.

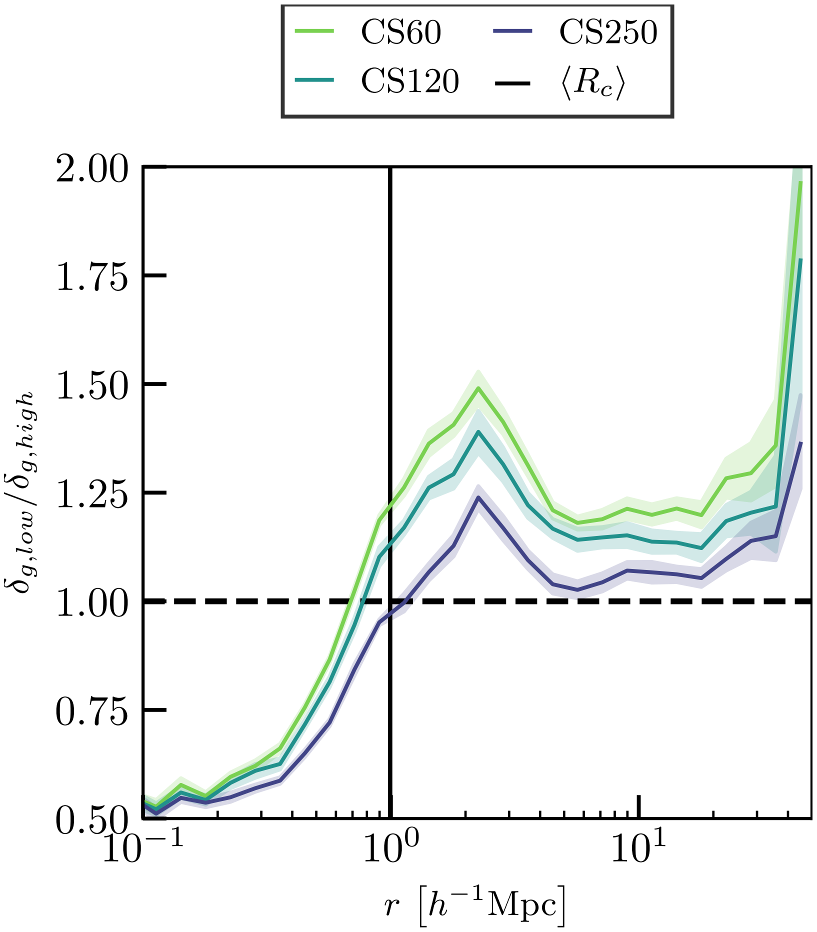

Figure 4.11: Ratios of the 3D galaxy number overdensity profiles of low- and high-concentration clusters for each of our three cluster samples with a vertical line indicating the mean cluster radius $\langle R_c \rangle$.

These differences are present both in the mass and in the galaxy number density profiles, and they affect the low- and high-concentration subsamples to differing degrees. In this section we focus specifically on galaxy number density profiles, compiling them in the two dimensions of projected separation and line-of-sight depth so that we can compare results for the two subsamples and isolate the distribution in depth of the galaxies which give rise to the difference in projected profiles.

Let R, as above, denote projected separation, and $q > 0$ denote line-of-sight separation, measured either in configuration space ($q = |d|$) or in redshift space ($q = |\pi|$). We define a set of cells of constant width in $\ln R$ and $\ln q$ and compile galaxy counts in these cells around the central galaxies of the low- and high-concentration subsamples of each of our cluster samples, $N_{lo}(R, q)$ and $N_{hi}(R, q)$ respectively.

In Figures 4.12 and 4.13 we show the quantity

$$\beta(R, q) = \frac{N_{lo}(R, q) - N_{hi}(R, q)}{\sum_q [N_{lo}(R, q) + N_{hi}(R, q) - N_c n_{gal} V(R, q)]}, \tag{4.5}$$

for the real-space and redshift space cases respectively. In this equation, N_c is the total number of clusters in the sample, n_{gal} is the mean space density of galaxies, and $V(R, q)$ is the volume of the cell at (R, q). Thus $2\sum_q \beta(R, q) = b_{lo}(R) - b_{hi}(R)$, where the assembly bias factors b_{lo} and b_{hi} are the ratios of the stacked 2D galaxy number overdensity profiles of the low- and high-concentration subsamples to that of the cluster sample as a whole. The distribution of β over q at fixed R thus indicates the distribution in depth of the difference in galaxy counts which gives rise to the apparent 2D assembly bias signal.

In the inner regions ($R < 400h^{-1}\,\mathrm{kpc}$) the projected profile of high c_{gal} clusters lies above that of low c_{gal} clusters for all three samples (see Figure 4.6). Figure 4.12 shows that, as expected, the additional galaxies which produce this excess lie in the inner regions of the clusters, with a median depth offset from the central galaxy of $150h^{-1}\,\mathrm{kpc}$ or less. In redshift space, the random motions within clusters move this excess out to $|\pi| \sim 700\,\mathrm{km\,s^{-1}}$, as shown in Figure 4.13.

Beyond $R = 400h^{-1}\,\mathrm{kpc}$ the behaviour switches and the projected profile of low c_{gal} clusters lies above that of high c_{gal} clusters (again see Figure 4.6). The galaxies which produce this excess lie in two different ranges of depth whose relative contribution varies both with R and with Δz_m. At $R < 2h^{-1}\,\mathrm{Mpc}$, a 'local' component centred near $R \sim |d| \sim \langle R_c \rangle$ contributes most of the excess low c_{gal} counts in CS60, about half of them in CS120, and a minority of them in CS250, producing much of the pronounced peak seen at these R in the profile ratios of Figure 4.6. A second component, distributed relatively uniformly over $\pm\Delta z_m$, the full allowed depth for potential cluster members, contributes excess counts to the low c_{gal} cluster profiles at all $R > R_c$ and is responsible for most of the large-scale assembly bias. It also dominates the excess counts near $\langle R_c \rangle$ in CS250. The systematic change in the relative weight of these two components with increasing R results in a shift in the median depth offset of the excess counts, indicated by the black solid lines in Figures 4.12 and 4.13. The increasing strength of the second component from CS60 to CS120 to CS250 is the cause of the increase in 2D assembly bias with Δz_m. Figure 4.13

shows that redshift space distortions significantly smear out these two components and make them more difficult to distinguish.

These results explain why strong assembly bias is seen in 2D for CS250 and CS120 (see Figure 4.6) but only a much weaker signal is seen in 3D (Figure 4.11). Many of the low-concentration clusters in these samples have significant foreground/background groups projected on their outer regions. These groups are distributed over the full depth $\pm\Delta z_m$, and are visible in Figures 4.12 and 4.13 as an excess in bins at large q and $R \sim R_c$. Galaxies correlated with these foreground/background groups then produce excess galaxy counts at similar q for all R values shown in the plot. Since the fall-off in these counts with R at the q of the background group is similar to that of galaxy counts at relatively small q correlated with the primary cluster, the induced apparent assembly bias is almost independent of R. The rise in 3D assembly bias seen at the largest r in Figure 4.11 is a result of beginning to pick up this additional correlated component in the counts around low-concentration clusters.

The strength of this effect clearly depends on the sensitivity of the cluster identification algorithm to projection effects at $R \sim R_c$. This in turn depends both on the effective Δz_m and on the weight assigned to potential members near the cluster edge. Hence, the apparent bias may differ between the real redMaPPer sample and our simulated samples. Nevertheless, the strong similarity seen in previous sections between the behaviour of our CS250 and CS120 samples and the SDSS sample analysed by More et al. (2016) and Miyatake et al. (2016) suggests that the assembly bias signal they found has a similar origin to that in the simulation. In the next section we will explore further the dependence of apparent splashback features on cluster definition and argue that the unexpected properties of the features detected by More et al. (2016) are a result of confusion with features imposed by the cluster selection procedure.

4.6 Cluster definition affects profile shape

We have argued above that the details of our redMaPPer-like algorithm leave an imprint on the stacked profiles of our simulated clusters. Although this is most evident in the strong peak at R_c in the profile ratios of Figure 4.6 and in the steep gradient interior to this radius induced by our separation of the two subsamples by concentration, c_{gal}, it is also visible in the crossing at R_c of the individual gradient profiles of Figure 4.2 and in their minima close to and on opposite sides of this radius. In this section we investigate these effects further by varying the value of R_c used to define clusters. Specifically, we set

$$R_c = 1.0\eta \left(\frac{\lambda}{\lambda_n(\eta)}\right)^{0.2} h^{-1}\,\mathrm{Mpc} \tag{4.6}$$

and we change η.

The variable normalisation $\lambda_n(\eta)$ in Equation 4.6 accounts for the fact that a given cluster will contain more galaxies within a larger projected radius. In the following we will consider $\eta = 2/3, 1$ (the value used in all earlier sections) and $3/2$. Based on the mean

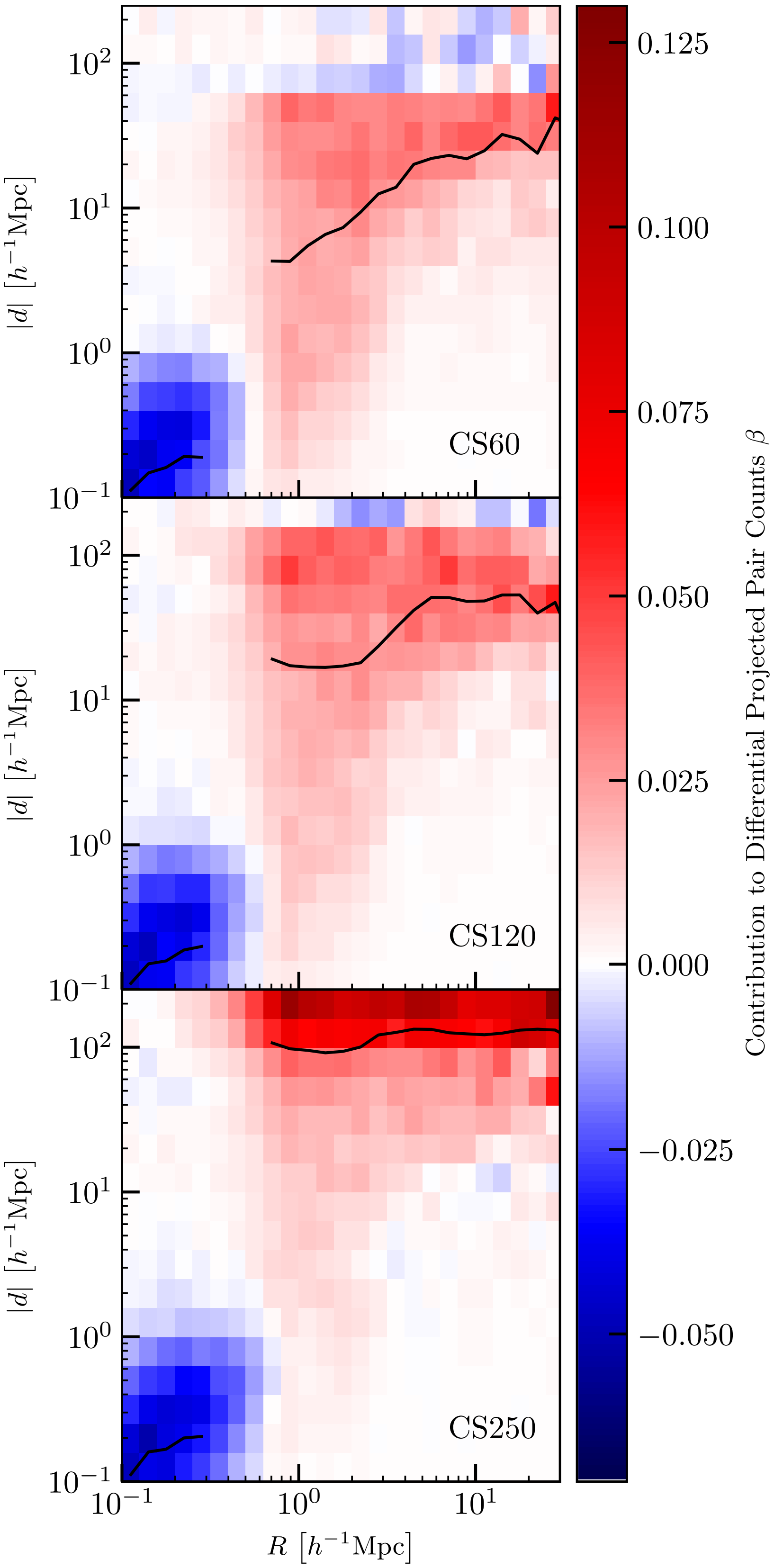

Figure 4.12: The quantity $\beta(R, q)$ of Equation 4.5 for the case $q = |d|$. This shows the distribution over depth q of the fractional difference between the projected galaxy count profiles of the low c_{gal} and high c_{gal} subsets of each of our three simulated cluster samples. The black curves give the median offset in depth of the excess counts as a function of R.

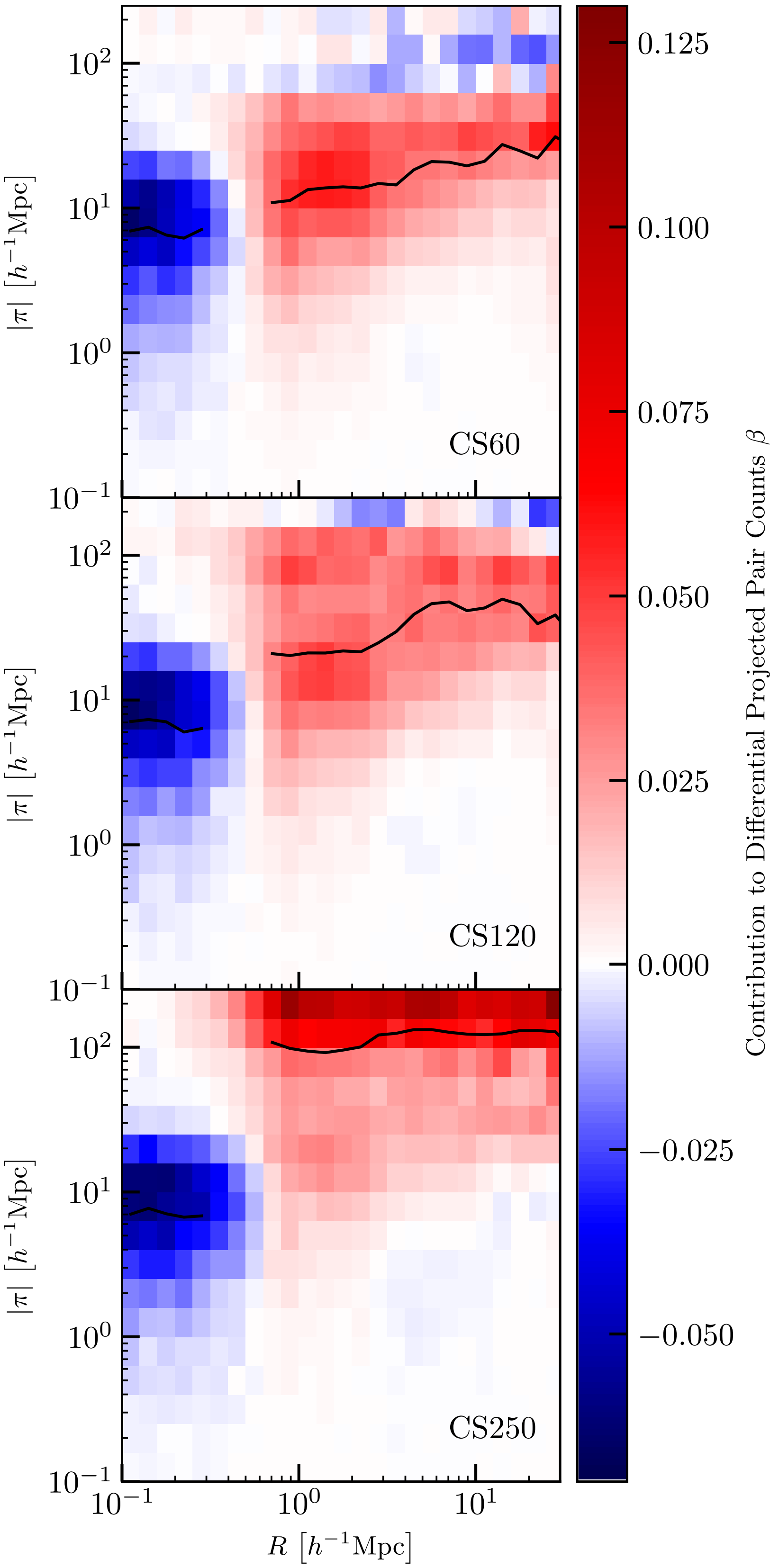

Figure 4.13: Identical to Figure 4.12 except for the redshift space case, $q = |\pi|$.

galaxy number overdensity stacks of Section 4.3.3, we take $\lambda_n\left(\eta = \frac{2}{3}\right) = 74$, $\lambda_n(1) = 100$, as before, and $\lambda_n\left(\eta = \frac{3}{2}\right) = 130$. For each choice of η we repeat the cluster selection and concentration calculation procedures of Sections 4.2.2.1 and 4.2.2.2. Since changing R_c changes the richness value λ assigned to each cluster, we shift the richness range defining our samples ($20 \leq \lambda \leq 100$ for $\eta = 1$) so that the total numbers of simulated clusters above the upper and lower limits remain unchanged. In the following we show results for $\Delta z_m = 250h^{-1}\,$Mpc only, since the two other cases behave very similarly.

Figure 4.14 repeats the observational and CS250 results from Figure 4.6 and compares them with predictions for $\eta = 2/3$ and $3/2$. The peak of the profile ratio increases strongly with η and shifts to match $\langle R_c \rangle$ in all three cases. Interestingly, the profile ratio for $\eta = 2/3$ peaks at a value of 1.8 at a radius where it is 0.8 for $\eta = 3/2$, and the ratio is unity for $\eta = 2/3$ at a radius where it is only 0.6 for $\eta = 3/2$. Thus, changing the limiting radius defining a cluster sample not only affects its stacked profiles in their outer parts, but also close to the centre. Beyond R_c, the secondary feature noted in Section 4.3.3 and apparently associated with true splashback effects is clearest for $\eta = 2/3$ and is very weak for $\eta = 3/2$. At large R, the strength of assembly bias increases noticeably with η. The stronger peak, the weaker splashback signal and the stronger large-scale assembly bias found with increasing η are all consistent with the expectation that projection effects should increase in importance when clusters are identified within larger radii, hence at lower projected overdensities. Also as expected, overall the SDSS results of More et al. (2016) behave most similarly to the $\eta = 1$ curves in Figure 4.14. Nevertheless the large scale ratios agree equally well with the ones using $\eta = 3/2$.

As shown in Figure 4.15, the logarithmic derivative of $\Delta\Sigma_g$ shows a strong and complex response to η. The middle panel here is essentially a repeat of Figure 4.2, while the upper and lower panels show similar plots for $\eta = 2/3$ and $\eta = 3/2$ respectively. A striking feature of these plots is that the slope profiles for the two subsamples always cross around $R = \langle R_c \rangle$ and at a value of about -1.4. The crossing 'coincidence' is mathematically equivalent to the fact that all the profile ratios have a maximum at $R \sim R_c$ in Figure 4.14, which itself is easily understood as a consequence our creating subsamples with identical distributions of λ but disjoint distributions of $c_{\rm gal}$, thus forcing the profile ratio to increase over the range $0 < R < \langle R_c \rangle$. The uniform slope value at curve crossing reflects the fact that this value equals the slope for the sample as a whole, which is quite slowly varying and close to -1.4 at these projected radii.

Within the crossing point, the slope for low-concentration clusters rises rapidly to a maximum of about $\gamma = -0.5$ at $R \sim \langle R_c \rangle$, while the slope for the high-concentration clusters drops to a minimum at approximately the same radius but with a value which decreases strongly with increasing η. This behaviour is clearly a consequence of our definition of $c_{\rm gal}$ and our separation of clusters into subsamples according its value. On larger scale, the slope profiles appear independent of η when R exceeds twice the largest value of $\langle R_c \rangle$ for the samples being compared. However, the curves for high- and low-concentration clusters differ both from each other and from those of More et al. (2016) in this regime. In the intermediate range, $\langle R_c \rangle < R < 2\langle R_c \rangle$, the shape of the curves is set by the need to interpolate between these two different behaviours, causing a minimum at or just outside

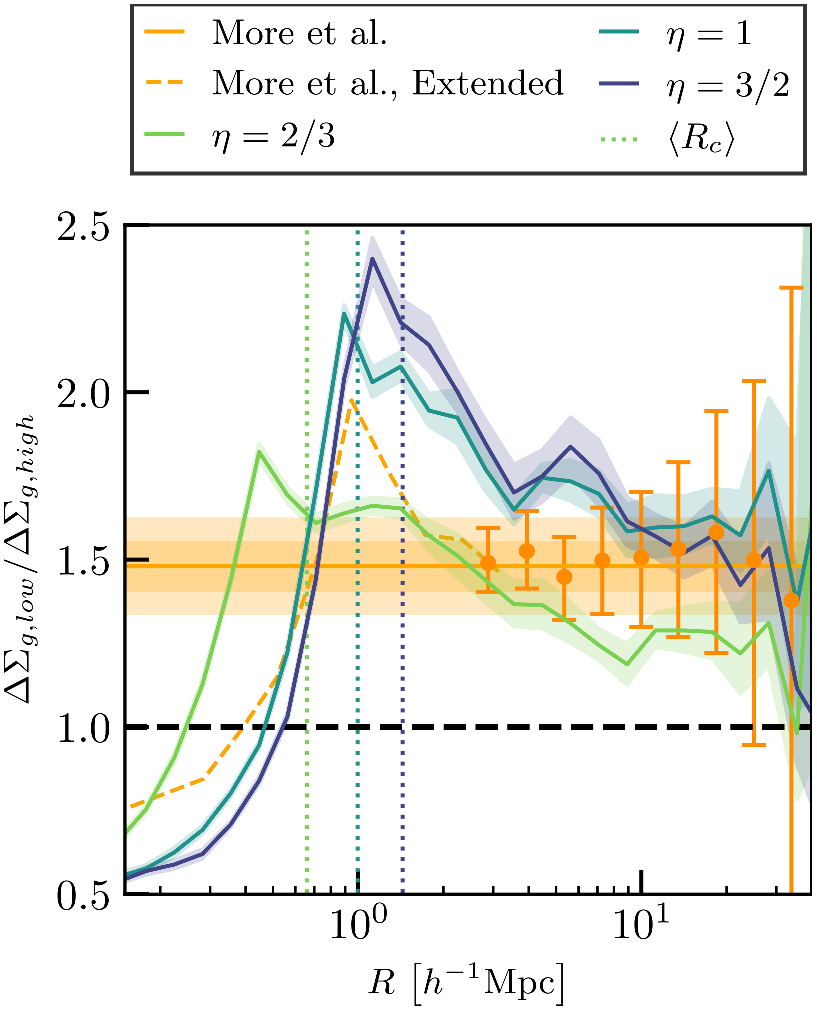

Figure 4.14: The ratio of the projected galaxy number density profiles of the low c_{gal} and high c_{gal} subsamples of CS250, taken from Figure 4.6, is compared with those found for cluster samples selected with the same value of Δz_m but with $\eta = 2/3$ and $3/2$ in Equation 4.6, rather than $\eta = 1$. Points with error bars and their continuation to smaller scales are the same as in Figure 4.6. Vertical lines indicate $\langle R_c \rangle$ for the three samples.

$\langle R_c \rangle$ and a maximum at slightly larger radius in the low- and high-concentration cases respectively.

In none of these panels are the simulated curves a good fit to the observed ones. The results for high c_{gal} clusters match quite well for $\eta = 3/2$, but the best fit for the low c_{gal} clusters is for $\eta = 1$, and even here the overall depth and the general shape of the features differ significantly. Given the strong sensitivity to the cluster identification algorithm and to the splitting by $c_{\rm gal}$, it is likely that these discrepancies reflect detailed differences between the real redMaPPer and concentration definition procedures and the simplified versions used here. It is clear that it will be very difficult to infer reliable information about splashback signals from data of this kind without a complete understanding of these effects.

4.7 Conclusions

In their analysis of a volume-limited sample of 8648 clusters selected by applying the redMaPPer algorithm to the SDSS/DR8 photometric data, More et al. (2016) detected strong assembly bias as a function of cluster concentration on projected scales $5h^{-1}\,\mathrm{Mpc} < R < 30h^{-1}\,\mathrm{Mpc}$, and substantial variations in the slope of cluster projected galaxy number density profiles in the range $500h^{-1}\,\mathrm{kpc} < R < 5h^{-1}\,\mathrm{Mpc}$ which they attributed to splashback effects. The assembly bias signal had previously been seen at lower signal-to-noise by Miyatake et al. (2016) in gravitational lensing data for the same cluster sample. By using a simplified version of the redMaPPer scheme on three orthogonal projections of publicly available galaxy catalogues from the Millennium Simulation, we have been able to identify up to 9196 clusters of similar richness, which we classify by concentration in a similar way to the SDSS studies. This allows us to carry out analyses directly analogous to those of More et al. (2016) and Miyatake et al. (2016) and to compare with results obtained from the full 3D information available for the simulation. This gives considerable insight into the features seen in the SDSS analysis.

The mean projected profiles of mass and galaxy number density which we find for the simulation are very similar to those found observationally, both for the cluster sample as a whole and for its low- and high-concentration subsamples. The apparent assembly bias on large scales agrees well with that observed, as does the shape of the ratio of the low- and high-concentration profiles which rises with decreasing projected radius R to a peak at the mean value of R_c, the limiting radius used to define clusters, before dropping precipitously to smaller scales. The variation with R of the logarithmic slope of the mean galaxy number density profiles shows a more complex structure than in SDSS, but reproduces the main features pointed out by More et al. (2016): the main minimum (the point where the profile is steepest) occurs at smaller radius than expected from the splashback studies of Diemer and Kravtsov (2014) and in addition the minima for the low- and high-concentration subsamples rank oppositely to the splashback expectation both in depth and in radius.

The observed large-scale assembly bias is best reproduced when all red galaxies projected onto a cluster (hence within $\pm 250h^{-1}\,\mathrm{Mpc}$ in depth) are considered as potential

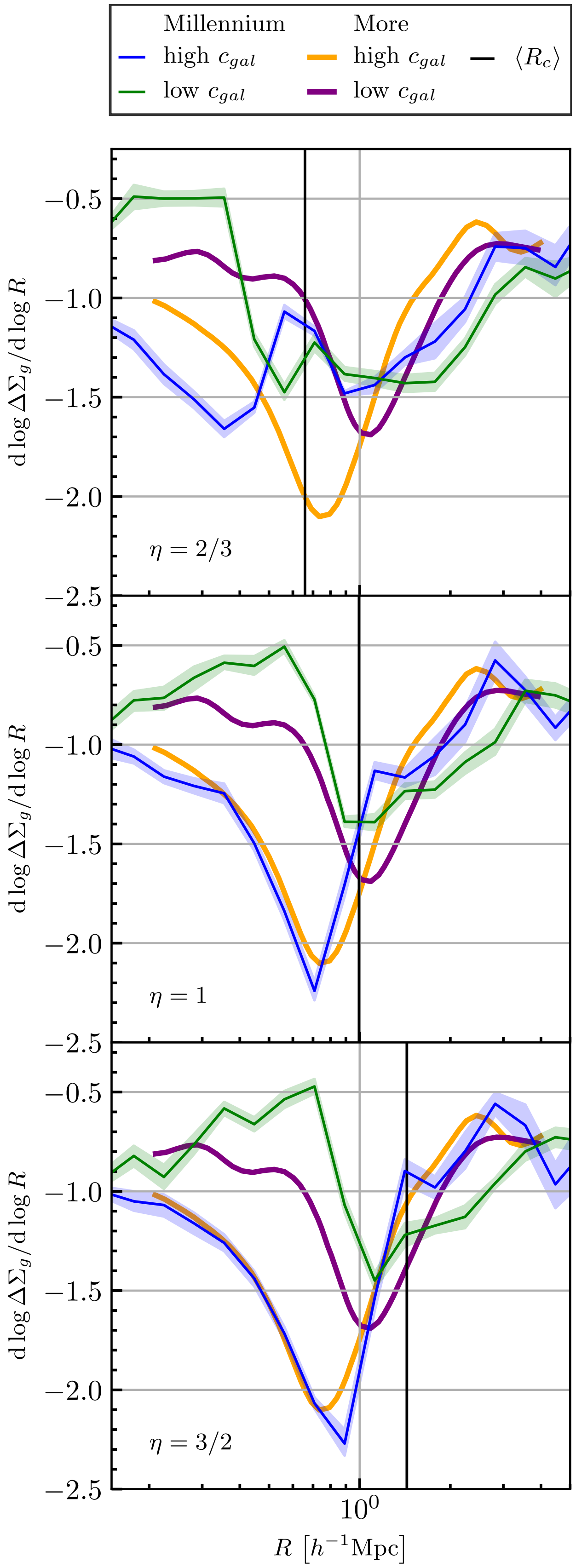

Figure 4.15: The logarithmic derivatives of simulated and observed $\Delta\Sigma_g$ profiles from Figure 4.2 are repeated in the middle panel and compared with results from simulated cluster catalogues with the same value of Δz_m but $\eta = 2/3$ and $3/2$ (top and bottom panels respectively). A solid vertical line in each panel indicates the value of $\langle R_c \rangle$ for the relevant sample.

members. The signal is slightly weaker if the maximal allowed depth offset is reduced to $120h^{-1}\,\mathrm{Mpc}$ and significantly weaker if it is reduced to $60h^{-1}\,\mathrm{Mpc}$. Such changes have negligible effect on the logarithmic slope profiles of stacked galaxy counts. Hence projection over relatively large depths appear to be a significant factor in apparent assembly bias but not in apparent splashback features.

The above results, derived by stacking simulated clusters in projection, can be compared to results obtained from a directly analogous analysis of the full 3D data. This shows some striking differences. The 3D assembly bias for separations between 3 and $30h^{-1}\,\mathrm{Mpc}$ is considerably smaller than that seen in 2D ($b \sim 1.15$ rather than $b \sim 1.5$) and varies in the opposite way with the maximum depth offset allowed for cluster members. The peak in the ratio of the galaxy number density profiles for low- and high-concentration clusters occurs at a substantially larger radius in 3D than in 2D ($r \sim 2.5h^{-1}\,\mathrm{Mpc}$ rather than $R \sim 800h^{-1}\,\mathrm{kpc}$). The logarithmic derivatives of the 3D mass and galaxy overdensity profiles vary more smoothly than in 2D, and show a single minimum which is at larger radius than in 2D and at the same position for the low- and high-concentration clusters. The ranking of the minima in depth remains opposite to that expected from splashback theory. (See the Appendix for a discussion of how cluster selection, scaling and stacking procedures can affect apparent splashback features).

The effects of projection and cluster definition on stacked cluster profiles can be clarified by examining them in the two-dimensional space of projected separation and line-of-sight depth. This allows identification of the depth ranges which give rise to the difference in projected counts around low- and high-concentration clusters. As expected, the galaxy excess at small projected radius which produces the high central surface density of high-concentration clusters is made up of objects which are close to the cluster centre also in 3D. In redshift space, these excess counts appear at offsets $\sim 800\,\mathrm{km\,s^{-1}}$, in the wings of the cluster velocity dispersion. At projected radii $500h^{-1}\,\mathrm{kpc} < R < 2h^{-1}\,\mathrm{Mpc}$, much of the projected count excess around low-concentration clusters comes from galaxies offset in depth by $\sim 1h^{-1}\,\mathrm{Mpc}$, apparently indicating that low-concentration clusters live in richer environments than their high-concentration analogues. At larger projected separation, the galaxies responsible for the strong assembly bias signal are distributed almost uniformly over the full depth accessible to potential cluster members, showing that they are correlated with background groups preferentially projected onto the low-concentration clusters, rather than with the clusters themselves. The overall effect of projection on 2D assembly bias clearly depends strongly both on the details of cluster and concentration definition and on the accuracy of the available photometric redshifts.

At projected radii $500h^{-1}\,\mathrm{kpc} < R < 3h^{-1}\,\mathrm{Mpc}$ where splashback effects are expected to be present, distant foreground and background galaxies contribute negligibly to projected cluster profiles. These are, however, strongly affected by the specific algorithms used to identify clusters and to classify them according to concentration. We demonstrate this explicitly by changing the limiting radius R_c within which red galaxies are counted as cluster members. Even though we take care to adjust parameters so that the abundance and typical mass of clusters are matched for different choices of limiting radius, we find that this radius is strongly imprinted on the mean projected profiles of the resulting samples. The effects are dramatic, both on the ratio of the profiles for low- and high-concentration clusters and on the shape of the logarithmic derivative profiles for the individual subsamples. It will be difficult to obtain reliable information about splashback without detailed understanding of such effects for the particular algorithms used to select an observed cluster sample.

4.A The effect of stacking procedures on apparent splashback signal

In Section 4.4.1 we noted that logarithmic derivative curves for the stacked 3D mass profiles of our clusters (Figure 4.8) differ in shape, particularly in the depth of the minimum, from those shown for objects of similar mass by Diemer and Kravtsov (2014) (DK14). A general difference in behaviour between mean and median of profile stacks was already mentioned in DK14. Here we investigate how the shapes of such profiles depend on the definition of the sample to be stacked and on the scaling and stacking procedures adopted.

In Figure 4.16, the purple curve is taken directly from DK14 where it is the one labelled $z = 0.25$ in the upper central panel of their Figure 4. It corresponds to haloes in a relatively narrow range of M_{200m}, selected at a redshift and with a mean mass which are close to those of the cluster sample analysed in this chapter. DK14 scaled the 3D mass profile of each cluster to its individual R_{200m} and then constructed the stack by taking the *median* value of density at each r/R_{200m}. The logarithmic derivative of the resulting profile is the quantity plotted. Note that it differs from the quantity plotted in Figure 4.8 in that DK14 did not subtract the mean background density from their stack. This has a significant effect beyond a few Mpc.

The light blue curve in Figure 4.16 corresponds to our full sample CS250, stacked in the same way as in Section 4.4.1, i.e. we constructed a spherically averaged mass profile around the central galaxy of each cluster, we averaged these profiles directly to obtain the stack, we scaled the result by the $\langle R_{200m}\rangle$ of the stack, and we then plotted its derivative. The curve effectively corresponds to an average of the two curves shown in Figure 4.8, except for differences at large r/R_{200m} due to the inclusion of the cosmic mean density. Its minimum value is about -2.7, just above the average of the values for the two curves in Figure 4.8 and considerably above the value found by DK14.

The orange curve in Figure 4.16 shows what happens if we scale the profile of each cluster in radius by its individual value of R_{200m} before stacking. This changes the shape of the curve, lowering its minimum slightly and moving it to slightly smaller radii. Not surprisingly, scaling before stacking results in a sharper transition between the one-halo and two-halo parts of the stacked profile.

If we stack these same scaled profiles by constructing their median at each r/R_{200m}, rather than their mean, we obtain the green curve. The minimum is now significantly deeper, although still not as deep as that found by DK14. The shape of the curve outside the minimum agrees very well with their results.

Finally, if we select halos directly from the Millennium Simulation with a narrow range of M_{200m} at $z = 0.24$, and we make a median stack after scaling each profile to its individual R_{200m} value, then we should be reproducing the halo selection and stacking procedures of DK14 almost exactly. The result is shown as a red curve in Figure 4.16. It now differs only slightly from the purple curve.

We suspect that these small residual discrepancies reflect differences in the effective smoothing associated with halo profile construction and differentiation. Overall, the results described here indicate that curves of this type are sensitive to how the halos are scaled and whether a mean or median stack is constructed. The minimum logarithmic slope is particularly sensitive to these factors, and changes in shape can also shift the position of the minimum by 10 or 20 per cent. We note that for individual observed clusters the value of R_{200m} is unknown, the full 3D information is not available, and the selection and definition effects on 2D profiles which we discuss in the main body of the chapter are large compared to the effects described here.

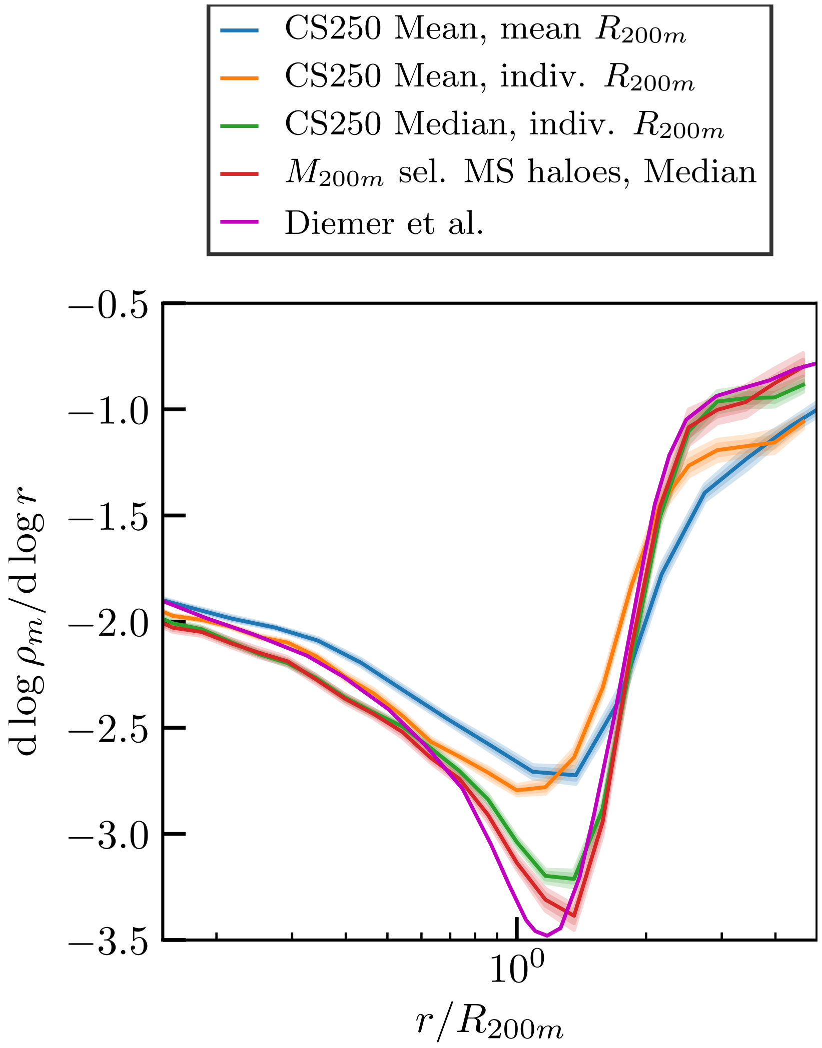

Figure 4.16: Logarithmic derivative curves for different definitions of the radially rescaled 3D mass density profile of simulated clusters are compared to the $z = 0.25$, $2 < \nu < 2.5$ curve given in DK14. For a description of the other curves the reader is referred to the text in Appendix 4.A.

Chapter 5

Conclusions

This thesis is, just as my time as a PhD student was, filled with a wide range of topics and it is therefore not easy to give a single overarching conclusion. Still, we can identify a few common themes that connect them: size estimation of surfaces, their percolation, and bias.

We obtained quite remarkable results regarding the percolation transition in the cosmic density field. Not only were we able to find the percolation transition at 6 to 7 times the mean density but also connect this to clustering properties of the density peaks that live in the percolating web.

Apart from the peculiarity of the bias of haloes attached to the cosmic web we also uncovered a complex landscape of bias values for haloes in mass-ρ_{lim} space. Separating haloes by ρ_{lim} does not only give us populations of group-sized haloes that are avoiding matter on large scales, but also the strongest known assembly bias effect when looking at quintiles in the properties.

While percolation has previously been discussed for the overlap-phase of reionisation, we could add a new perspective to it by connecting it to the local size of bubbles. More importantly we proposed a novel way of measuring this local size and showed how it evolved with progressing reionisation. We also used it connect it to the clustering of bubbles for which we gave a first measurement of bias in simulations using this new measure. This way we could also show how the onset of the formation of superbubbles from overlapping bubbles reflects on the density distribution within them and their larger environment.

In the preceding chapter we saw how difficult it can be to observe assembly bias and the splashback radius (another example of a percolating surface in the universe, see Stücker et al. (2018)). This shows that a lot of work will have to be invested to improve the means to detect the effects we discussed in this thesis.

The Road Forward

The next goal in our endeavour of quantifying morphology and connecting it to processes in the cosmos is a convergence of the techniques and approaches presented in this work.

We would like to find a description of an object delimited by a surface that represents its full morphology, with which we mean topology and sizes. The medial axis transform (MAT) (Blum, 1967) is a tool to achieve this description.

It transforms a given $(d-1)$-dimensional closed surface embedded in a d-dimensional Euclidean space into its $(d-1)$-dimensional skelteon. If this skeleton is decorated with a radius at every point, the surface can be reconstructed by taking the union of balls of said radius each anchored at its corresponding point. Thus it is a homeomorphic representation of reduced complexity.

We can treat the MAT just as the MBS from Section 3.3.5 with the additional knowledge of the connectivity between the centres. Thus, we obtain an object that also directly contains the topology of the described surface. While the MAT itself is very complex, maybe too complex for direct meaningful analysis, it is very easy to simplify as at its core it is just a graph representing the sphere centres and their connections and in its most stripped down form it only retains branching points in loops to encode the bare topology of the described surface. We believe such a description of morphology would be universally useful and applicable in astrophysical problems.

Apart from extending the methodology we would also apply the presented methodology to more phenomena. Just as there is a hierarchy of peaks, there is a hierarchy of voids. The tools developed in chapter 2 can equivalently applied to their study. The same is true for our investigations of reionisation which so far have been restricted to the ionisation fraction. These will be extended to the 21cm lines under the effects of observations with current and future instruments.

It is hard enough to find a splashback surface in simulations (Mansfield et al., 2017; Diemer, 2017) but by following the dark matter sheet we can investigate the morphology of caustics in the universe and hopefully understand more about the connection between their shape and the underlying processes such as galaxy formation. In general we will try to connect the galaxy population to the morphology of the structure that hosts them. This might allow us to find better observables telling us about the cosmic web and its collapse.

Bibliography

Adhikari, S., Dalal, N., and Chamberlain, R. T. (2014). Splashback in accreting dark matter halos. *Journal of Cosmology and Astro-Particle Physics*, 11:019.

Allgood, B., Flores, R. A., Primack, J. R., Kravtsov, A. V., Wechsler, R. H., Faltenbacher, A., and Bullock, J. S. (2006). The shape of dark matter haloes: dependence on mass, redshift, radius and formation. *Monthly Notices of the Royal Astronomical Society*, 367(4):1781–1796.

Aubert, D., Pichon, C., and Colombi, S. (2004). The origin and implications of dark matter anisotropic cosmic infall on ˜L* haloes. *Monthly Notices of the Royal Astronomical Society*, 352:376–398.

Bag, S., Mondal, R., Sarkar, P., Bharadwaj, S., and Sahni, V. (2018). The shape and size distribution of H II regions near the percolation transition. *Monthly Notices of the Royal Astronomical Society*, 477:1984–1992.

Bagla, J. S. (2005). Cosmological N-Body simulation: Techniques, Scope and Status. *Current Science*, 88:1088–1100.

Barkana, R. and Loeb, A. (2001). In the beginning: The first sources of light and the reionization of the universe. *Physics Reports*, 349:125–238.

Becker, R. H., Fan, X., White, R. L., Strauss, M. A., Narayanan, V. K., Lupton, R. H., Gunn, J. E., Annis, J., Bahcall, N. A., Brinkmann, J., Connolly, A. J., Csabai, I., Czarapata, P. C., Doi, M., Heckman, T. M., Hennessy, G. S., Ivezić, Z., Knapp, G. R., Lamb, D. Q., McKay, T. A., Munn, J. A., Nash, T., Nichol, R., Pier, J. R., Richards, G. T., Schneider, D. P., Stoughton, C., Szalay, A. S., Thakar, A. R., and York, D. G. (2001). Evidence for Reionization at z 6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar. *The Astronomical Journal*, 122(6):2850–2857.

Behroozi, P. S., Wechsler, R. H., and Wu, H.-Y. (2013). The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *The Astrophysical Journal*, 762(2):109.

Bertschinger, E. (1985). Self-similar secondary infall and accretion in an Einstein-de Sitter universe. *The Astrophysical Journal Supplement Series*, 58:39–65.

Blum, H. (1967). A transformation for extracting new descriptors of shape. *Models for the perception of speech and visual form*, 19(5):362–380.

Bond, J. R., Kofman, L., and Pogosyan, D. (1996). How filaments of galaxies are woven into the cosmic web. *Nature*, 380:603–606.

Borzyszkowski, M., Porciani, C., Romano-Díaz, E., and Garaldi, E. (2017). ZOMG - I. How the cosmic web inhibits halo growth and generates assembly bias. *Monthly Notices of the Royal Astronomical Society*, 469(1):594.

Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., and Mahesh, N. (2018). An absorption profile centred at 78 megahertz in the sky-averaged spectrum. *Nature*, 555(7694):67–70.

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., and Lemson, G. (2009). Resolving Cosmic Structure Formation with the Millennium-II Simulation. *Monthly Notices of the Royal Astronomical Society*, 398(3):1150–1164. arXiv: 0903.3041.

Bromm, V. and Larson, R. B. (2004). The First Stars. *Annual Review of Astronomy and Astrophysics*, 42:79–118.

Bruzual, G. and Charlot, S. (2003). Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*, 344:1000–1028.

Busch, P., Eide, M. B., Kakiichi, K., and Ciardi, B. (2019). Opening Reionisation: Morphological Characterisation of Ionised Bubbles via Binary Fields. in prep.

Busch, P. and White, S. D. M. (2017). Assembly bias and splashback in galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 470(4):4767–4781.

Busch, P. and White, S. D. M. (2019a). Assembly Bias of Density Peaks in the Millennium Simulations with the Tessellation Level Tree. in prep.

Busch, P. and White, S. D. M. (2019b). The Structure of Density Peaks in the Millennium Simulations with the Tessellation Level Tree. in prep.

Busch, P. and White, S. D. M. (2019c). The Tessellation Level Tree: characterising the nested hierarchy of density peaks and their spatial distribution in cosmological N-body simulations. in prep.

Cautun, M., Cai, Y.-C., and Frenk, C. S. (2016). The view from the boundary: A new void stacking method. *Monthly Notices of the Royal Astronomical Society*, 457:2540–2553.

Chen, Z., Xu, Y., Wang, Y., and Chen, X. (2018). Stages of Reionization as revealed by the Minkowski Functionals. *arXiv e-prints*, 1812:arXiv:1812.10333.

Ciardi, B., Bianchi, S., and Ferrara, A. (2002). Lyman continuum escape from an inhomogeneous interstellar medium. *Monthly Notices of the Royal Astronomical Society*, 331:463–473.

Ciardi, B., Bolton, J. S., Maselli, A., and Graziani, L. (2012). The effect of intergalactic helium on hydrogen reionization: Implications for the sources of ionizing photons at z>6. *Monthly Notices of the Royal Astronomical Society*, 423:558–574.

Ciardi, B. and Ferrara, A. (2005). The First Cosmic Structures and Their Effects. *Space Science Reviews*, 116:625–705.

Ciardi, B., Stoehr, F., and White, S. D. M. (2003). Simulating intergalactic medium reionization. *Monthly Notices of the Royal Astronomical Society*, 343:1101–1109.

Coles, P. P. and Lucchin, F. (2002). *Cosmology: The Origin and Evolution of Cosmic Structure, 2nd Edition*. Wiley-Blackwell, 2nd edition edition.

Dalal, N., White, M., Bond, J. R., and Shirokov, A. (2008). Halo Assembly Bias in Hierarchical Structure Formation. *The Astrophysical Journal*, 687:12–21.

Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292:371–394.

DES Collaboration (2018). Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *Physical Review D*, 98(4):043526.

Desjacques, V., Jeong, D., and Schmidt, F. (2018). Large-scale galaxy bias. *Physics Reports*, 733:1–193.

Diemer, B. (2017). The Splashback Radius of Halos from Particle Dynamics. I. The SPARTA Algorithm. *The Astrophysical Journal Supplement Series*, 231(1):5.

Diemer, B. and Kravtsov, A. V. (2014). Dependence of the Outer Density Profiles of Halos on Their Mass Accretion Rate. *The Astrophysical Journal*, 789(1):1.

Diemer, B., Mansfield, P., Kravtsov, A. V., and More, S. (2017). The Splashback Radius of Halos from Particle Dynamics. II. Dependence on Mass, Accretion Rate, Redshift, and Cosmology. *The Astrophysical Journal*, 843(2):140.

Dubois, Y., Peirani, S., Pichon, C., Devriendt, J., Gavazzi, R., Welker, C., and Volonteri, M. (2016). The HORIZON-AGN simulation: morphological diversity of galaxies promoted by AGN feedback. *Monthly Notices of the Royal Astronomical Society*, 463:3948–3964.

Dvornik, A., Cacciato, M., Kuijken, K., Viola, M., Hoekstra, H., Nakajima, R., van Uitert, E., Brouwer, M., Choi, A., Erben, T., Fenech Conti, I., Farrow, D. J., Herbonnet, R., Heymans, C., Hildebrandt, H., Hopkins, A. M., McFarland, J., Norberg, P., Schneider, P., Sifón, C., Valentijn, E., and Wang, L. (2017). A KiDS weak lensing analysis of assembly bias in GAMA galaxy groups. *Monthly Notices of the Royal Astronomical Society*, 468:3251–3265.

Eide, M. B., Graziani, L., Ciardi, B., Feng, Y., Kakiichi, K., and Di Matteo, T. (2018a). The epoch of cosmic heating by early sources of X-rays. *Monthly Notices of the Royal Astronomical Society*, 476:1174–1190.

Eide, M. B., Graziani, L., Ciardi, B., Feng, Y., Kakiichi, K., Di Matteo, T., and Busch, P. (2018b). The epoch of reionization. in preparation.

Faltenbacher, A. and White, S. D. M. (2010). Assembly Bias and the Dynamical Structure of Dark Matter Halos. *The Astrophysical Journal*, 708(1):469.

Fan, X., Strauss, M. A., Richards, G. T., Hennawi, J. F., Becker, R. H., White, R. L., Diamond-Stanic, A. M., Donley, J. L., Jiang, L., Kim, J. S., Vestergaard, M., Young, J. E., Gunn, J. E., Lupton, R. H., Knapp, G. R., Schneider, D. P., Brandt, W. N., Bahcall, N. A., Barentine, J. C., Brinkmann, J., Brewington, H. J., Fukugita, M., Harvanek, M., Kleinman, S. J., Krzesinski, J., Long, D., Neilsen, Jr., E. H., Nitta, A., Snedden, S. A., and Voges, W. (2006). A Survey of z>5.7 Quasars in the Sloan Digital Sky Survey. IV. Discovery of Seven Additional Quasars. *The Astronomical Journal*, 131:1203–1209.

Fillmore, J. A. and Goldreich, P. (1984). Self-similar gravitational collapse in an expanding universe. *The Astrophysical Journal*, 281:1–8.

Frenk, C. S. and White, S. D. M. (2012). Dark matter and cosmic structure. *Annalen der Physik*, 524(9-10):507–534.

Friedrich, M. M., Mellema, G., Alvarez, M. A., Shapiro, P. R., and Iliev, I. T. (2011). Topology and sizes of H ii regions during cosmic reionization. *Monthly Notices of the Royal Astronomical Society*, 413(2):1353–1372.

Furlanetto, S. R. and Oh, S. P. (2016). Reionization Through the Lens of Percolation Theory. *Monthly Notices of the Royal Astronomical Society*, 457(2):1813–1827.

Gao, L., Springel, V., and White, S. D. M. (2005). The age dependence of halo clustering. *Monthly Notices of the Royal Astronomical Society*, 363:L66–L70.

Gao, L. and White, S. D. M. (2007). Assembly bias in the clustering of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 377:L5–L9.

Giri, S. K., Mellema, G., and Ghara, R. (2018). Optimal identification of H II regions during reionization in 21-cm observations. *Monthly Notices of the Royal Astronomical Society*, 479(4):5596.

Gleser, L., Nusser, A., Ciardi, B., and Desjacques, V. (2006). The morphology of cosmological reionization by means of Minkowski functionals. *Monthly Notices of the Royal Astronomical Society*, 370:1329–1338.

Gnedin, N. Y. (2014). Cosmic Reionization On Computers I. Design and Calibration of Simulations. *The Astrophysical Journal*, 793(1):29.

Graziani, L., Maselli, A., and Ciardi, B. (2013). CRASH3: Cosmological radiative transfer through metals. *Monthly Notices of the Royal Astronomical Society*, 431(1):722.

Greig, B. and Mesinger, A. (2018). 21CMMC with a 3D light-cone: The impact of the co-evolution approximation on the astrophysics of reionization and cosmic dawn. *Monthly Notices of the Royal Astronomical Society*, 477(3):3217.

Gunn, J. E. and Gott, III, J. R. (1972). On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *The Astrophysical Journal*, 176:1.

Gunn, J. E. and Peterson, B. A. (1965). On the Density of Neutral Hydrogen in Intergalactic Space. *The Astrophysical Journal*, 142:1633–1641.

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. *Monthly Notices of the Royal Astronomical Society*, 413:101–131.

Hahn, O., Porciani, C., Dekel, A., and Carollo, C. M. (2009). Tidal effects and the environment dependence of halo assembly. *Monthly Notices of the Royal Astronomical Society*, 398:1742–1756.

Han, J., Cole, S., Frenk, C. S., Benitez-Llambay, A., and Helly, J. (2018). HBT+: an improved code for finding subhaloes and building merger trees in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 474:604–617.

Han, J., Jing, Y. P., Wang, H., and Wang, W. (2012). Resolving subhaloes' lives with the Hierarchical Bound-Tracing algorithm. *Monthly Notices of the Royal Astronomical Society*, 427:2437–2449.

Hearin, A. P., Watson, D. F., Becker, M. R., Reyes, R., Berlind, A. A., and Zentner, A. R. (2014). The dark side of galaxy colour: evidence from new SDSS measurements of galaxy clustering and lensing. *Monthly Notices of the Royal Astronomical Society*, 444:729–743.

Henriques, B. M. B., White, S. D. M., Thomas, P. A., Angulo, R., Guo, Q., Lemson, G., Springel, V., and Overzier, R. (2015). Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses. *Monthly Notices of the Royal Astronomical Society*, 451:2663–2680.

Kaiser, N. (1984). On the spatial correlations of Abell clusters. *The Astrophysical Journal Letters*, 284:L9–L12.

Kakiichi, K., Graziani, L., Ciardi, B., Meiksin, A., Compostella, M., Eide, M. B., and Zaroubi, S. (2017a). The concerted impact of galaxies and QSOs on the ionization and thermal state of the intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 468:3718–3736.

Kakiichi, K., Majumdar, S., Mellema, G., Ciardi, B., Dixon, K. L., Iliev, I. T., Jelić, V., Koopmans, L. V. E., Zaroubi, S., and Busch, P. (2017b). Recovering the H II region size statistics from 21-cm tomography. *Monthly Notices of the Royal Astronomical Society*, 471:1936–1954.

Kapahtia, A., Chingangbam, P., Appleby, S., and Park, C. (2018). A novel probe of ionized bubble shape and size statistics of the epoch of reionization using the contour Minkowski Tensor. *Journal of Cosmology and Astro-Particle Physics*, 2018(10):011.

Khandai, N., Di Matteo, T., Croft, R., Wilkins, S., Feng, Y., Tucker, E., DeGraf, C., and Liu, M.-S. (2015). The MassiveBlack-II simulation: the evolution of haloes and galaxies to z ˜ 0. *Monthly Notices of the Royal Astronomical Society*, 450:1349–1374.

Knebe, A., Pearce, F. R., Lux, H., Ascasibar, Y., Behroozi, P., Casado, J., Moran, C. C., Diemand, J., Dolag, K., Dominguez-Tenreiro, R., Elahi, P., Falck, B., Gottlöber, S., Han, J., Klypin, A., Lukić, Z., Maciejewski, M., McBride, C. K., Merchán, M. E., Muldrew, S. I., Neyrinck, M., Onions, J., Planelles, S., Potter, D., Quilis, V., Rasera, Y., Ricker, P. M., Roy, F., Ruiz, A. N., Sgró, M. A., Springel, V., Stadel, J., Sutter, P. M., Tweed, D., and Zemp, M. (2013). Structure finding in cosmological simulations: The state of affairs. *Monthly Notices of the Royal Astronomical Society*, 435(2):1618.

Komatsu, E., Dunkley, J., Nolta, M. R., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Limon, M., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Meyer, S. S., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2009). Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation. *The Astrophysical Journal Supplement Series*, 180:330–376.

Kosheleva, O., Cabrera, S. D., Gibson, G. A., and Koshelev, M. (1997). Fast implementations of fuzzy arithmetic operations using fast Fourier transform (FFT). *Fuzzy Sets and Systems*, 91(2):269–277.

Kraljic, K., Arnouts, S., Pichon, C., Laigle, C., de la Torre, S., Vibert, D., Cadiou, C., Dubois, Y., Treyer, M., Schimd, C., Codis, S., de Lapparent, V., Devriendt, J., Hwang, H. S., Le Borgne, D., Malavasi, N., Milliard, B., Musso, M., Pogosyan, D., Alpaslan, M., Bland-Hawthorn, J., and Wright, A. H. (2018). Galaxy evolution in the metric of the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 474:547–571.

Kreisch, C. D., Pisani, A., Carbone, C., Liu, J., Hawken, A. J., Massara, E., Spergel, D. N., and Wandelt, B. D. (2018). Massive Neutrinos Leave Fingerprints on Cosmic Voids.

Lacey, C. and Cole, S. (1994). Merger Rates in Hierarchical Models of Galaxy Formation - Part Two - Comparison with N-Body Simulations. *Monthly Notices of the Royal Astronomical Society*, 271:676.

Lazeyras, T., Musso, M., and Schmidt, F. (2017). Large-scale assembly bias of dark matter halos. *Journal of Cosmology and Astro-Particle Physics*, 03:059.

Lee, J. and Hoyle, F. (2015). An Observational Signal of the Void Shape Correlation and its Link to the Cosmic Web. *The Astrophysical Journal*, 803:45.

Lemson, G. and the Virgo Consortium (2006). Halo and galaxy formation histories from the millennium simulation: Public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the LambdaCDM cosmogony. *ArXiv Astrophysics e-prints*, pages arXiv:astro–ph/0608019.

Libeskind, N. I., van de Weygaert, R., Cautun, M., Falck, B., Tempel, E., Abel, T., Alpaslan, M., Aragón-Calvo, M. A., Forero-Romero, J. E., Gonzalez, R., Gottlöber, S., Hahn, O., Hellwing, W. A., Hoffman, Y., Jones, B. J. T., Kitaura, F., Knebe, A., Manti, S., Neyrinck, M., Nuza, S. E., Padilla, N., Platen, E., Ramachandra, N., Robotham, A., Saar, E., Shandarin, S., Steinmetz, M., Stoica, R. S., Sousbie, T., and Yepes, G. (2018). Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 473:1195–1217.

Lin, Y., Oh, S. P., Furlanetto, S. R., and Sutter, P. M. (2016a). The distribution of bubble sizes during reionization. *Monthly Notices of the Royal Astronomical Society*, 461:3361–3374.

Lin, Y.-T., Mandelbaum, R., Huang, Y.-H., Huang, H.-J., Dalal, N., Diemer, B., Jian, H.-Y., and Kravtsov, A. (2016b). On Detecting Halo Assembly Bias with Galaxy Populations. *The Astrophysical Journal*, 819(2):119. arXiv: 1504.07632.

Lithwick, Y. and Dalal, N. (2011). Self-similar Solutions of Triaxial Dark Matter Halos. *The Astrophysical Journal*, 734:100.

Liu, A., Tegmark, M., Bowman, J., Hewitt, J., and Zaldarriaga, M. (2009). An improved method for 21-cm foreground removal. *Monthly Notices of the Royal Astronomical Society*, 398:401–406.

Loeb, A. and Zaldarriaga, M. (2004). Measuring the Small-Scale Power Spectrum of Cosmic Density Fluctuations through 21cm Tomography Prior to the Epoch of Structure Formation. *Physical Review Letters*, 92:211301.

LOFAR Collaboration (2013). LOFAR: The LOw-Frequency ARray. *Astronomy and Astrophysics*, 556:A2.

Ma, X., Hopkins, P. F., Kasen, D., Quataert, E., Faucher-Giguère, C.-A., Keres, D., Murray, N., and Strom, A. (2016). Binary stars can provide the ‘missing photons’ needed for reionization. *Monthly Notices of the Royal Astronomical Society*, 459:3614–3619.

Madau, P. and Haardt, F. (2015). Cosmic Reionization after Planck: Could Quasars Do It All? *The Astrophysical Journal Letters*, 813:L8.

Mansfield, P., Kravtsov, A. V., and Diemer, B. (2017). Splashback Shells of Cold Dark Matter Halos. *The Astrophysical Journal*, 841(1):34.

Maselli, A., Ferrara, A., and Ciardi, B. (2003). CRASH: a radiative transfer scheme. *Monthly Notices of the Royal Astronomical Society*, 345:379–394.

Massara, E., Villaescusa-Navarro, F., Viel, M., and Sutter, P. M. (2015). Voids in massive neutrino cosmologies. *Journal of Cosmology and Astro-Particle Physics*, 11:018.

Meerburg, P. D., Dvorkin, C., and Spergel, D. N. (2013). Probing Patchy Reionization through τ-21 cm Correlation Statistics. *The Astrophysical Journal*, 779:124.

Mellema, G., Koopmans, L. V. E., Abdalla, F. A., Bernardi, G., Ciardi, B., Daiboo, S., de Bruyn, A. G., Datta, K. K., Falcke, H., Ferrara, A., Iliev, I. T., Iocco, F., Jelić, V., Jensen, H., Joseph, R., Labroupoulos, P., Meiksin, A., Mesinger, A., Offringa, A. R., Pandey, V. N., Pritchard, J. R., Santos, M. G., Schwarz, D. J., Semelin, B., Vedantham, H., Yatawatta, S., and Zaroubi, S. (2013). Reionization and the Cosmic Dawn with the Square Kilometre Array. *Experimental Astronomy*, 36:235–318.

Mertens, F. G., Ghosh, A., and Koopmans, L. V. E. (2018). Statistical 21-cm signal separation via Gaussian Process Regression analysis. *Monthly Notices of the Royal Astronomical Society*, 478(3):3640.

Mesinger, A., Furlanetto, S., and Cen, R. (2011). 21cmfast: A fast, seminumerical simulation of the high-redshift 21-cm signal. *Monthly Notices of the Royal Astronomical Society*, 411(2):955–972.

Miyatake, H., More, S., Takada, M., Spergel, D. N., Mandelbaum, R., Rykoff, E. S., and Rozo, E. (2016). Evidence of Halo Assembly Bias in Massive Clusters. *Physical Review Letters*, 116:041301.

Mo, H. J. and White, S. D. M. (1996). An analytic model for the spatial clustering of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 282:347–361.

More, S., Diemer, B., and Kravtsov, A. (2015). The splashback radius as a physical halo boundary and the growth of halo mass. *The Astrophysical Journal*, 810(1):36. arXiv: 1504.05591.

More, S., Kravtsov, A. V., Dalal, N., and Gottlöber, S. (2011). The Overdensity and Masses of the Friends-of-friends Halos and Universality of Halo Mass Function. *The Astrophysical Journal Supplement Series*, 195:4.

More, S., Miyatake, H., Takada, M., Diemer, B., Kravtsov, A. V., Dalal, N. K., More, A., Murata, R., Mandelbaum, R., Rozo, E., Rykoff, E. S., Oguri, M., and Spergel, D. N. (2016). Detection of the Splashback Radius and Halo Assembly Bias of Massive Galaxy Clusters. *The Astrophysical Journal*, 825:39.

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1995). Simulations of X-ray clusters. *Monthly Notices of the Royal Astronomical Society*, 275:720–740.

Neyrinck, M. C. (2008). ZOBOV: a parameter-free void-finding algorithm. *Monthly Notices of the Royal Astronomical Society*, 386(4):2101–2109. arXiv: 0712.3049.

Neyrinck, M. C., Gnedin, N. Y., and Hamilton, A. J. S. (2005). VOBOZ: an almost-parameter-free halo-finding algorithm. *Monthly Notices of the Royal Astronomical Society*, 356:1222–1232.

Ocvirk, P., Gillet, N., Shapiro, P. R., Aubert, D., Iliev, I. T., Teyssier, R., Yepes, G., Choi, J.-H., Sullivan, D., Knebe, A., Gottlöber, S., D'Aloisio, A., Park, H., Hoffman, Y., and Stranex, T. (2016). Cosmic Dawn (CoDa): The First Radiation-Hydrodynamics Simulation of Reionization and Galaxy Formation in the Local Universe. *Monthly Notices of the Royal Astronomical Society*, 463(2):1462.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley, Chichester ; New York, 2 edition edition.

Pacucci, F., Ferrara, A., Grazian, A., Fiore, F., Giallongo, E., and Puccetti, S. (2016). First identification of direct collapse black hole candidates in the early Universe in CANDELS/GOODS-S. *Monthly Notices of the Royal Astronomical Society*, 459(2):1432–1439.

Pandey, B., White, S. D. M., Springel, V., and Angulo, R. E. (2013). Exploring the non-linear density field in the Millennium Simulations with tessellations - I. The probability distribution function. *Monthly Notices of the Royal Astronomical Society*, 435:2968–2981.

Pawlik, A. H., Rahmati, A., Schaye, J., Jeon, M., and Dalla Vecchia, C. (2017). The Aurora radiation-hydrodynamical simulations of reionization: Calibration and first results. *Monthly Notices of the Royal Astronomical Society*, 466(1):960.

Pillepich, A., Springel, V., Nelson, D., Genel, S., Naiman, J., Pakmor, R., Hernquist, L., Torrey, P., Vogelsberger, M., Weinberger, R., and Marinacci, F. (2018). Simulating galaxy formation with the IllustrisTNG model. *Monthly Notices of the Royal Astronomical Society*, 473(3):4077.

Planck Collaboration (2014). Planck 2013 results. I. Overview of products and scientific results. *Astronomy and Astrophysics*, 571:A1.

Planck Collaboration (2018). Planck 2018 results. I. Overview and the cosmological legacy of Planck.

Rahmati, A. and Schaye, J. (2018). The mean free path of hydrogen ionizing photons during the epoch of reionization. *Monthly Notices of the Royal Astronomical Society*, 478(4):5123.

Riess, A. G., Casertano, S., Yuan, W., Macri, L., Anderson, J., MacKenty, J. W., Bowers, J. B., Clubb, K. I., Filippenko, A. V., Jones, D. O., and Tucker, B. E. (2018). New Parallaxes of Galactic Cepheids from Spatially Scanning the Hubble Space Telescope : Implications for the Hubble Constant. *The Astrophysical Journal*, 855(2):136.

Rosdahl, J., Katz, H., Blaizot, J., Kimm, T., Michel-Dansac, L., Garel, T., Haehnelt, M., Ocvirk, P., and Teyssier, R. (2018). The SPHINX cosmological simulations of the first billion years: The impact of binary stars on reionization. *Monthly Notices of the Royal Astronomical Society*, 479:994–1016.

Rykoff, E. S., Koester, B. P., Rozo, E., Annis, J., Evrard, A. E., Hansen, S. M., Hao, J., Johnston, D. E., McKay, T. A., and Wechsler, R. H. (2012). Robust Optical Richness Estimation with Reduced Scatter. *The Astrophysical Journal*, 746:178.

Rykoff, E. S., Rozo, E., Busha, M. T., Cunha, C. E., Finoguenov, A., Evrard, A., Hao, J., Koester, B. P., Leauthaud, A., Nord, B., Pierre, M., Reddick, R., Sadibekova, T., Sheldon, E. S., and Wechsler, R. H. (2014). redMaPPer. I. Algorithm and SDSS DR8 Catalog. *The Astrophysical Journal*, 785:104.

Sana, H., de Mink, S. E., de Koter, A., Langer, N., Evans, C. J., Gieles, M., Gosset, E., Izzard, R. G., Le Bouquin, J.-B., and Schneider, F. R. N. (2012). Binary Interaction Dominates the Evolution of Massive Stars. *Science*, 337:444.

Schaap, W. E. and van de Weygaert, R. (2000). Continuous fields and discrete samples: reconstruction through Delaunay tessellations. *Astronomy and Astrophysics*, 363:L29–L32.

Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., Jenkins, A., Rosas-Guevara, Y. M., White, S. D. M., Baes, M., Booth, C. M., Camps, P., Navarro, J. F., Qu, Y., Rahmati, A., Sawala, T., Thomas, P. A., and Trayford, J. (2015). The EAGLE project: Simulating the evolution and assembly of galaxies and their environments. *Monthly Notices of the Royal Astronomical Society*, 446(1):521.

Shandarin, S. F. and Zeldovich, Y. B. (1989). The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium. *Reviews of Modern Physics*, 61(2):185–220.

Sheth, R. K. and Tormen, G. (1999). Large-scale bias and the peak background split. *Monthly Notices of the Royal Astronomical Society*, 308:119–126.

Shi, X. (2016). The outer profile of dark matter halos: an analytical approach. *Monthly Notices of the Royal Astronomical Society*, 459(4):3711–3720. arXiv: 1603.01742.

Shim, J., Lee, J., and Li, B. (2014). MASSIVE GRAVITY WRAPPED IN THE COSMIC WEB. *The Astrophysical Journal*, 784(1):84.

Smith, A., Bromm, V., and Loeb, A. (2017). The first supermassive black holes. *Astronomy and Geophysics*, 58(3):3.22.

Springel, V. (2005). The cosmological simulation code GADGET-2. *Monthly Notices of the Royal Astronomical Society*, 364:1105–1134.

Springel, V. (2010). E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Monthly Notices of the Royal Astronomical Society*, 401:791–851.

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*, 435:629–636.

Springel, V., White, S. D. M., Tormen, G., and Kauffmann, G. (2001). Populating a cluster of galaxies - I. Results at [formmu2]z=0. *Monthly Notices of the Royal Astronomical Society*, 328:726–750.

Stücker, J., Busch, P., and White, S. D. M. (2018). The median density of the Universe. *Monthly Notices of the Royal Astronomical Society*, 477:3230–3246.

Sutter, P. M., Lavaux, G., Hamaus, N., Pisani, A., Wandelt, B. D., Warren, M., Villaescusa-Navarro, F., Zivick, P., Mao, Q., and Thompson, B. B. (2015). VIDE: The Void IDentification and Examination toolkit. *Astronomy and Computing*, 9:1.

Tinker, J. L., George, M. R., Leauthaud, A., Bundy, K., Finoguenov, A., Massey, R., Rhodes, J., and Wechsler, R. H. (2012). The Correlated Formation Histories of Massive Galaxies and Their Dark Matter Halos. *The Astrophysical Journal Letters*, 755:L5.

Trenti, M. and Hut, P. (2008). Gravitational N-body Simulations.

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G. F., Nelson, D., and Hernquist, L. (2014). Introducing the Illustris Project: Simulating the coevolution of dark and visible matter in the Universe. *Monthly Notices of the Royal Astronomical Society*, 444(2):1518–1547. arXiv: 1405.2921.

Vogelsberger, M., White, S. D. M., Mohayaee, R., and Springel, V. (2009). Caustics in growing cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 400:2174–2184.

Wang, L., Weinmann, S. M., De Lucia, G., and Yang, X. (2013). Detection of galaxy assembly bias. *Monthly Notices of the Royal Astronomical Society*, 433:515–520.

Wang, Y., Xu, Y., Wu, F., Chen, X., Wang, X., Kim, J., Park, C., Lee, K. G., and Cen, R. (2015). Topology of neutral hydrogen distribution with the Square Kilometre Array. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 33.

Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., and Allgood, B. (2006). The Dependence of Halo Clustering on Halo Formation History, Concentration, and Occupation. *The Astrophysical Journal*, 652:71–84.

Weinreich, G. and Hughes, V. W. (1954). Hyperfine Structure of Helium-3 in the Metastable Triplet State. *Physical Review*, 95(6):1451–1460.

White, S. D. M. and Rees, M. J. (1978). Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering. *Monthly Notices of the Royal Astronomical Society*, 183:341–358.

Wise, J. H., Regan, J. A., O'Shea, B. W., Norman, M. L., Downes, T. P., and Xu, H. (2019). Formation of massive black holes in rapidly growing pre-galactic gas clouds. *Nature*, 566:85–88.

Xu, Y., Yue, B., and Chen, X. (2017). IslandFAST: A Semi-numerical Tool for Simulating the Late Epoch of Reionization. *The Astrophysical Journal*, 844(2):117.

Yang, X., Mo, H. J., and van den Bosch, F. C. (2006). Observational Evidence for an Age Dependence of Halo Bias. *The Astrophysical Journal Letters*, 638:L55–L58.

Zel'dovich, Y. B. (1970). Gravitational instability: An approximate theory for large density perturbations. *Astronomy and Astrophysics*, 5:84–89.

Zu, Y., Mandelbaum, R., Simet, M., Rozo, E., and Rykoff, E. S. (2017). On the level of cluster assembly bias in SDSS. *Monthly Notices of the Royal Astronomical Society*, 470(1):551–560.

Acknowledgements

I would like to thank my supervisors Simon White and Benedetta Ciardi for their continued support, advice and stimulating discussions over the last slightly more than three years. Without their help and frequent moral support this thesis would never have seen the light of day.

The two people without whom I would not have seen it myself are my parents to whom I do not only owe a great upbringing and support in my educational endeavours, but also my drive to learn and wish to increase my understanding of the world. They, just as Charlie and Susi, have supported me with some contentual distance but more than matching personal affection.

I also owe a great deal of gratitude to my comrades in suffering Aoife and Jens from which I learnt a great deal not just about astrophysics. Without them and Martin, Matteo, Max, Ricard, Timo and Marius my last years would not have been as bearable.

Finally I would like to thank Stefan Gottlöber for taking me on my first tours of the dark universe and Volker Springel for first opening my eyes to its existence and the ways we can use computers to understand it.