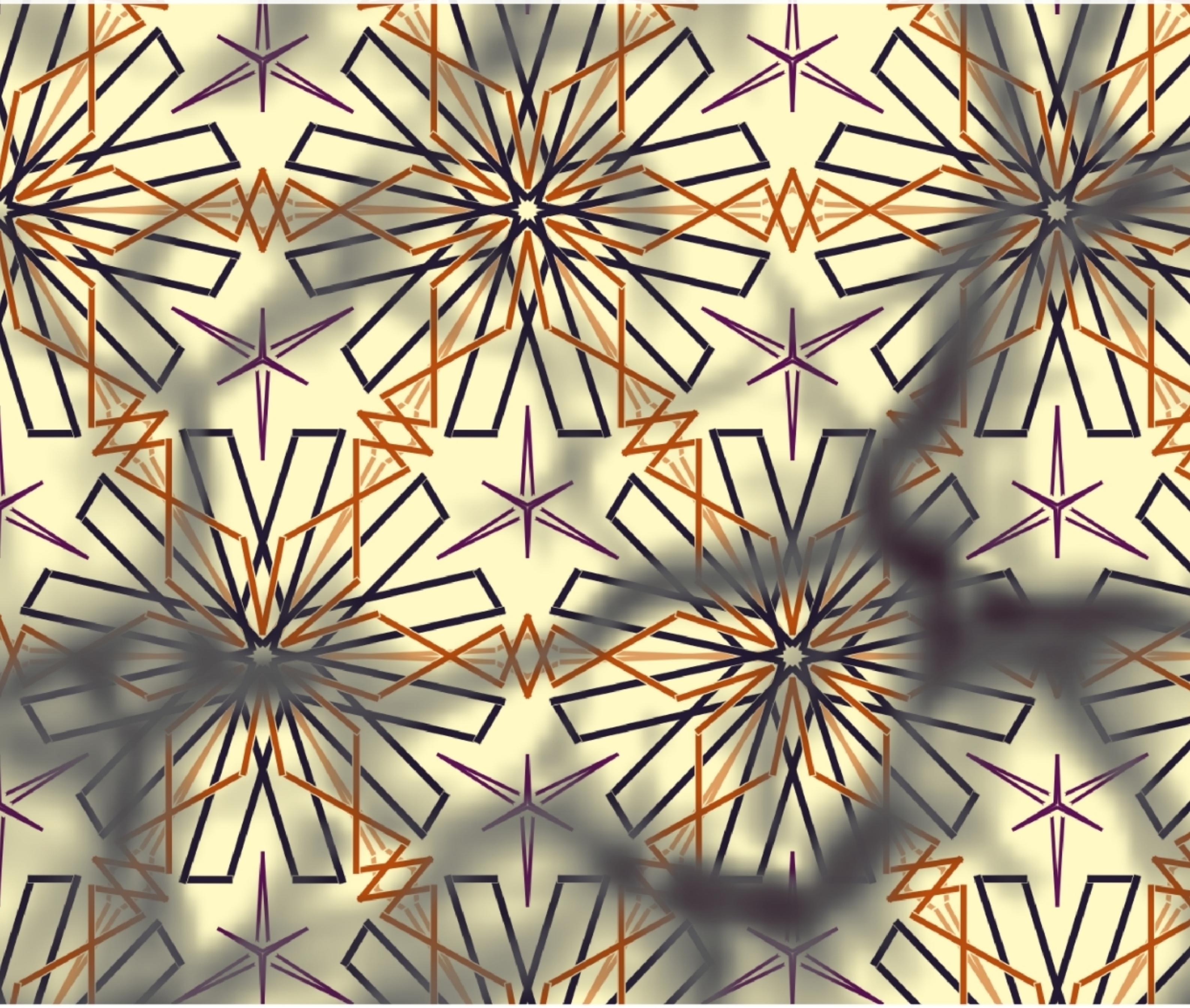
university of groningen
faculty of mathematics and natural sciences
astronomy

Velocity fields:
probes and agents of Cosmic Web evolution

Mathijs van de Mast



Velocity fields:
probes and agents of Cosmic Web evolution

Mathijs van de Mast
Supervisor: Rien van de Weygaert
Second reader: Marco Spaans

Abstract

We have studied the density and velocity fields from the Cosmogrid simulation (Ishiyama et al., 2013). Our methods include a Delaunay Tessellation Field Estimation (DTFE) of density and velocity fields from the simulation particles. We decomposed the velocity gradient into divergence, shear and vorticity, and classified six different components of the cosmic web on the basis of the eigenvectors of the deformation tensor.

This thesis presents the spatial and statistical distributions of these quantities, and a decomposition of these fields into the contributions from different web components. We have studied the correlations between density and various velocity-related quantities, and followed the redshift evolution of parameters for the lognormal fits to the statistical distributions.

From these results, we find various forms of evidence for hierarchical evolution of cosmic structures; we determine the extent to which density and velocity divergence are correlated; we explore the formation and interactions between different structures that make up the cosmic web; we specifically probe the evolution of anisotropic structures; and follow the time evolutions of density, divergence, shear and vorticity. Lastly, we have identified a few artefacts resulting from the data and methods; and we assert the merit of the DTFE algorithm.

Contents

1 Introduction

The cosmic web

Our milky way — the spiral galaxy in which we spend all of our time — is only one of billions of galaxies that occupy the observable Universe. At far larger scales, the way matter is distributed in space is referred to as the *cosmic web*, the central subject in the study of cosmological structure formation (Bond et al., 1996). Redshift surveys have revealed that the Universe at Megaparsec (Mpc) scales is populated by regions of elevated density (Aghanim & Baccigalupi, 2015; Colless et al., 2003; York et al., 2000; Huchra et al., 2012; Saunders et al., 2000; Drinkwater et al., 2010) — where galaxies *cluster* together. These clusters form the *nodes* of the cosmic web. In the space between them, galaxies are gathered into filamentary structures that span from cluster to cluster. Nodes are also connected by sheet-like formations of galaxies, referred to as *walls* (Bond et al., 1996). In between these various types of dense features are the *voids* — vast regions of space where matter is sparse. These mostly empty regions dominate the Universe by volume, but contain only a minute fraction of all the matter in existence (e.g., Gregory & Thompson, 1978; Jõeveer et al., 1978; Kauffmann & Fairall, 1991; Rojas et al., 2005).

In reality, these weblike structures appear in various sizes. Across a very broad range of scales, the Universe is pervaded by a hierarchy of nodes, filaments, walls and voids (e.g., Kirshner et al., 1981; Jõeveer et al., 1978; Bond et al., 1996; Jenkins et al., 1998; Sheth & van de Weygaert, 2004; Colberg et al., 2005; Springel et al., 2005; Dolag et al., 2006; van de Weygaert & Schaap, 2009)[1]. For example, the tiniest filaments — or *tendrils* — consist of only a handful of galaxies, embedded in voids (Alpaslan et al., 2014). Only at scales upwards of the *homogeneity scale* does the Universe assume a homogeneous appearance. Estimates of this scale range from ~ 70 Mpc (Hogg et al., 2005; Sarkar et al., 2009; Scrimgeour et al., 2012; Sylos Labini et al., 2009) to several hundreds of Megaparsecs — see Jones et al. (2004) and references therein[2]. Nodes, filaments, walls and voids can be found throughout this whole hierarchy of spatial scales, the smaller objects embedded within larger ones. The formation and evolution of this hierarchical structure is the topic of investigation in this study.

Structure formation I — Gravitational instability

If we start with the assumption of a perfectly smooth initial matter distribution in the Universe, this induces a homogeneous gravity everywhere. Whatever the general expansion or contraction of the Universe may be, the relative position of any particle has no reason to change one way or another. However, this is an unstable equilibrium: an arbitrarily small local deviation from uniformity can be the seed for the growth of structure[3]. Overdensities attract mass, which in

[1]Computer simulations have also shown the formation of weblike structures, see (Doroshkevich et al., 1980; Melott, 1983; Pauls & Melott, 1995; Shapiro et al., 1983; Sathyaprakash et al., 1996)

[2]To add to the confusion around the maximum scale at which structures appear, Gamma-ray Bursts have been observed to be clustered at scales around 2000-3000 Mpc (Horváth et al., 2014).

[3]In fact, the initial conditions for structure formation are generally regarded to be the primordial quantum fluctuations, expanded to macroscopic scales by the cosmological *inflation* (Mukhanov & Chibisov, 1981; Guth & Pi, 1982; Hawking, 1982; Linde, 1982; Starobinsky,

turn further increase the density contrast. As long as pressure forces are insufficient to counteract this infall of matter, the overdensity can decouple from the general expansion of the Universe: it collapses into a gravitationally bound object (Icke, 1973; Peebles, 1980). Conversely, the gravitational force in an underdense region, surrounded by overdensities, will point outward. Matter will stream out of such regions, creating voids.

Initially, these deviations from homogeneous gravity are small, and so are the displacements of matter. As the unbound collapse of structures intensifies, much of the matter will be displaced more and more, in various directions. At relatively small scales, this will be noticeable relatively early on. At larger scales, matter displacement becomes significant at a later time, when the particle velocities have increased enough.

At any spatial scale, then, there is an initial epoch in which the displacement of matter is still limited, this period is referred to as the *linear regime* (e.g. Dekel, 1994, and references therein). This comprises a relatively large portion of the time in which structure forms, on any scale. From then onwards, the growth of structure will gradually turn *nonlinear* (e.g., Fry & Ma, 2001).

The time around which this happens at any spatial scale depends on the *power spectrum* of the spatial distribution of density (Peebles, 1980). Section 2.6 will treat this topic in more detail, but suffice it to say that — under the currently determined power spectrum index — the spatial distribution of matter is more 'clumpy' at small scales than at large scales. For that reason, the transition to nonlinear structure growth occurs later at higher spatial scales — at present day, the limiting spatial scale is roughly 8 Mpc. In other words: objects at small scales collapse earlier than large ones. This allows the growth of structure to be *hierarchical*[4] (Bond et al., 1996).

Structure formation II — Hierarchical interplay

The nodes, filaments and walls created at different scales are by no means static entities. Evolution of these structures is characterised by the interplay between various cosmic web components at various scales. For example, filaments act as matter conduits, feeding mass into the ever condensing clusters (see e.g. Summers, 1993; van Haarlem & van de Weygaert, 1993). From a certain point, *alignment* between minor sub-filaments tends to increase — due to the same anisotropic gravity fields that span filaments between clusters — causing those filaments to merge into larger-scale filaments (Bond et al., 1996; Aragón-Calvo et al., 2007b).

A complementary view to the picture of hierarchical collapse and merging of massive features is built upon voids. In this paradigm, the voids are seen as as key actors in the formation of structure. The dynamics of the empty regions has yielded valuable insights (see e.g. van de Weygaert (1991, 2002); Sheth & van de Weygaert (2004); Platen et al. (2008); van de Weygaert et al. (2010); Aragon-Calvo & Szalay (2013); Padilla et al. (2014); Sutter et al. (2014); Ceccarelli et al.

1982; Bardeen et al., 1983). The minute density fluctuations — of ~ 10 ppm in magnitude — which are probed by the *cosmic microwave background radiation* form the seeds for all structure growth (Smoot et al., 1992; Bennett et al., 2003; Spergel et al., 2007)

[4] A 'bottom-up' hierarchy of structures was previously hypothesised (Press & Schechter, 1974; Peebles, 1980) and competed with a 'top-down' picture of structure formation (Zel'dovich, 1970; Zel'dovich et al., 1982; Arnold et al., 1982; Klypin & Shandarin, 1983), where walls are the first structures to form, followed by filaments and then nodes.

(2015); Lambas et al. (2016) and many more sources).

Voids account for $\sim 95\%$ of the volume (Kauffmann & Fairall, 1991; El-Ad et al., 1996; El-Ad & Piran, 1997; Rojas et al., 2005). Some void regions, due to a deep depression in density, expand at a *super-Hubble* rate, exceeding the expansion of the Universe[5]. By number, the overwhelming majority of smaller sub-voids are squeezed between expanding voids and the massive components of the cosmic web. These voids may collapse along one or two axes. In this study, these cases will be referred to as *oblate* $\sim$ and *prolate collapsing void regions*, respectively.

When matter is squeezed between merging voids, this typically results in sheet-like features (Dubinski et al., 1993) — compressed in one direction by the voids, but expanding in the other two. Such scenarios are characterised by the *bending* of the voids' velocity outflows into that sheet. It is for this reason that we study anisotropic velocity flows, to probe the formation of anisotropic structures.

Structure formation III — Weaving the cosmic web

Similarly, anisotropic velocity flows are a major aspect of the formation of filaments.

When density peaks form and grow, they induce a multipolar gravity field. Between two density peaks, each of them acts as an attractor, while the surrounding underdense regions form depressions, effectively pushing matter away. In such scenarios, matter in the environment is encouraged to flow into the shaft directly between the density peaks (Bond et al., 1996). This is how filamentary bridges are formed, connecting neighbouring nodes[6]. Bond et al. (1996) have christened this mechanism the *weaving* of anisotropic structures between peaks in density. As such, the term *cosmic web* is very descriptive, not only of the appearance but also of the formation of large scale structure.

The collapse of gravitationally bound features is generally characterised by an increase in anisotropy. Even the most subtly aspherical bound regions have a major and a minor spatial axis. Collapse along the minor axis is augmented as the centre of attracting mass is closer to most particles along the minor axis than to most of those along the major axis. This progressively increases the eccentricity (van de Weygaert, 2006). Collapse along the shortest axis leaves a flattened object — *pancake-like*, in the official jargon. Further collapse along the second shortest axis forms an elongated feature.

Bond et al. (1996) have determined on theoretical grounds that peaks in density are the first features to emerge from nearly-homogeneous initial conditions. The filamentary features connecting them form afterwards.

Velocity flows

As the previous subsections aim to illustrate: the formation of various components of the cosmic web is characterised by various types of velocity flows. Due to gravitational instability, density peaks increase by the infall of matter, and underdense regions grow by a divergence of velocity flows. Anisotropic structures like filaments and walls form by the bending of velocity flows into

[5] This always happens in void regions with a flat density profile.

[6] Bond et al. state that walls form at a later stage, out of the rest of the matter left between voids.

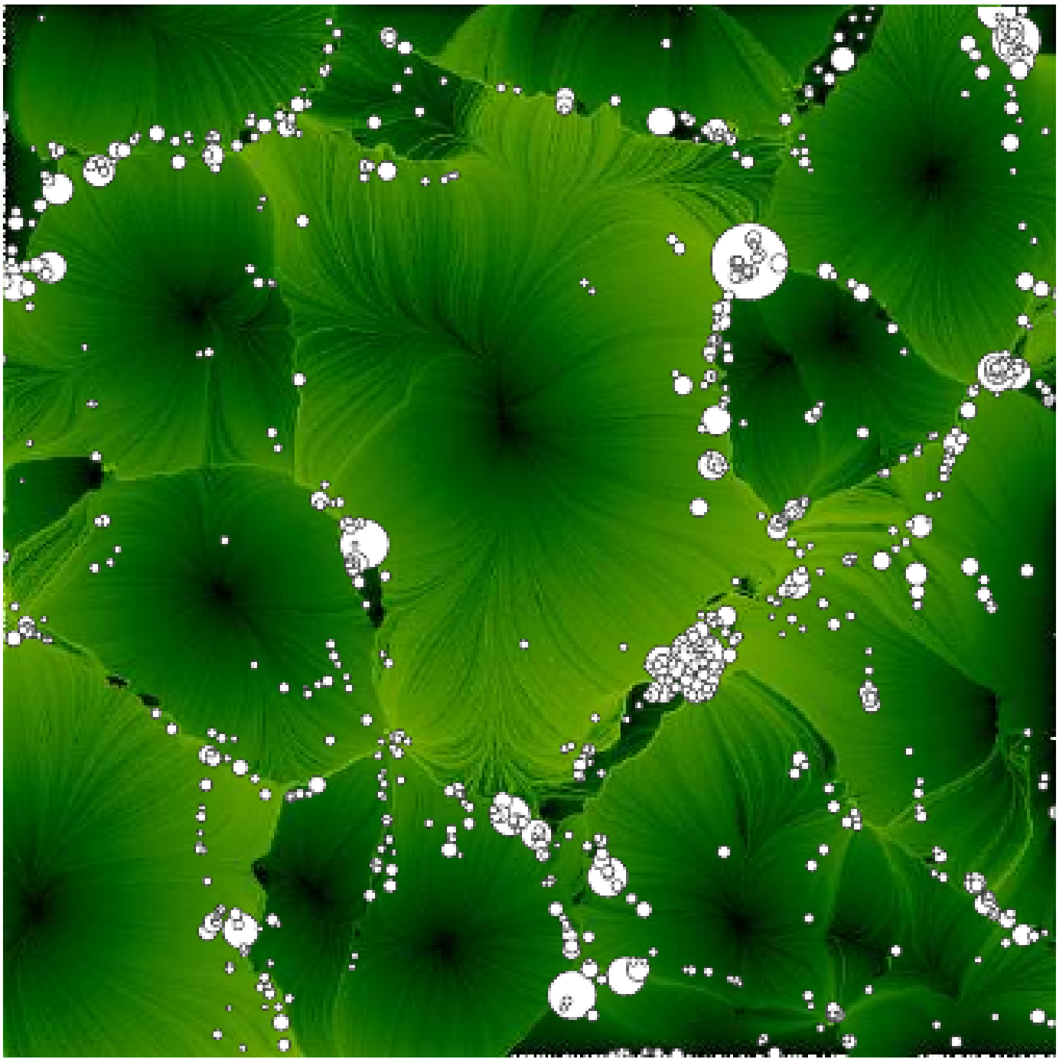


Figure 1: An example of the commonly occurring scenario where overdensities and underdensities erect a quadrupolar gravity field. Notice how matter from the environment is pulled into the linear region directly between the overdensities. Notice, also, the shearing motion of the matter flowing from the underdensities into the filamentary structure. This image was taken from Aragon-Calvo & Szalay (2013), who indicate dark matter halos from their simulation as white circles. They have used the *particle advection* technique for visualising the velocity stream lines.

anisotropic patterns. Therefore, this study approaches the formation and evolution of structure in the framework of velocity flows.

Whether considering the formation and hierarchical merging of massive components, or focusing on voids in a dynamical description, velocity flows — linear and nonlinear — are a key aspect of structure formation. The arrangement of matter into structures manifests in the form of velocity flows, and they have proven to be excellent probes of structure formation as well. Various theoretical frameworks have been formulated to link spatial distributions of matter to velocity flows, and velocity flows to the formation and evolution of cosmic structures — refer to Zel'dovich (1970); Peebles (1980) for seminal publications, and for more recent investigations e.g. to Nusser et al. (1991); Gramann (1993); Cautun et al. (2013); Libeskind et al. (2014). In these studies, important distinctions are made between linear and nonlinear velocity growth — see sections 2.3 and 2.5 — and between *potential* and *rotational* velocity flow components.

The most basic scenario in which a potential flow can be illustrated would be that of an isotropic density peak surrounding a sparser region. Here, the distribution of mass induces an inward-pointing *gravitational field*, and therefore a radial well in the *velocity potential* field. As a result, matter — provided zero initial rotation — falls radially into the dense region. In this scenario, the scalar density and velocity potential fields are proportional at any point in space, as are the vector gravity and velocity fields. In a more general example of potential flow, matter is not distributed isotropically, and these proportionalities no longer hold. Still, the velocity field is induced purely by the distribution of density. It can be expressed completely as the mathematical *gradient* of the scalar velocity potential. This means that the mathematical *curl* of the velocity field vanishes[7].

In an anisotropic density distribution, the resulting anisotropic gravitational force field will induce *shearing* motions of matter in the velocity flows: the gradient of the velocity potential field will generally curve into different directions, and trajectories of matter elements will be *bent*. In the context of the cosmological large scale structure, a particularly suitable example of this is given by a typical system of a filament connecting two clusters and flanked by two voids, see figure 1. Whereas matter around the clusters falls inward radially, due to depressions in the potential field, matter near the midpoint of the filament undergoes a *shearing* motion. This is because the filament crosses a *saddle point* in the potential field. Matter flowing out of the void generally travels in curved paths, navigating around that saddle point on way or the other.

Instances of shearing motion in the velocity field, then, betray the existence of anisotropic gravity fields — and therefore anisotropic structures. The main strategy in this study uses the formal mathematical definition of the velocity deformation — see section 2.7 — to detect anisotropic features in the data set. In the same vein, the velocity *divergence* is identified through mathematical methods, to follow the expansion and contraction of structures.

What distinguishes a potential flow from a velocity field with a non-zero *rotational component* is that rotating features like vortices do not appear. The picture of structure formation described so far leaves velocity flows without any

[7]Section 2.7 introduces the mathematical foundations for this subject, and section 2.2 provides a more detailed treatment.

rotational component. As for the evolution of rotational flow, the Kelvin circulation theorem[8] states that it is conserved: Under ordinary conditions, rotation can be neither created nor destroyed. In linear perturbation theory — see equation 31 — it is established that any existing vorticity in the linear regime of structure formation can only decay rapidly (Peebles, 1980). Starting from a potential flow, this leaves no possibility for rotation to occur.

To say that all velocity flows in the Universe are potential flows, however, would be too hurried a conclusion. As long as displacement of matter is linear on a certain scale, there will be no spatial overlap between several different velocity flows on that scale, and so potential flow is preserved throughout this stage. In the nonlinear regime, though, there can be significant crossing of matter streams. This occurrence is referred to as *shell crossing* (Mücket, 1985; Hellaby & Lake, 1985; Shandarin et al., 2012; Laigle et al., 2015), and section 2.9 discusses it in detail. The presence of various directional influences at the same location can produce *vorticity* (Pichon & Bernardeau, 1999; Pueblas & Scoccimarro, 2009). The detection of vorticity could have interesting implications for the study of cluster formation and perhaps even spiral galaxies. Yet this pursuit is not a trivial one: as Hahn et al. (2015) have pointed out, the vorticity signal determined from velocity flow measurements can in some part consist of a *projection effect* between the different streams, rather than actual vorticity — see the discussion leading up to equation 96.

Outline of this thesis

In this thesis, we document our investigations of velocity flows in the Cosmogrid simulation (Ishiyama et al., 2013). The rest of the present section will further introduce the goings-on in this field of study. Then, section 2 presents the theoretical background that this study is founded upon. The more strategical issues of simulation and field estimation are introduced in sections 3 and 4, respectively.

Section 5 traces out the practical procedures this study employed, the results of which are presented in section 6. Section A compiles the relevant statistical certainties and philosophical caveats, and section B presents a brief overview and distills the conclusions from this study.

1.1 Cosmological context

The expanding Universe

Of the four fundamental forces of nature, it is undisputedly gravity that dominates on large scales. An evaluation of the dynamics of the Universe, then, relies on a proper general relativistic treatment of gravity. In General relativity, gravity is a metric force, determined by the curvature of spacetime according to the *Einstein field equations.*

Given the Universe's principal homogeneity and isotropy, there are only three possible metrics it can assume. These are all isotropic and homogeneous, they expand or contract in time at some rate a, and exhibit one of three possible

[8] This theorem applies to fluids. The approach of matter in the universe as a *cosmic fluid* is introduced in the theory of gravitational instability by Peebles (1980) — see section 2.2. The conservation of circulation is limited to systems without anisotropic stresses. These are generated in *shell crossing*, see section 2.9.

types of curvature k. The expansion factor $a(t)$ describes the rate of expansion of the Universe, and is normalised at present day t_0 to be $a(t_0) = 1$. This means that any *proper* position $\mathbf{r}$ can be written in terms of the expansion factor and its *comoving* position $\mathbf{x}$ by:

$$\mathbf{r} = a\mathbf{x}. \tag{1}$$

The spherically symmetric spatial part of the metric requires a uniform curvature, described by $k > 0$ for a *closed* Universe, $k = 0$ for a *flat* one, or $k < 0$ for an *open* Universe. Then, the *Robertson-Walker metric* gives the spacelike spacetime interval ds in spherical coordinates (r, θ, ϕ):

$$ds^2 = c^2dt^2 - a^2(t)\left(dr^2 + R_c^2 S_k^2\left(\frac{r}{R_c}\right)\left[d\theta^2 + sin^2(\theta)d\phi^2\right]\right), \tag{2}$$

where c is the speed of light, R_c is the radius of curvature, and the curvature function $S_k\left(\frac{r}{R_c}\right)$ is given by:

$$S_k\left(\frac{r}{R_c}\right) = \begin{cases} \sin\left(\frac{r}{R_c}\right) & k = +1 \\ \frac{r}{R_c} & k = 0 \\ \sinh\left(\frac{r}{R_c}\right) & k = -1 \end{cases} \tag{3}$$

Armed with this metric and the Einstein field equations, which relate the local spacetime curvature to the local energy density, Friedmann derived a pair of seminal equations in 1922. Relating the time evolution of the expansion factor to the density ρ, pressure p and the cosmological constant Λ, the *Friedmann-Robertson-Walker-Lemaître* equations:

$$\begin{cases} \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda}{3} \\ \frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2} + \frac{\Lambda}{3}. \end{cases} \tag{4}$$

The quantity $\frac{\dot{a}}{a}$ is the *Hubble parameter* H,

$$H \equiv \frac{\dot{a}}{a},$$

which relates the distance by which galaxies are separated from us to rate of expansion. The discovery of this — to first order — linear relationship by Edwin Hubble, only in 1929, was the first confirmation of the theory of an expanding Universe. Just as equation 1 relates the proper and comoving positions of a point, its physical velocity is:

$$\dot{\mathbf{r}} = H\mathbf{r}. \tag{5}$$

Since the scale factor relates the proper position of a particle to its comoving position, its time derivative $\dot{a}$ relates the physical velocity to that same proper position. Its dimensionality is thus velocity times reciprocal distance. The Hubble parameter is thus expressed in km/s/Mpc. As Georges Lemaître has pointed out — prior to Hubble's observations! — these relations imply that the Universe has started as an extremely hot and dense entity. The Big Bang.

All formation of structure on cosmological scales happens on the background of our expanding Universe. These processes are governed by gravity. Above the scales of homogeneity, even the largest structures can no longer be discerned. The Universe then appears like a homogeneous and isotropic body (Hogg et al., 2005; Sarkar et al., 2009; Scrimgeour et al., 2012; Sylos Labini et al., 2009). The *cosmological principle* is still valid at smaller scales, though, in the sense that the probabilistic spatial distribution of matter is uniform throughout space (Bardeen et al., 1986).

1.2 Investigations of structure formation and velocity flows

So far, the study of structure formation has branched out into various strategies, and various tools have been developed. A full analytical description of structure formation has proven to be a challenge beyond our capabilities, but various *perturbation theories* — applied to the theory of gravitational instability — have yielded very useful approximations (Peebles, 1980, 1993). This study will be limited to linear perturbation theory, which is applicable throughout the linear and mildly nonlinear regime.

Linear perturbation theory

The essence of perturbation theory is that tiny variations — *perturbations* — in a variable can be approximated e.g. as a linear or higher order deviation from its base value. Any higher order power of this small deviation will then be negligible, as will any higher order product of various perturbation quantities. This allows us to remove higher order expressions from equations, a potentially crucial simplification that holds for as long as the perturbations in these quantities stay sufficiently close to zero. Section 2.3 presents the linear Eulerian version and section 2.5 the Lagrangian perturbation theory.

This is an important reason why we are interested in the distinction between linear and nonlinear phases of structure formation. In the linear phase — during which most of structure formation occurs — perturbation theory can facilitate a very reliable treatment of the evolution of quantities like density, gravity and velocity. As the growth of physical quantities becomes more and more nonlinear, approximations from lower-order perturbation theories will get shoddy. This thesis will review the application of linear perturbation theory to two complementary views of fluid mechanics: the Eulerian and Lagrangian ones. The former considers the time evolution and thoroughfare of quantities in a stationary location; the latter travels along with a fluid element and describes its motion and deformations.

Simulations

Efforts towards analytical descriptions are complemented by numerical strategies — N-body simulations. Cosmological N-body simulations have been around for decades, and have been unhalting in their technical improvements.

Redshift surveys have yielded large quantities of data to form a basis of structure formation research (e.g. Saunders et al., 2000; York et al., 2000; Colless et al., 2003; Tegmark et al., 2004; Huchra et al., 2005; Drinkwater et al., 2010; Huchra et al., 2012; Aghanim & Baccigalupi, 2015), but they are not without limitations. Our field of view is cluttered with foreground objects and effects; fainter galaxies are far less visible at greater distances; and *redshift distortions* arise from the degeneracy between cosmological expansion and proper motion of galaxies (see Lonsdale & Barthel (1986); Hamilton (1998) and references therein). Due to this last effect, the distribution of galaxies in cosmic web formations is enhanced differently in radial and azimuthal directions[9]. This brings immediate complications and uncertainties into structure formation studies.

The ever increasing power and availability of computation have been a catalyst for the *simulation* of structure formation. N-body simulations start with a large number of particles in their *initial conditions*: they typically sit in a regular grid, and are given a Gaussian random[10] initial displacement and velocity. From there, particle displacements are calculated in iterative evaluations of the adopted equations of motion. Section 3 provides a more complete description of these methods. More and more complex and realistic simulations have been conducted in the recent decades — see references in section 3 — and they circumvent the problems of field of view, visibility and redshift distortions. Trading "fingers of God" for "eyes of God", simulations allow the locations and velocities of particles to be determined to arbitrary precision. Furthermore, as simulations trace out the time evolution of structures in all of space in parallel, we can follow the evolution of one and the same structure throughout the course of cosmic time. This is not possible through observations, where a view of an earlier time frame — i.e. higher redshift — necessitates a focus to greater distances.

It has to be stated very clearly, though, that simulation is by no means a replacement for observation. Most importantly, simulations will never provide data on what the Universe really looks like. They only evaluate structure formation under the theoretical assumptions that they are built upon. As such, it would be more suitable to see them as test Universes for our theories — a simulation is only as useful as the theoretical framework it incorporates. Reality is an incredibly complex system: although most of structure formation is well modelled by gravity alone, the influences of myriad physical components like pressure, radiation, dark energy, general relativity, magnetism, etc. are very real — see section 3. To equip a simulation with a proper implementation all of these mechanisms is beyond the limits of our abilities. Furthermore, there are various nontrivial complications inherent to the way simulations operate. One prominent example of this is the discretised nature of the space and time that they define. While the particle locations and velocities can be defined to machine precision, the densities and gravitational forces are only calculated at regular points in a grid of some resolution (Hockney & Eastwood, 1981; Barnes & Hut, 1986). Among other complications, this leads to errors in the calculation of particle displacements. Sections 3 and A treat these limitations in more detail.

Even so, N-body simulations have led to invaluable insights. They form an extremely useful complementary approach to the study, in parallel to observations. This project is founded upon the results of the Cosmogrid simulation (Ishiyama et al., 2013), which simulates the formation of structure in a volume of 30 Mpc, by means of 2048^3 dark matter particles, in a standard ΛCDM cosmology.

[9] One prominent example of redshift distortions is the radial elongation of galaxy clusters due to the spread in radial velocity components, which induce a range of spectral shifts on top of their cosmological redshift. This radial elongation effect is referred to as "fingers of god".

[10] see section 2.6 for an introduction

Field estimation

At any given time frame in a simulation, the locations and velocities of all the particles can be listed. One of the most important steps in the analysis of simulation data is to distill from these locations and velocities a field — e.g. a density or velocity field — which is sampled at a regular set of grid points. This step is called *field estimation*, and there is a variety of approaches to this — e.g. Hockney & Eastwood (1981); Bernardeau & van de Weygaert (1996); Schaap & van de Weygaert (2001); van de Weygaert & Schaap (2009). The present study appeals to a method that has gained particular favour in the recent years: the *Delaunay Tessellation Field Estimator* (Bernardeau & van de Weygaert, 1996) — accounts of its nature, modus operandi, and assessment can be found in sections 4.4, 4.5 and A.

Separating spatial scales

As a last introductory note, there is a convenient way to disentangle the matter and velocity distributions from different distance scales: This is done by the process of *filtering* the spatial distribution at a scale of choice. The procedure is explained in section 4.1, and has the effect of smoothing out all structure at all scales below a chosen value. This allows for the analysis of structure formation at various scales.

By virtue of the hierarchical nature of the structure of the Universe, the evolution of a region at any particular scale is dependent only on the distribution on that scale and above — i.e. it is not necessary, to determine the activity on all finer scales (Peebles, 1980; Little et al., 1991). This is a very fortunate fact, as an accurate reconstruction of all smaller scale structures would border on the impossible.

2 Theoretical Background

2.1 Overview

This section aims to provide a description of the theoretical background of relevance to this study. We start with a treatment fluid dynamics in general instability — section 2.2 — which forms the very basis of studies in cosmic structure formation. Here, we will stumble upon an obstacle in solving the fluid equations; this is solved in a *linear perturbation theory* — section 2.3 — by linearising the equations. Then, section 2.5 makes a switch from an *Eulerian* perspective to a *Lagrangian* one. We formulate *Lagrangian perturbation theory* and the *Zel'dovich approximation* in an attempt at describing the nonlinear evolution of structure. Sections 2.2, 2.3 and 2.5 are largely based on (Peebles, 1980, 1993; Zel'dovich, 1970) and lecture notes by Rien van de Weygaert.

Next, a framework for the statistics of various field quantities is developed in section 2.6. Here we introduce *Gaussian random fields* (Bardeen et al., 1986), and an important statistical descriptor, the *power spectrum*, which is of importance to the way structures evolve over cosmic time. The related view of *lognormal random fields* (Coles & Jones, 1991) is introduced too.

Sections 2.7 and 2.8 provide an overview of velocity flows, where the *velocity gradient* is decomposed into *divergence*, *shear* and *vorticity*. Density- and velocity-related fields are used for the disentanglement of different cosmic web components — i.e. nodes, filaments, walls and voids — in the so-called *T-web* and *V-web* classifiers. Section 2.8 traces a conceptual genealogy of the web classification algorithm used in this study. The most important publications in the history of this topic are (Hahn et al., 2007; Forero-Romero et al., 2009; Hoffman et al., 2012; Cautun et al., 2013).

The role vorticity in structure formation is explored in section 2.9. We get a taste of the intricacies and nontrivialities in the study of vorticity in cosmic velocity flows (Pichon & Bernardeau, 1999; Pueblas & Scoccimarro, 2009; Hahn et al., 2015).

Following this Theoretical Background section, section 3 introduces the computational world of cosmological N-body simulations. Various approaches to the task of *field estimation* are covered separately in section 4.

2.2 Cosmic structure formation

The fluid approximation

While, fundamentally, the distribution of matter in the Universe is discretised, we approximate it as a continuous *cosmic fluid.* Whether this approximation is reliable depends on the nature of the matter distribution: For the baryonic component, the discrete nature of particulate matter is of importance only at microscopic levels, until the formation of condensed objects like stars and galaxies — which forbid a fluid approximation at scales below $\sim$ a hundred kpc at lower redshifts. For dark matter, the limiting scales remain very small during all of cosmic history. Our study of velocity flows and structure formation is limited to dark matter, and concentrates on Megaparsec scales, so that a fluid approximation is very adequate.

Approximating the cosmic matter distribution as a fluid allows us to describe its evolution in the *fluid equations* (Peebles, 1993). These equations describe properties of the fluid — e.g. density and velocities — in the context of *gravitational instability*: the spatial distribution of density induces a *gravitational potential*, which in turn influences the velocity field. Note that a full description must include pressure forces between matter elements. However, these are not of influence on the dynamics of dark matter at the scales in this study. Therefore, pressure terms will be left out of the treatment in this thesis.

Thus, the physical quantities of interest are position $\mathbf{r}$, momentum $\mathbf{p}$, density ρ, gravitational potential Φ and gravitational acceleration $\mathbf{g}$.

Fluid equations

A fluid can be generally described by the *phase space distribution function* $f(\mathbf{r}, \mathbf{p}, t)$, which follows the distribution of matter in physical space $\mathbf{r}$, momentum space $\mathbf{p}$ and time t. Emerging from the conservation of energy and mass, the *Vlasov equation* describes the evolution of the distribution function:

$$\frac{\partial f}{\partial t} + \mathbf{p} \cdot \nabla f - \nabla\Phi \cdot \frac{\partial f}{\partial \mathbf{p}} = 0. \tag{6}$$

Taking the first cumulants of this equation — i.e. multiplying it by the first powers of momentum and integrating over momentum space — yields equations that govern the density 7, velocity 8 and the gravitational field 9. These three equations are widely applied in the study fluid dynamics:

Firstly, the *continuity equation* ensures conservation of mass:

$$\frac{\partial \rho}{\partial t} + \nabla_r \cdot \rho\mathbf{u} = 0 \tag{7}$$

Secondly, the *Euler equation* links the gravitational forces acting on a mass element to its velocity:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\nabla_r \Phi \tag{8}$$

Lastly, the *Poisson equation* determines the gravitational field as sourced by the matter distribution.

$$\nabla_r^2 \mathbf{\Phi} = 4\pi G \sum_l (1 + 3w_l)\rho_l \tag{9}$$

Do note two things about this system of equations:

- We assume dark matter to be pressureless and ignore other sources of pressure. In a more complete description, pressure would appear as a term coupled with density in the Euler equation. Particularly in the matter-dominated era — during which most of structure formation occurs — pressure forces are negligible. Also, these equations leave out magnetic fields on cosmic scales, and the minute general relativistic effects of radiation and dark energy — after all, $\Omega_{r,0} \simeq 10^{-5}$.

- Equations 7, 8 and 9 comprise what we call the *Eulerian* picture of the fluid equations. Section 2.5 will present a complementary view, referred to as the *Lagrangian* view.

Comoving coordinates

Equations 7, 8 and 9 describe the evolution of fluid elements in physical coordinates. In an expanding Universe, it is useful to switch to a system of *comoving* coordinates. If at any given time, the Universe has expanded by a factor $a(t)$ relative to its size at the present day

$$a(t_0) \equiv a_0 \equiv 1,$$

the *comoving positions* $\mathbf{x}$ of an element will be

$$\mathbf{x} \equiv \frac{\mathbf{r}}{a(t)}. \tag{10}$$

This way, a comoving coordinate system allows us to describe the displacement of fluid elements with respect to the expanding background (Peebles, 1993) — a successful description of this displacement equals a description of structure formation.

The procedure is to write the quantities — density, velocity, gravitational potential and gravitational acceleration — in terms of their deviations from the cosmic background. These deviations are called *perturbations*. The *density perturbation* δ is defined in terms of its deviation from the universal density[11] $\rho_u(t)$:

$$\delta(\mathbf{x}, t) \equiv \frac{\rho(\mathbf{x}, t) - \rho_u(t)}{\rho_u(t)} \tag{11}$$

note that this confines δ to $[-1; \infty\rangle$. The *peculiar velocity* $\mathbf{v}$ is the change in comoving position:

$$\mathbf{v} = a(t)\dot{\mathbf{x}}, \tag{12}$$

and as such describes velocity relative to the Hubble expansion.

The gravitational potential can be written as the sum of a background potential Φ_u due to the Hubble expansion plus the perturbative gravitational potential ϕ. The potential perturbation is thus given by

$$\phi(\mathbf{x}, t) = \Phi(\mathbf{x}, t) - \frac{1}{2} a\ddot{a}x^2. \tag{13}$$

The peculiar gravitational acceleration is the gradient of the potential perturbation, and related to the peculiar velocity:

$$\mathbf{g}(\mathbf{x}, t) \equiv -\frac{\nabla\phi}{a} \tag{14}$$

Rewriting these equations for perturbation quantities in comoving coordinates, they become[12]:

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla_x \cdot \underbrace{(1+\delta)\,\mathbf{v}} = 0 \tag{15}$$

$$\frac{\partial\mathbf{v}}{\partial t} + \underbrace{\frac{1}{a}(\mathbf{v}\cdot\nabla_x)\mathbf{v}} + \frac{\dot{a}}{a}\mathbf{v} = -\frac{1}{a}\nabla_x\phi \tag{16}$$

[11] We will assume the adiabatic perturbation mode, in which the density of matter and radiation are completely coupled, although other options exist.

[12] The $\underbrace{\text{horizontal braces}}$ are included to mark the nonlinear couplings between perturbation quantities. These are simplified in section 2.3.

$$\nabla_x^2 \phi = 4\pi G a^2 \rho_u \delta \tag{17}$$

Here, the operator ∇_x refer to differentiation in comoving space — as opposed to the derivatives in physical space, in equations 7, 8 and 9.

As a last adaptation, it may be noted that the comoving perturbative Euler equation can be simplified by switching to an alternative time-differentiation. The *Lagrangian derivative*[13] $\frac{d}{dt}$ contains a time derivative, but also travels along with the trajectory of a fluid element. Acting upon a field $f(\mathbf{x}, t)$, the comoving Lagrangian derivative follows:

$$\frac{d}{dt} f(\mathbf{x}, t) \equiv \frac{\partial}{\partial t} f(\mathbf{x}, t) + \frac{1}{a}(\mathbf{v} \cdot \nabla) f(\mathbf{x}, t) \tag{18}$$

Upon inspection, the first two terms of the euler equation 16 can be replaced by $\frac{d}{dt}\mathbf{v}$, and we can even simplify even further:

$$\frac{d}{dt}(a\mathbf{v}) = -\nabla\phi \tag{19}$$

Notice that the Euler equation also equals the gravitational acceleration:

$$-\nabla\phi = a\mathbf{g}(\mathbf{x}, t). \tag{20}$$

2.3 Linear Perturbation Theory

The Fluid equations provide a good framework to study the flow of matter and the formation of structure. However, they contain terms that are nonlinear combinations of perturbative quantities, which proves to be an immense complication in solving them. To be specific, the expressions marked with $\underbrace{\text{horizontal braces}}$ render an analytical solution to the full perturbative fluid equations impossible — with known analytical methods.

There is, however, a practical workaround to this unfortunate complication: Linear Perturbation Theory. In the linear regime, perturbations are very small — i.e. $\delta << 1$ — and that means that higher order combinations of them are negligible. The essence of linear perturbation theory lies in the omission of the nonlinear combinations, which greatly simplifies the expressions, and allows an approximate analytical description of structure formation for as long as the physical quantities evolve linearly.

The description of linear perturbation theory in this section originates from seminal work by Peebles (1980, 1993).

Linear solutions

The matter dominated epoch is the most instrumental phase of structure formation, and most matter in the Universe appears to be *collisionless* dark matter. Therefore, our initial evaluation of structure growth will start with the fluid equations for matter perturbations, ignoring pressure effects. A *linearisation* of the continuity and Euler equations — the Poisson equation 17 remains unchanged — yields:

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla_x \cdot \mathbf{v} = 0; \tag{21}$$

[13] a.k.a. *convective* derivative

and

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a}\mathbf{v} = -\frac{1}{a}\nabla_x \phi. \tag{22}$$

With a bit of algebraic persuasion, these linearised equations may be manipulated into a second order partial differential equation for the growth in δ:

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = \frac{3}{2}\Omega_0 H_0^2 \frac{1}{a^3}\delta. \tag{23}$$

This equation contains no spatial derivatives, meaning that the evolution is independent of location, and so any solution can be separated into independent spatial $\Delta(\mathbf{x})$ and temporal $D_\delta(t)$ components. Two solutions exist,

$$\delta(\mathbf{x}, t) = D_{\delta,1}(t)\Delta_1(\mathbf{x}) + D_{\delta,2}(t)\Delta_2(\mathbf{x}), \tag{24}$$

where $\Delta(\mathbf{x})$ describes the spatial distribution of density perturbations, and the *density growth factor* $D_\delta(t)$ plays the same role[14] as δ in equation 23.

The actual evolution of $D_\delta(t)$ depends on the background cosmology. When the time evolutions of

$$2\frac{\dot{a}}{a} \qquad \text{and} \qquad \frac{3}{2}\Omega_0 H_0^2 \frac{1}{a^3} \tag{25}$$

are determined and inserted into equation 23, two solutions for $D_\delta(t)$ can be found.

For example, an Einstein-de Sitter (EdS) Universe is characterised by $\Omega(t) = 1$ and $H(t) = H_0$, so that the expansion factor follows

$$a(t) = \left(\frac{3}{2}H_0 t\right)^{2/3}. \tag{26}$$

This implies that the $a(t)$-dependent expressions 25 become

$$\frac{2}{3t} \qquad \text{and} \qquad \frac{2}{3t^2},$$

respectively. The resulting differential equation for $\delta(t)$ will then have the temporal solutions

$$\begin{cases} D_1(t) \propto t^{2/3}; \\ D_2(t) \propto t^{-1}. \end{cases} \tag{27}$$

A density growth factor $D_\delta(t) \propto t^\alpha$ for $\alpha < 0$ comprises a *decaying mode solution*, and will result in the extinction of structure formation. A density growth factor for which $\alpha > 0$ will only increase, and such a *growing mode solution* will dominate structure formation. In the EdS Universe example, we see that $D_1(t)$ grows, and $D_2(t)$ decays. Thus, it is usually sufficient to distill the linear solution to the fluid equations down to only the growing mode $D_\delta(t)$.

We proceed to the growth factor for gravitational potential $D_\phi(t)$. The

[14] $D_\delta(t)$ being the sole temporal component of $\delta(t)$, and equation 23 being a purely temporal PDE

Poisson equation for the potential perturbation ϕ is completely determined by matter density perturbations.

$$\nabla^2\phi = \frac{3}{2}\Omega_m H^2 a^2 \delta_m(\mathbf{x}, t),$$

which can be solved by Greens's formula. Time-dependence enters via $\Omega_m H^2 a^2$ and δ_m. In the linear regime, $\delta_m(\mathbf{x}, t) \propto D_\delta(t)$, and for the potential perturbation, the growth factor $D_\phi(t)$ holds:

$$\phi(\mathbf{x}, t) = D_\phi(t)\phi_0(\mathbf{x}) = \frac{D_\delta(t)}{a(t)}\phi_0(\mathbf{x}), \tag{28}$$

so we see that $D_\phi(t)$ differs from $D_\delta(t)$ only by a factor $a(t)$.

The gravitational acceleration perturbation $\mathbf{g}(\mathbf{x}, t) \equiv -\frac{\nabla\phi}{a}$, and can thus be related to the potential perturbation. In a very similar fashion, the *gravity growth factor* $D_g(t)$ can be determined:

$$D_g(t) = \frac{D_\delta(t)}{a^2(t)}.$$

In most cases, while $D_\delta(t)$ keeps increasing, it is the hubble expansion that causes the peculiar gravity to decrease nonetheless.

Velocity growth

In the general picture of structure formation, matter flows out from underdense regions, and towards overdense regions. Figure 2 — from a study by Stanonik et al. (2009) — provides a visual explanation, it displays velocity flow lines and density contours from an N-body simulation. This image makes clear how the velocity flows stretch from low to high density regions, and bend along with the anisotropic density distributions.

To study the evolution of peculiar velocity $\mathbf{v}$, recall the Euler equation in terms of the Lagrangian derivative of $a\mathbf{v}$, equation 19. The Euler equation — $\frac{d}{dt}(a\mathbf{v}) = -\nabla\phi$ — incorporates the net effect on velocity against the background of an expanding Universe. Note that the latter term in $\frac{d}{dt}(a\mathbf{v})$ is nonlinear:

$$a(\mathbf{v}\cdot\nabla)\mathbf{v},$$

so that it can be omitted in linear theory. The linearised Euler equation can then be written with just a partial time derivative:

$$\frac{\partial(a\mathbf{v})}{\partial t} = -\nabla\phi. \tag{29}$$

Presently, we use two vector identities. The curl of the gradient of any scalar field always vanishes, as does the divergence of the curl of any vector field:

$$\nabla \times (\nabla f) = \mathbf{0};$$

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0.$$

Therefore, any vector field can be written in terms of a component with zero curl — the *potential* component $v_\parallel = \nabla f$ — and a component with zero divergence — the *rotational* component $v_\perp = \nabla \times \mathbf{A}$. If we express peculiar velocity as

$$\mathbf{v} = \mathbf{v}_\parallel + \mathbf{v}_\perp,$$

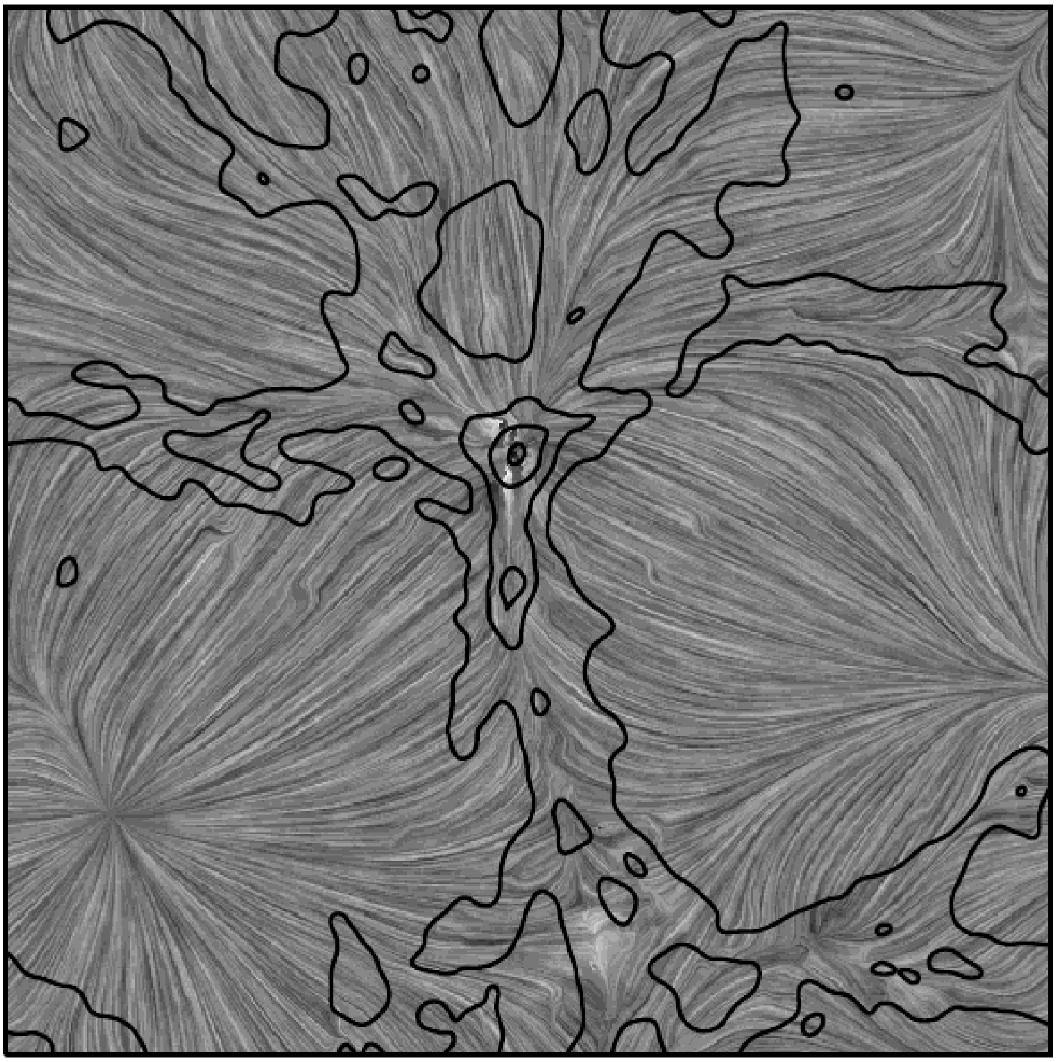


Figure 2: Velocity flow lines from an N-body simulation (Stanonik et al., 2009). Superimposed are the density contours. Notice how velocity flows point outward from the underdensities; bend into filamentary and wall-like structures, and concentrate to density peaks. Image source: (Stanonik et al., 2009).

then the Euler equation can be split into two components; and the aforementioned identities yield that:

$$\begin{cases} \frac{\partial}{\partial t}(a\mathbf{v}_{\parallel}) & = -\nabla\phi \\ \frac{\partial}{\partial t}(a\mathbf{v}_{\perp}) & = 0 \end{cases} \tag{30}$$

As a consequence, in the linear regime the rotational component will always decay:

$$\mathbf{v}_{\perp} \propto \frac{1}{a}. \tag{31}$$

For this reason, vorticity (see section 2.7) does not occur at the spatial and temporal scales that comprise the linear regime. For example, notice that no rotational flow occurs in figure 2.

Now that velocity and gravitational acceleration are both gradients of a potential, their relation can be found via the Poisson equation. Keeping in mind that $\mathbf{g}(t) \propto D_g(t) \propto \frac{D_\delta(t)}{a^2}$:

$$\mathbf{v} = a\frac{\partial}{\partial t}\left(\frac{\mathbf{g}}{a\pi G\rho_u a}\right) = \frac{1}{D_\delta}\frac{dD_\delta}{dt}\left(\frac{\mathbf{g}}{4\pi G\rho_u}\right), \tag{32}$$

where we can take $\frac{1}{D_\delta}\frac{dD_\delta}{dt}$ to be

$$H(t)\frac{a}{D_\delta}\frac{dD_\delta}{da} \equiv Hf. \tag{33}$$

We have just defined

$$f \equiv \frac{a}{D_\delta}\frac{dD_\delta}{da}, \tag{34}$$

the *linear velocity growth factor*, and it is of extreme importance within linear perturbation theory. It directly relates velocity to gravitational acceleration:

$$\mathbf{v} = \frac{Hf}{4\pi G\rho_u}\mathbf{g}. \tag{35}$$

It has been approximated as a function of Ω_m. For Ω_m below unity, an extremely good approximation is given by (Linder, 2006)

$$f(\Omega_m) \simeq \Omega_m^\gamma, \tag{36}$$

where

$$\gamma = 0.55 + 0.05(\omega + 1),$$

and ω is the equation of state parameter for the dominant component.

From the linearised continuity equation, we can derive the growth factor for velocity D_v in a matter-dominated Universe:

$$D_v(t) = aD_\delta H f(\Omega_m). \tag{37}$$

Via equation 35, the dimensionless velocity growth factor $f(\Omega_m)$ provides a way to observationally determine Ω_m!

From the linearised continuity equation, we can derive an immediate relation between density perturbation and velocity divergence. In terms of the velocity growth factor. This is done by establishing the proportionality

$$\frac{\partial \delta}{\partial t} = \frac{\dot{D}_\delta}{D_\delta}\delta,$$

where

$$\frac{\dot{D}_\delta}{D_\delta} = Hf(\Omega),$$

yielding

$$\delta = -\frac{\nabla \cdot \mathbf{v}}{Haf(\Omega).} \tag{38}$$

2.4 Nonlinearity

In describing the growth of structure outside the linear regime[15], the approximations made in linear perturbation theory are no longer warranted. Perturbation quantities approach and exceed unity, and *power transfer* starts to occur. Expressions that are linear functions of perturbation quantities — e.g. linear terms in the fluid equations — have a property that is very important for the formation of structure: their evolution on any distance scale is completely independent from that on every other scale. This becomes apparent when the evolution of these quantities is expressed in Fourier transforms: the Fourier modes **k** are independent from each other. See, for example, the linearised continuity equation in Fourier space:

$$\frac{d}{dt}\hat{\delta}(\mathbf{k}) - \frac{1}{a} i\mathbf{k} \cdot \hat{\mathbf{v}}(\mathbf{k}) = 0 \tag{39}$$

In contrast, the nonlinear terms in the fluid equations — marked with $\underbrace{\text{vertical braces}}$ in equations 15 and 16 — lack this Fourier mode independence. Their Fourier transforms involve the mixing of Fourier modes from different perturbation quantities. For example, the full nonlinear Fourier version of equation 39 acquires a *mode-coupling term*:

$$\frac{d}{dt}\hat{\delta}(\mathbf{k}) - \frac{1}{a} i\mathbf{k} \cdot \hat{\mathbf{v}}(\mathbf{k}) - \frac{1}{a}\int \frac{d\mathbf{k}'}{(2\pi)^3} i\hat{\delta}\mathbf{k}' \cdot \hat{\mathbf{v}}(\mathbf{k} - \mathbf{k}') = 0 \tag{40}$$

The physical consequence of that nonlinear term — which comes into view as soon as $\hat{\delta}$ and $\hat{\mathbf{v}}$ approach or exceed unity — is that the Fourier contributions from different wavenumbers start influencing each other. A concrete example of this is the nonlinear collapse of high density regions, where the Fourier modes corresponding to the spatial extent of the peak are enhanced. Another important example is the influence from large scale structures onto small scale structures embedded into them: e.g. a small overdensity embedded in a large overdense region will collapse more quickly.

The collapse of large-scale structures influences the growth of smaller-scale substructures. If the opposite would happen, small-scale structures would affect their surroundings — there would be transfer from high to low frequency modes. In that case, it would introduce the necessity that a feature's substructure be known completely in order to determine its evolution. Fortunately, this does not

[15] id est: later epochs and smaller distance scales

occur if the steepness of the fluctuation *power spectrum* is low enough (Peebles, 1980; Little et al., 1991),

$$n(k) = \frac{d \log P(k)}{d \log k} < 4. \tag{41}$$

2.5 Lagrangian and Zel'dovich theory

The Eulerian and Lagrangian approaches

Faced with the generic task of describing the motions of energy and matter in a certain volume, the Eulerian approach is to consider the evolution of any location, fixed in space. This evolution consists of the local changes in the system, and of the flow of energy and matter through the element under scrutiny.

The fluid equations, as presented in sections 2.2 and 2.3, are a clear example of Eulerian thinking. The continuity equation links the local density to the in- and outflow of matter; the Euler equation links the local velocity to changes in the local gravity potential; and the poisson equation links that to the local density.

As an alternative to the Eulerian philosophy, the Lagrangian approach is a more dynamic way of monitoring the evolution of a system. The core idea is to travel along with an element of the fluid, and describe the time evolution of its location and deformation. To this end, it is convenient to follow the evolution of a moving fluid element with the Lagrangian derivative — equation 18 — rather than just a partial time-derivative. A few aspects of the Lagrangian perspective of fluid dynamics will be formulated momentarily; and in it, the forces acting on a fluid element become more readily apparent.

The deformation of a fluid element can be described by the three modes of the spatial velocity derivative, introduced in section 2.7. The divergence measures the expansion or compression of a fluid element, and consequently changes in density. The shear tensor measures deformations in different directions: its eigenvectors describe the principal axes, and corresponding eigenvalues measure the strength of deformation along them. The overall rotation of a fluid element is measured by the vorticity. Two-dimensional equivalents of these three modes of deformation are illustrated in figure 3

Lagrangian fluid equations

Initially, our Lagrangian description of a pressureless fluid dynamics assumes pure laminar flow, in which shell crossing — see section 2.9 — does not occur. This description consists of a Lagrangian version of the continuity, Euler and Poisson equations, as well as equations governing the divergence, vorticity[16] and shear of fluid elements.

While section 2.7 elaborates on the different modes of the velocity gradient that exist, it is useful to introduce them briefly at this point. The velocity gradient $\nabla \mathbf{v}$ is a 3×3 tensor, which can be decomposed into three components: a scalar *divergence* θ, a tensor *shear* σ and a tensor *vorticity* ω. These adhere

[16]The vorticity equation is not treated in this thesis, but suffice it to say that it does not allow the growth of vorticity from an irrotational primordial state, as long as shell crossing does not occur. This is in agreement with the Eulerian description.

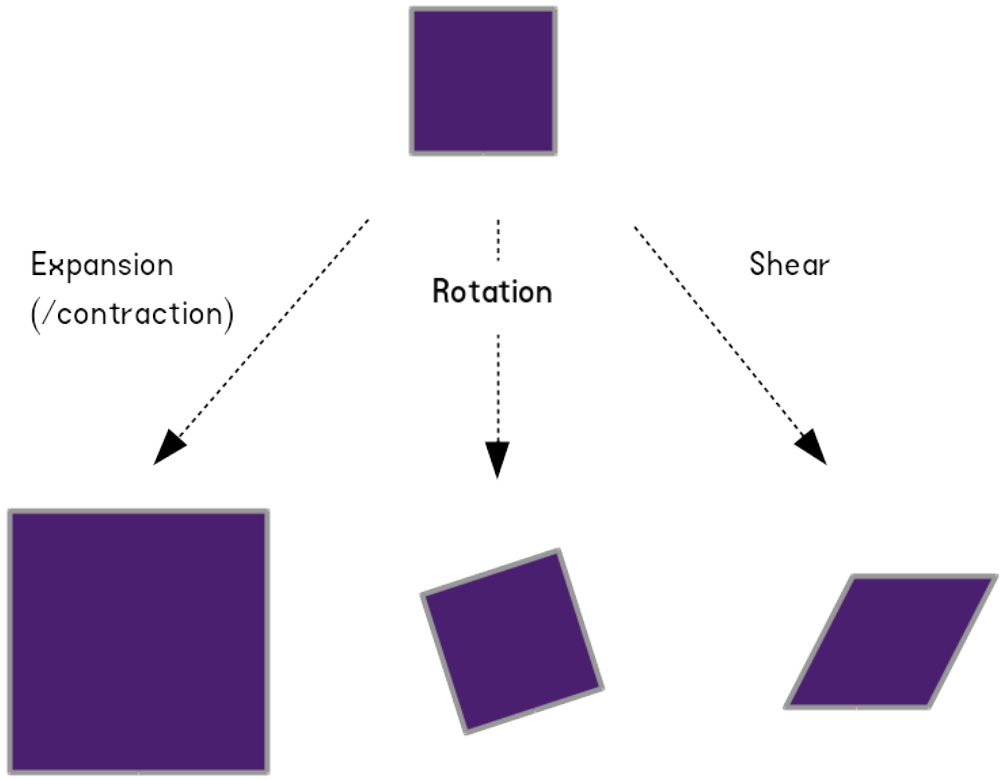


Figure 3: Three modes of deformation of fluid elements represented in two dimensions. A square fluid element can undergo changes in size, rotation and shearing — the latter effect can be seen as a shortening and expansion along two perpendicular axes, not necessarily aligned with the fluid element's principal axes. Any linear combination of these effects can apply to a fluid element at a time. Note that the three-dimensional case allows extra degrees of freedom to rotation axes, and shear planes.

to the following definitions:

$$\theta = \frac{1}{a}\nabla \cdot \mathbf{v} = \frac{1}{a}(\partial_x v_x + \partial_y v_y + \partial_z v_z) \tag{42}$$

$$\sigma_{ij} = \frac{1}{a}\frac{1}{2}(\partial_j v_i + \partial_i v_j) - \frac{1}{a}\frac{1}{3}(\nabla \cdot \mathbf{v})\delta_{ij} \tag{43}$$

$$\omega_{ij} = \frac{1}{a}\frac{1}{3}(\partial_j v_i - \partial_i v_j) \tag{44}$$

In each of these, the division by the scale factor a is to obtain the comoving quantity. From a Lagrangian perspective, the conservation of mass is ensured by considering a fluid element of fixed mass. Its density is then determined purely by the expansion, so that the continuity equation becomes:

$$\frac{d\delta}{dt} + a(1+\delta)\theta = 0 \tag{45}$$

The Euler equation has already been rewritten in terms of the Lagrangian derivative, refer to equation 19. The dependence of the potential on the matter distribution is the same in Eulerian and Lagrangian treatments, so the Poisson equation 17 is unaltered.

The Euler and Poisson equations combine into the *Raychaudhuri* equation (Raychaudhuri, 1955), which describes the evolution of divergence:

$$\frac{d\theta}{dt} + 2\frac{\dot{a}}{a}\theta + \frac{1}{3}\theta^2 + \sigma^{ij}\sigma_{ij} - 2\omega^{ij}\omega_{ij} = -4\pi G\rho_u\delta, \tag{46}$$

it shows — in combination with the continuity equation — that shear accelerates the collapse of a fluid element, while vorticity counteracts it. Appearing only in quadratic terms, these two constitute purely nonlinear effects.

Lastly, the evolution of shear is dependent on the *gravitational tidal field*[17] T:

$$T_{ij}(\mathbf{x}) = \frac{1}{a^2}\left(\partial_i\partial_j\phi - \frac{1}{3}\nabla^2\phi\delta_i j\right). \tag{47}$$

The continuity and Euler equations combine into an equation linking the shear mode of deformation of fluid elements to the gravitational tidal field:

$$\frac{d\sigma_{ij}}{dt} + 2\frac{\dot{a}}{a}\sigma_{ij} + \frac{2}{3}\theta\sigma_{ij} + \sigma_{ik}\sigma^k_j - \frac{1}{3}\delta_{ij}(\sigma^{kl}\sigma_{kl}) = -T_{ij}. \tag{48}$$

This is an important equation for the purposes of our study: it shows how the tidal field induces shear in velocity flows. As can be seen from these Lagrangian fluid equations, all quantities used are locally determined, except for the gravitational field ϕ, which depends on the entire mass distribution in all of space.

The Zel'dovich approximation

In order to construct a framework for describing the nonlinear growth of structure purely from local quantities, the *Zel'dovich approximation* (Zel'dovich, 1970) takes the complete evolution of density, velocity gradient and gravity gradient of any mass element to be fully determined by the initial conditions,

[17] The tidal tensor field T_{ij} is the traceless part of the *deformation tensor* — equation 56. Note that the tidal field and the shear field are equivalent in — and only in — the linear regime.

and independent of other mass elements. Despite the inane appearance of this assumption, the Zel'dovich approximation has proven to be very apt, and an instrumental contribution to the history of this study.

For a particle following a trajectory $\mathbf{x}(t)$, the initial comoving position $\mathbf{q} \equiv \mathbf{x}(t = t_{init})$ is a unique indicator of that trajectory. Zel'dovich approximates to first order the distance travelled at any time relative to $\mathbf{q}$, denoted $\mathbf{x}^{(1)}$. By the grace of mass conservation, the density $\rho(\mathbf{x}, t)$ of a mass element at any time and position $d\mathbf{x}$ can be related to the initial density, which equals the average density of the Universe: $\rho(\mathbf{x}, t)d\mathbf{x} = \rho_u(t)d\mathbf{q}$. Then, the density perturbation $\delta(\mathbf{x}, t)$ can be linked to the determinant of the Jacobian matrix of the mass element's displacement and deformation relative to its initial position:

$$1 + \delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t)}{\rho_u(t)} = \left\| \frac{\partial \mathbf{x}}{\partial \mathbf{q}} \right\|^{-1}. \tag{49}$$

The central assumption made in the Zel'dovich approximation is that the Jacobian matrix $\frac{\partial \mathbf{x}}{\partial \mathbf{q}}$ is determined to first order:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{q}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{\partial \mathbf{x}^{(1)}}{\partial \mathbf{q}}, \tag{50}$$

discarding higher order terms $\frac{\partial \mathbf{x}^{(n)}}{\partial \mathbf{q}}$. Then, its determinant is approximated in terms of the divergence in Lagrangian initial space $\mathbf{q}$, to $1 + \nabla_{\mathbf{q}} \cdot \mathbf{x}^{(1)}$. So that

$$\delta^{(1)}(\mathbf{x}, t) = -\nabla_{\mathbf{q}} \cdot \mathbf{x}^{(1)}. \tag{51}$$

Entering this result into the Poisson equation yields our much desired localised expression for the potential perturbation $\phi^{(1)}$. Subsequently, the assumption that the displacement $\mathbf{x}^{(1)}$ is completely potential — i.e. $\nabla \times \mathbf{x}^{(1)} = 0$ — , the Poisson equation can be manipulated to give the potential gradient:

$$\nabla \phi^{(1)} = -4\pi G \rho_u a^2 \mathbf{x}^{(1)}. \tag{52}$$

Via the Euler equation, this implies that the trajectory $\mathbf{x}^{(1)}$ is governed by:

$$\frac{d^2 \mathbf{x}^{(1)}}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\mathbf{x}^{(1)}}{dt} = 4\pi G \rho_u \mathbf{x}^{(1)}, \tag{53}$$

This is equivalent to the partial differential equation 23, for the linear growth of density perturbations, and has the same solutions. As we have in section 2.3, we discard the decaying mode solution, and we separate the solely t-dependent $D_\delta(t)$ from a solely position-dependent quantity $\boldsymbol{\psi}(\mathbf{q})$. This is a purely potential vector field — like $\mathbf{x}^{(1)}$ under our current assumptions — and can thus be written as the gradient of a *displacement potential field* Ψ:

$$\mathbf{x} = \mathbf{q} - D_\delta(t) \nabla \Psi(\mathbf{q}). \tag{54}$$

We can now relate this potential field gradient to the peculiar velocity:

$$\mathbf{v} \equiv a\frac{d\mathbf{x}}{dt} = -aD_\delta H f(\Omega) \nabla \Psi,$$

while we already have an invaluable relation linking peculiar velocity to gravity (equation 35). Combining these yields that the displacement potential field is

$$\Psi(\mathbf{q}) = \frac{2}{3D_\delta a^2 H^2 \Omega}\phi(\mathbf{x}, t), \tag{55}$$

still a time-independent quantity.

There is a way to retrieve the sec displacement field $\boldsymbol{\psi}(\mathbf{q})$ from the spatial distribution of the density perturbations. This utilises the relation between the Fourier transforms of the density perturbation and the gravitational potential field at the present epoch,

$$\hat{\phi}_0(\mathbf{k}) = -\frac{3}{2}\Omega_0 H_0^2 \frac{1}{k^2}\hat{\delta}_0(\mathbf{k}).$$

Combining this with our relation between the gravitational potential and the displacement field (from equation 55) results in $\boldsymbol{\psi}(\mathbf{k}) = -i\mathbf{e}_k\hat{\delta}(\mathbf{k})$. This allows the displacement vector field $\boldsymbol{\psi}(\mathbf{q})$ to be calculated from the initial density perturbations, determining approximately the formation of structure throughout the linear regime, and even resulting in a reasonable approximation beyond.

Zel'dovich deformation tensor and anisotropic collapse

To this end, we introduce the *deformation tensor*[18]:

$$\psi_{ij} = \frac{D_\delta(t)}{a(t)}\frac{\partial^2\Psi}{\partial q_i \partial q_j} \tag{56}$$

It is always possible to define a coordinate system in which this tensor can be written in a purely diagonal form. The axes then point along the principal directions of the deformation, and the tensor is described completely by its eigenvalues λ_1, λ_2 and λ_3. The linearly extrapolated density perturbation $\underline{\delta}(t)$ can then be written as:

$$\underline{\delta}(t) = a(t)\sum_m \lambda_m. \tag{57}$$

As a result of this approximation, it is possible to take the tidal field to be linearly proportional to the shear, (Hui & Bertschinger, 1996), with proportionality $-\frac{3\Omega H^2}{2Hf(\Omega)}$. This proportionality violates equations 46 and 48, but allows for a localised approach to simulating structure formation.

Only at the point where different streams of matter from different starting points $\mathbf{q}$ meet each other does the Zel'dovich formalism really break down. The approximation of immutable displacements sends mass elements onward on their same old trajectory after this crossing takes place, which is incongruous with the gravitational attraction between the streams that is induced in reality. The main limitation of the Zel'dovich formalism is that it does not reproduce the self-gravitation of structures.

A note of credit: the hierarchical formation of filaments and nodes is immediately predicted from the Zel'dovich formalism, which was formulated in 1970. Twenty-six years before the term 'cosmic web' was coined — see the introduction.

[18]Notice, through equation 55, that ψ is proportional to the Hessian of ϕ.

As Nusser et al. (1991) have noted, building from the Zel'dovich formalism, the extraction of a density from the velocity field can be done in two distinct ways. One option, solving the continuity equation, results in a *continuity density* ρ_c. Resting on the concept of Zel'dovich displacements, the continuity equation can be written in terms of the Eulerian density δ_x and the Lagrangian density δ_q:

$$\rho_x(\mathbf{x})d^3x = \rho_q d^3q \tag{58}$$

Nusser et al. derive the continuity density as follows:

$$\delta_c(\mathbf{x}) = \left\| \frac{\partial \mathbf{q}}{\partial \mathbf{x}} \right\| - 1 \tag{59}$$

$$= \left\| I - D_\delta \frac{\partial \boldsymbol{\psi}}{\partial \mathbf{x}} \right\| - 1 \tag{60}$$

$$= \left\| I - (Hf)^{-1} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \right\| - 1, \tag{61}$$

where I is the identity matrix. As long as shell crossing does not occur, and the velocity field is purely potential, the deformation tensor $\frac{\partial \boldsymbol{\psi}}{\partial \mathbf{x}}$ can be diagonalised at any given x. The continuity density is then a function of the eigenvalues $\lambda_1, \lambda_2, \lambda_3$:

$$\delta_c = (1 - D_\delta \lambda_1)(1 - D_\delta \lambda_2)(1 - D_\delta \lambda_3) - 1. \tag{62}$$

The continuity density ensures that mass is conserved, but in general fails to conserve momentum.

Alternatively, solving the equation of motion will conserve momentum but generally violate the conservation of mass. Nusser et al. derive the *dynamic density* ρ_d by inserting the Zel'dovich motions into the dynamical Euler-Poisson equation. The resulting density

$$\delta_d(\mathbf{x}) = -(Hf)^{-1} \nabla \cdot \mathbf{v} = -D_\delta M_1. \tag{63}$$

Like the continuity density, the derivation of the dynamic density rests on the assumption of no shell crossing, but holds for rotational flows as well. The relation between velocity and density in the quasi-linear regime is now a probe of the differences between Zel'dovich displacements and the displacements brought about by real gravity.

2.6 Gaussian random fields

For cosmological purposes, Gaussian Random Fields (GRFs) are the core concept in the statistical analysis of matter distributions. The primordial density distribution — that forms the initial conditions for all cosmic structure formation — is assumed to be Gaussian random. It is the product of the *inflation* of quantum fluctuations.

The assumption that primordial density fields — and those derived from density — are Gaussian rests on a few pillars. Among them are the fact that inflation is assumed to generate Gaussian fluctuations; and the *central limit theorem*, which states that the superposition of a large number of variables independently drawn from the same distribution is Gaussian. Also, convincing deviations from Gaussianity have yet to be found (e.g. Planck Collaboration et al., 2015b).

A field that follows a homogeneous and isotropic Gaussian Random distribution can be statistically specified completely by one single descriptor: the *power spectrum*. It is within the framework of GRFs that we can specify the most important properties of structure formation — e.g. the relative sizes of collapsed structures; the manner of hierarchical structure formation; and the (non)reciprocal influence exerted between large and small scale structures. Particularly the properties of the initial density — or gravitational potential — distribution are of key interest, as they determine the most important aspects of all structure formation that follows.

The analysis of physical quantities over a broad range of scales via a more straightforward computational route — e.g. simulations — can become problematic. For example, the highest density peaks will always be rare, the semi-analytical approach of GRFs is more suitable for such purposes. Seminal cosmological research on the topic of GRFs has been done by Adler (1981); Vanmarcke (1983); Peacock & Heavens (1985) and particular credit goes to Bardeen et al. (1986).

Gaussian random fields are fields of variables from a Gaussian distribution. The initial conditions of structure growth — id est: the density fluctuations resulting from inflation, observable through the Cosmic Microwave Background radiation — are assumed to be Gaussian. This assumption has so far withstood rigorous testing (see e.g. Fergusson & Shellard, 2009; pla, ????; Planck Collaboration et al., 2015b).

In studies where measurements can be taken from various realisations of an underlying physical system, a simple *ensemble average* over a suitable number of realisation can straightforwardly provide a measurement and a corresponding confidence level. However, astronomical observations are limited to only one Universe. Fortunately, the *ergodic theorem* states that a *spatial average* over a number of sufficiently small patches — sub-horizon scale — is an adequate substitute to an ensemble average[19].

The main statistical descriptor of a GRF is the *power spectrum*. In physical terms, it is a measure of the 'clumpiness' of a realisation of the field. More generally, a random (non)Gaussian field can be specified completely by the *covariance matrix* and the mean of the variables. After introducing these, the GRF power spectrum will be defined. Following that, we will apply this theory to cosmological structure formation, and present a more generally applicable alternative: the *lognormal* distribution.

The covariance matrix and the correlation function

For a general M-dimensional random field $f : R^M \to C$, the N*-point probability function* gives the probability that N specific points in space $\{\mathbf{x_1}...\mathbf{x}_N\}$ satisfy N specific values $f(\mathbf{x}_i) = f_i \quad \forall i \in [1;N]$ — i.e. $f(\mathbf{x_1}) = f_1$ and $f(\mathbf{x_2}) = f_2$ and ... and $f(\mathbf{x}_N) = f_N$. This N-point probability function looks like:

$$P_N = P[f(\mathbf{x_1}), f(\mathbf{x_2}), \ldots, f(\mathbf{x}_N)]df_1 df_2 \ldots df_N \tag{64}$$

[19] Adler (1981) has shown that this requires the power spectrum — introduced momentarily — to be continuous in k.

The *covariance* function $\xi(\mathbf{x}_1, \mathbf{x}_2)$ is a measure of the statistical dependence between two points $\mathbf{x}_1$ and $\mathbf{x}_2$, it is defined as the following *ensemble average*:

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \langle f(\mathbf{x}_1) f(\mathbf{x}_2) \rangle. \tag{65}$$

If the field is homogeneous and isotropic, the only thing that really matters is the distance between the two points, and the correlation function $\xi(|\mathbf{x}_1 - \mathbf{x}_2|)$ gives a general feel of how smooth the field is. A large correlation between points far apart implies a smooth landscape, low correlation between nearby points characterises a more Himalayan scene. It can easily be shown that the correlation of a point with itself equals the variance of the field's underlying distribution: $\xi(\mathbf{0}) = \sigma^2$.

Gaussian Random Field

For a *Gaussian field* with standard deviation σ, the 1-point function is

$$P_1(y) = \frac{1}{\sqrt{2\pi}} \sigma^2 e^{-\frac{y^2}{2\sigma^2}}. \tag{66}$$

The N-point function is:

$$P_N \prod_i^N dy_i = \frac{1}{(2\pi)^{N/2} \sqrt{\|\mathbf{M}\|}} \exp\left\{ -\frac{1}{2} \sum_i^N \sum_j^N y_i (M_{ij}^{-1}) y_j \right\} \prod_i^N dy_i \tag{67}$$

where $\mathbf{M}$ is the *covariance matrix* — and $\|\mathbf{M}\|$ its determinant:

$$\mathbf{M} = \begin{pmatrix} \xi(\mathbf{0}) & \xi(\mathbf{x}_1, \mathbf{x}_2) & \cdots \\ \xi(\mathbf{x}_2, \mathbf{x}_1) & \xi(\mathbf{0}) & \\ \vdots & & \ddots \end{pmatrix}. \tag{68}$$

Power spectrum

Equation 65 can be worked out to:

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \int \frac{d\mathbf{k}}{(2\pi)^3} P(\mathbf{k}) e^{-i\mathbf{k}\cdot|\mathbf{x}_1 - \mathbf{x}_2|}, \tag{69}$$

where $P(\mathbf{k})$ is the *power spectrum*, defined as

$$(2\pi)^3 P(\mathbf{k}_1) \delta_D(\mathbf{k}_1 - \mathbf{k}_2) = \langle \hat{f}(\mathbf{k}_1) \hat{f}^*(\mathbf{k}_2) \rangle = \langle |\hat{f}(\mathbf{k}_1)| |\hat{f}(\mathbf{k}_2)| \rangle, \tag{70}$$

where $\delta_D(\mathbf{k})$ is the Dirac delta function, and the superscript * denotes complex conjugation. The power spectrum is the Fourier transform of the correlation function.

The power spectrum of the initial density field of the Universe is extremely influential for the formation of structure — see e.g. Little et al. (1991). It measures the contribution of different spatial wavenumbers $\mathbf{k}$ to the fluctuations in the field. If there is a lot of contribution from high wavenumbers, small scale fluctuations will play an important role. They will collapse early, allowing for a hierarchical *bottom-up* structure formation scenario. See the left column in figure 4 for an illustration. However if there is a cutoff at some $\mathbf{k}$, above which

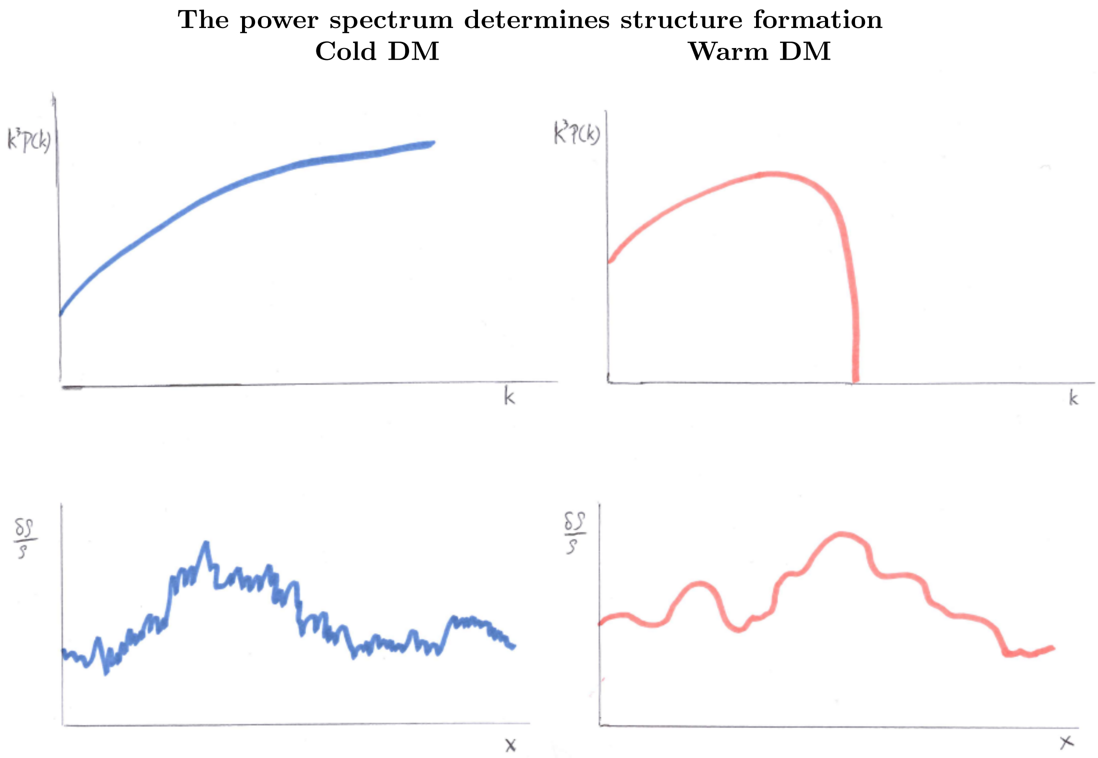


Figure 4: Two different power spectra (top row) and possible corresponding one-dimensional density distributions (bottom row). The panels on the left correspond to a ‘cold dark matter’ cosmology, in which high frequency components dominate the spectrum, allowing for small scale activity and hierarchical structure formation. The panels on the right show a ‘warm dark matter’ type of spectrum, where structure below a cutoff scale does not occur.

the power spectrum is zero, the longer wavelength components will dominate. This implies a *top-down* scenario for the formation of structure — illustrated in the right column of figure 4. A power spectrum where high frequencies dominate — e.g. a power law with a spectral index $n > -3$ — is associated with a *cold dark matter* cosmology (Peebles, 1982; Bond & Szalay, 1983). A power spectrum dominated by low frequencies corresponds to *warm dark matter*.

Note that the Fourier transform $\hat{f}(\mathbf{k})$ contains all the information of a GRF realisation, and is in general a complex number. Written in terms of its phase and amplitude, $\hat{f}(\mathbf{k}) = |\hat{f}(\mathbf{k})|e^{i\theta(\mathbf{k})}$. Equation 70 shows that the power spectrum of a field f is independent of its phases $\theta(\mathbf{k})$. While the power spectrum contains crucial statistical information of a GRF, it is no descriptor of its field value at any specific location.

GRFs in structure formation

From linear perturbation theory (section 2.3) it can be shown that the assumption of initial density perturbations δ being a Gaussian field leads to the velocity $\mathbf{v}$ and gravitational potential ϕ being Gaussian as well, following

$$P_{\mathbf{v}}(\mathbf{k}) = (Haf(\Omega))^2 \frac{P(\mathbf{k})}{\mathbf{k}^2}, \tag{71}$$

and

$$P_{\phi}(\mathbf{k}) = (\frac{3}{2}\Omega H^2 a^2)^2 \frac{P(\mathbf{k})}{\mathbf{k}^4}. \tag{72}$$

This implies that the density field is dominated by small-scale fluctuations, velocity by medium scale fluctuations, and gravitational potential by the larger scale distribution of matter.

The primordial power spectrum — containing the seeds of structure growth — emanates from the *inflationary* phase of cosmic history. This is predicted to be a *Harrison-Zel'dovich spectrum*, a spectrum increasing linearly by wavenumber, given by:

$$P(k) = Ak. \tag{73}$$

A more general case is a power law spectrum

$$P(k) = Ak^n,$$

characterised by a slope A and a spectral index n. In combination with equation 72, this means that the variance of the gravitational potential σ_ϕ^2 obeys:

$$\sigma_\phi^2 \propto A \int \frac{d \log k}{(2\pi)^2} k^{n-1}. \tag{74}$$

Therefore, a Harrison-Zel'dovich spectrum, with $n = 1$, results in a scale-free contribution to potential perturbations.

Departing from the primordial state, the formation of structure involves an evolution of the power spectrum. Physical processes generally have different influences on different frequency ranges. During the linear regime, the distributions retain a Gaussian shape, and the various Fourier components evolve independently from each other. Still, processes like *Silk damping* may result in a low-pass filtering.

The lognormal distribution

While Gaussian random fields appear an adequate model for distributions in the linear phase of structure formation, deviations form Gaussianity appear in the nonlinear regime. An important example is the unbounded increase of density peaks under gravitational instability, and the spatial growth of underdense regions. The density distribution then evolves away from a Gaussian curve.

To model this non-Gaussian distribution, Coles & Jones (1991) have suggested a framework centred upon the *lognormal* distribution — illustrated in figure 46 — to approximate the distribution of density perturbations evolving away from initial conditions. The lognormal distribution is different from the Gaussian, but can be arbitrarily close to it given certain combinations of its parameters. For fields following a lognormal distribution, the same holds as for GRFs: they can be statistically characterised completely by one covariance function. In contrast to the Gaussian distribution, though, the lognormal is not completely specified by its moments.

The lognormal distribution $Y(\mathbf{r})$ is obtained by transforming a Gaussian $X(\mathbf{r})$ as follows:

$$Y(\mathbf{r}) = \exp[X(\mathbf{r})] \tag{75}$$

An underlying Gaussian distribution with mean μ and variance σ^2 then yields the following lognormal one-point distribution function:

$$f_1(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log y - \mu)^2}{2\sigma^2}\right\} \frac{dy}{y}. \tag{76}$$

Nota bene that μ and σ^2 do not correspond to the mean and variance of the lognormal distribution. It can easily be shown that the n-th moment about the origin, μ'_n, follows

$$\mu'_n = \exp[n\mu + n^2\sigma^2/2],$$

so that the mean μ'_1 is

$$\mu'_1 = e^{\mu+\sigma^2/2}. \tag{77}$$

The variance — the 2$^{\text{nd}}$ moment about the mean, μ_2 — is then given by

$$\mu_2 = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

For the purposes of this study, it is useful to determine the median, which is given by

$$\text{median}(\mu) = e^{\mu}. \tag{78}$$

Having defined the one-point distribution function, a *lognormal random field* is determined by the covariance matrix $\mathbf{M}$:

$$f_n(y_1, ..., y_n) = \frac{1}{\sqrt{(2\pi)^n \|\mathbf{M}\|}} \exp\left\{-\frac{1}{2}\sum_{i,j} \mathbf{M}_{ij}^{-1} \log(y_i)\log(y_j)\right\} \prod_{i=1}^{n} \frac{1}{y_i}. \tag{79}$$

This field, too, can be specified by a correlation function $\xi(\mathbf{r})$. The reader is referred to (Coles & Jones, 1991) for a detailed description.

Notice, from the one-point distribution function 76, that the values in a lognormal distribution are confined to the support $y \in \langle 0; \infty\rangle$, which works in its favour as a potential descriptor[20] for the density perturbation $\delta+1$. Notice, also, that by taking a sufficiently large mean relative to the variance, the lognormal distribution can approach arbitrarily closely to a Gaussian. This property is of obvious importance, since the initial conditions are assumed to be Gaussian.

The fact that observations of galaxies on the celestial plane can be well approximated by the lognormal distribution function (Hubble, 1934; Peebles, 1980) is another motivation for the lognormal model. Also, the continuity equation implies that the density grows as an exponential function of velocity divergence.

2.7 Velocity flows

The velocity field in a cosmological volume is a three dimensional vector field, meaning that for each of three components, a spatial derivative with respect to any of three spatial directions can be defined. The velocity gradient tensor is the *Jacobian* of the velocity field:

$$\nabla \mathbf{v} = \begin{pmatrix} \frac{\partial v_x}{\partial x} & \frac{\partial v_x}{\partial y} & \frac{\partial v_x}{\partial z} \\ \frac{\partial v_y}{\partial x} & \frac{\partial v_y}{\partial y} & \frac{\partial v_y}{\partial z} \\ \frac{\partial v_z}{\partial x} & \frac{\partial v_z}{\partial y} & \frac{\partial v_z}{\partial z} \end{pmatrix}. \tag{80}$$

This tensor contains all the information on the spatial change in velocity, but there are three convenient quantities that describe different modes of the velocity gradient: *divergence* θ, *shear* σ and *vorticity* ω. The velocity gradient is retrieved by the following summation:

$$\partial_j v_i = \frac{1}{3}\theta\delta_{ij} + \sigma_{ij} + \omega_{ij}, \tag{81}$$

where δ_{ij} is the Kronecker delta.

The illustrations in figures 5 and 7 are from the wikimedia commons, created by Jorge Stolfi, and licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

Divergence

Divergence is defined as the sum of the diagonal components of $\nabla \mathbf{v}$.

$$\nabla \cdot \mathbf{v} \equiv \partial_x v_x + \partial_y v_y + \partial_z v_z. \tag{82}$$

In comoving space, there is the divergence θ:

$$\theta = \frac{1}{a}\nabla \cdot \mathbf{v} \tag{83}$$

It is a scalar field, that measures the amount of outward pointing velocity from a given point — a negative divergence indicates the velocity field converging towards a given point. See panel (b) in figure 5 for an illustration of positive divergence. It is this mode, together with the shear mode, that characterises a potential flow — see section 2.3. In cosmological scenarios, the prominent cause for negative divergence is a peak in density, e.g. a halo or a galaxy cluster, whose

[20] The lognormal model is a particularly simple descriptor for positive definite distributions, and offers an analytical evaluation of various properties.

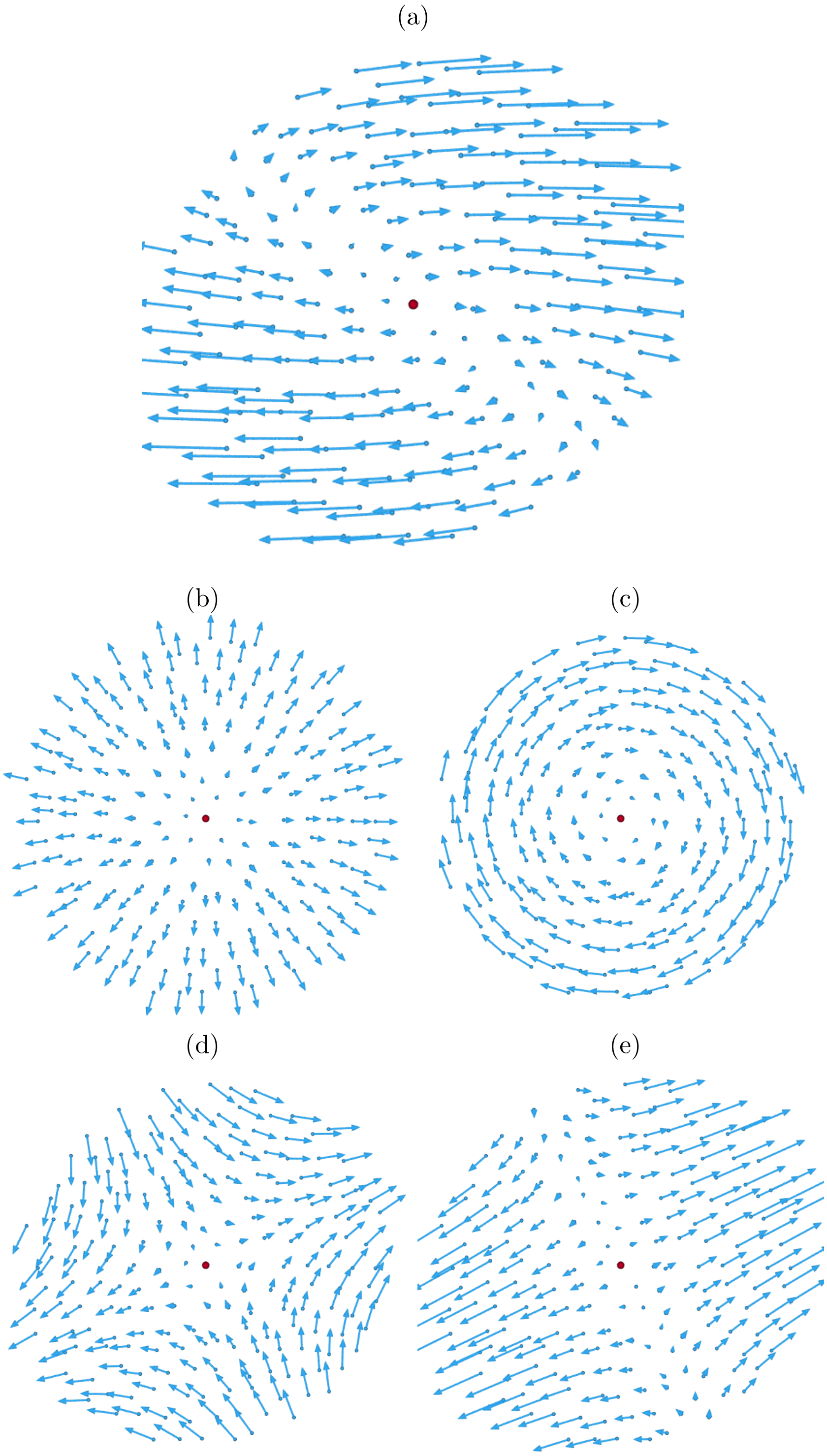


Figure 5: Decomposition of a two-dimensional sample velocity field (panel a) into its divergence (panel b), vorticity (panel c), and traceless shear (panel d). Panel (e) shows the symmetric component of the velocity gradient, which is the shear plus the divergence. Note that the generalisation into three dimensionality is simple for the case of divergence: it is measured by the total inward and outward flow in all spatial directions. Three-dimensional vorticity is — like the two-dimensional case shown — characterised by one single rotational axis. Two- and three-dimensional shearing motion can take on a large variety of appearances. See text and figures 6 and 7 for details. Image: Jorge Stolfi

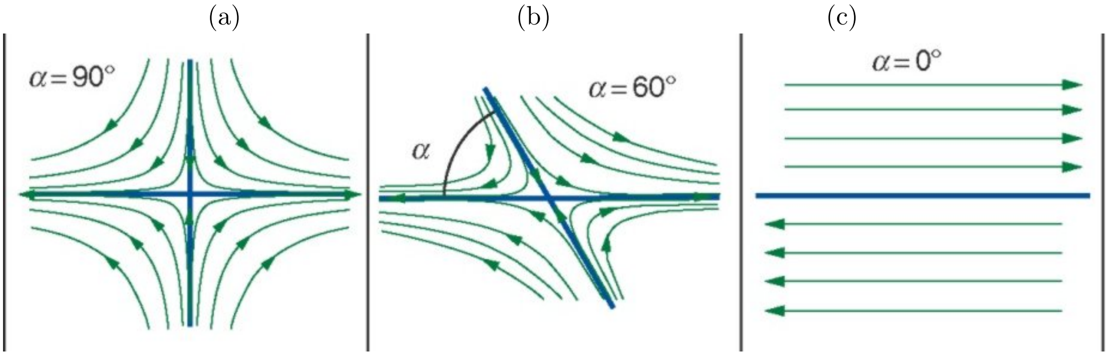


Figure 6: Illustrations of pure shear (a), general shear (b) and simple shear (c). Source: (Fossen, 2016).

depressed gravitational potential manifests a region of gravitational acceleration pointing towards itself.

A positive divergence is usually related to the presence of an underdensity. The continuity equation relates the expansion or contraction of a velocity field to the density gradient:

$$\frac{1}{\rho}\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{v}. \tag{84}$$

This is a fundamental hydrodynamical relation, that holds in linear and non-linear conditions. The typically large regions of space between clusters, filaments and walls initially comprise only a slight underdensity. As the positive density perturbations attract matter and collapse, the underdense regions are drained and expanded. It is possible for the outflow of matter from these regions to create a divergence greater than the *Hubble expansion*, but the greatest possible divergence is determined by the greatest possible underdensity[21] $\delta = -1$ (Romano-Díaz & van de Weygaert, 2007)

$$\theta_{\max} = 1.5\Omega^{0.6}.$$

See the introduction for a brief description of voids.

The physical divergence Θ of a velocity field will be the comoving divergence θ plus the global divergence due to the hubble expansion $\Theta_H = 3H(t)$. The hubble divergence is simply to be subtracted from Θ, to obtain the comoving quantity.

Within the framework of linear perturbation theory, the divergence of the velocity field is directly related to the density field — see equation 38.

Shear

The shear σ of a three-dimensional velocity field is a 3 by 3 tensor. The ij component is defined as

$$\sigma_{ij} = \frac{1}{2}(\partial_j v_i + \partial_i v_j) - \frac{1}{3}(\nabla \cdot \mathbf{v})\delta_{ij}. \tag{85}$$

[21] Romano-Díaz & van de Weygaert show that 1.5 is the difference between H in Universes characterised by $\Omega = 0$ and $\Omega = 1$. Note that void interiors characterised by a flat and minimal density profile "locally mimic the behaviour of an $\Omega = 0$ universe".

In comoving space, this quantity is divided by the scale factor a. From the above expression, it is readily visible that it is always a symmetric tensor — $\sigma_{ij} = \sigma_{ji}$. The shear defined in equation 85 is *traceless*, i.e.:

$$\sigma_{xx} + \sigma_{yy} + \sigma_{zz} = 0.$$

As a symmetric and traceless tensor, σ can be specified completely by five independent components. Physically, shear measures the change of velocity like in panel (d) of figure 5. Like divergence, shear can exist in a purely potential velocity field.

The scalar quantity of *shear magnitude* can be defined as

$$|\sigma| = \sqrt{\sigma_{xx}^2 + \sigma_{yy}^2 + \sigma_{zz}^2 + 2\sigma_{xy}^2 + 2\sigma_{yz}^2 + 2\sigma_{xz}^2}. \tag{86}$$

This quantity traces the total amount of shear in any direction.

Note that the shear flow illustrated in figure 5 — panel (d) — is a special symmetric case, referred to as *pure shear*. More generally, two-dimensional velocity flows can also exhibit *simple shear* — see panel (c) in figure 6. A combination of the two results in *general shear* — panel (b) in figure 6. The extension to three dimensions allows for various other shear patterns, all of them can be specified by five independent components of the shear tensor.

Panel (e) in figure 5 shows the complete *symmetric part of the velocity gradient*:

$$\frac{1}{2}(\partial_j v_i + \partial_i v_j).$$

This quantity is also referred to as *strain*, and it incorporates both shear and divergence. As such it specifies completely the potential component of any velocity flow. The strain tensor is symmetric but not traceless, so it is determined by six independent components.

In the context of cosmological structure formation, the shear mode in velocity flows is of great relevance in various ways. Firstly, shear is induced by anisotropic gravitational potentials; and a manifestation of anisotropic velocity flows in structure formation. Particularly, it is a major aspect of the formation of filamentary and sheet-like structures. As such, the shear mode is a potent agent and probe of structure formation. See figures 1 and 2 for visualisations of this principle through N-body simulations. Equation 48 governs the evolution of velocity shear, which depends on the tidal field T. It reflects that shear is induced by anisotropy in the gravitational potential. See Bond et al. (1996) for a more detailed discussion.

As an inextricable consequence of this, the shear mode of velocity flows can be used to identify the components of the cosmic web — i.e. nodes, filaments, walls and voids. Section 2.8 elaborates upon this. This study, too, relies on the shear tensor for the classification of cosmic web components, see section 5.4 for an explanation of our methods.

Furthermore, in Lagrangian theory — section 2.5 — it is derived that shear has a nonlinear effect on the gravitational collapse of structures. The Raychaudhuri equation 46 shows that shear accelerates the collapse of fluid elements.

Vorticity

As with the shear, the vorticity ω of a three-dimensional velocity field is a 3 by

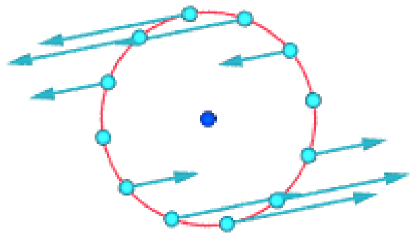


Figure 7: Velocity vorticity resulting from a simple shearing motion like in panel (c) of figure 6. Image: Jorge Stolfi

3 tensor. It follows the definition:

$$\omega_{ij} = \frac{1}{2}(\partial_j v_i - \partial_i v_j). \tag{87}$$

Vorticity, too, is divided by a to obtain the comoving tensor. This expression clearly results in the vanishing of the diagonal elements, and shows that vorticity measures the asymmetry in the gradient tensor, the difference between the cyclic and contra-cyclic components. Vorticity is also called the *antisymmetric* component of the velocity gradient. When $\partial_i v_j$ and $\partial_j v_i$ do not balance out, a rotation about the k-axis is created. Panel (c) in figure 5 gives a two-dimensional visualisation of this. In three dimensions, the rotation vector and velocity gradient are related by the Levi-Civita symbol of their indices.

$$\omega^k = \epsilon^{kji} \partial_j v_i.$$

Note that vorticity is not only generated by 'vortex-like' velocity patterns such as the one in figure 5. In fact, vorticity can be a part of a field characterised by simple shear — panel (c) in figure 6. From the perspective of a point in the middle of that diagram, there is a disbalance between the upper and lower flows; velocities at neighbouring points may follow a pattern illustrated in figure 7.

From Lagrangian theory, we learn that vorticity has an inhibiting effect on the collapse of structures. Notice that vorticity ω appears with the opposite sign from shear σ in the Raychaudhuri equation 46. As with the shear, this is only a nonlinear effect.

A potential flow has zero vorticity. As described in section 2.3, equation 30, the rotational component of the cosmological velocity field in the *linear regime* decays rapidly. Any vorticity in the Universe has to be generated at a later epoch. It is only when different velocity streams start to overlap spatially that vorticity can be generated — this scenario is referred to as *shell crossing* and *multi-streaming*, see section 2.9 for a discussion.

Vorticity plays only a side role in this study. While it is certainly of interest in nonlinear structure formation, there are nontrivial hurdles to be crossed in the analysis of vorticity measurements. Section 2.9 will explain how vorticity measurements can be contaminated by a certain *projection effect*, due to the projection of velocity streams from phase space onto physical space (Hahn et al., 2015). A proper analysis of vorticity falls outside the scope of this project.

2.8 Velocity-based web classification

When trying to work out the type of environment a matter element is embedded in at a certain scale[22], one very straightforward approach would be to consider the local density. After all, there clearly is a coarse density dependence between voids, walls, filaments, and nodes[23]. This method has been employed by Lemson & Kauffmann (1999); Macciò et al. (2006); Maulbetsch et al. (2007); Cautun et al. (2013), but it is not the most reliable option, as very large overlap exists between the densities spanned by different web components.

From a visual assessment of the cosmic web — through observations and simulations — it is undeniable that the boundaries between nodes, filaments, walls and voids are blurry. An unambiguous definition marking the boundaries of these types of structures does not exist. This poses an epistemological challenge in the development of classification algorithms. It is inevitable that such algorithms will rely on free parameters that — in the words of Hoffman et al. (2012) — "cannot be determined from first principles".

Among the parameters a web classification algorithm depends on is the spatial scale at which structures are to be identified. Aragón-Calvo et al. (2007a) has developed a way to circumvent this problem, by studying a volume at a range of scales. This technique is called the *multiscale morphology filter* (MMF), and its essence lies in defining a *scale space*. This is an extension of three-dimensional space, by a fourth dimension which specifies the length scale at which space is filtered. Cautun et al. (2013) extended this strategy in the formulation of the web classifiers NEXUS and NEXUS+. The latter applies a lognormal filter kernel to the density field, which results in a good resolution of structures spanning several orders of magnitude in spatial scales. The local geometry of a mass field is quantified by the Hessian — the second order derivative — of the density field. The Nexus algorithms operate by determining the Hessian of a tracer field[24], and performing an eigendecomposition on that Hessian tensor. A study of the eigenvalues allows for a local classification of the dimensionality of the environment, and thus offers a way to identify cosmic web components.

Investigations of the eigenvalues of the shear tensor (e.g. Doroshkevich & Shandarin, 1978a,b) are by no means a novelty, but the following literature study will demonstrate their usefulness to the purpose of web classification. Velocity-based web classification allows for a resolution high enough to study the formation of haloes and galaxies (Hoffman et al., 2012). A study of the geometry of velocity and potential fields allows a view on the dynamics of structure formation. Various studies have focused on velocity fields as tracers of the cosmic web (e.g. Hahn et al., 2007; Forero-Romero et al., 2009; Hoffman et al.,

[22] One may want to do this since certain trends have been observed between properties of (satellite) galaxies and sub-haloes and their types of environments (Dressler, 1980; Knebe et al., 2004; Blanton et al., 2005; Avila-Reese et al., 2005; Gao et al., 2005; Libeskind et al., 2005; Maulbetsch et al., 2007; Libeskind et al., 2011; Forero-Romero et al., 2011).

[23] This is very intuitive, it has been reported in many studies (Hahn et al., 2007; Cautun et al., 2013), and it is readily visible in our results, see figure 41.

[24] In the case of Nexus, this can be the density, tidal tensor, divergence or shear.

2012; Cautun et al., 2013; Tempel et al., 2014; Cautun et al., 2014; Fisher et al., 2016).[25]

T-web: Tidal tensor eigenvalues

Hahn et al. (2007) have taken an important step in the development of this method for the classification of cosmological environments. The equation of motion can be linearised (see also van de Weygaert & Bertschinger, 1996), yielding[26]:

$$\ddot{\mathbf{x}}_i = -T_{ij}(\mathbf{x}_k)(x_j - x_{k,j}), \tag{88}$$

where T_{ij} is the tidal tensor, the Hessian of the gravitational potential

$$T_{ij} \equiv \partial_i \partial_j \phi.$$

From this, Hahn et al. determine, to first order, the dynamics around $\mathbf{x}_i$, from the three eigenvalues of T_{ij}. They argue that the eigenvalue signature of the force field is related to the local morphology of the cosmic web. A positive eigenvalue indicates a contraction along the direction of its corresponding eigenvector; a negative eigenvalue indicates expansion. Therefore number of positive eigenvalues — which can be $[0 - 3]$ for the non-traceless tidal tensor — is linked to the dimensionality of the structure near $\mathbf{x}_i$.

# positive eigenvalues:	environment classification:
0	Void
1	Wall
2	Filament
3	Node

In 2009, Forero-Romero et al. improved this method with the requirement that an eigenvalue exceeds a certain positive threshold λ_{th}, before it is counted towards the classification signature. It was argued that this filters out the cases in which only a very minimal collapse occurs — i.e. where $1 >> \lambda_i > 0$. If the threshold λ_{th} is too low, it will result in an overrepresentation of collapsed regions, and an underrepresentation of voids. Conversely, a very high λ_{th} results in a volume dominated by voids, where filamentary regions no longer reach from nodes to neighbouring nodes[27]. Figure 9 — Figure 1 in (Forero-Romero et al., 2009) — gives a concise visual explanation of this. For a properly normalised T_{ij}, Forero-Romero et al. find empirically that a threshold $\lambda_{th} \in [0.2; 0.4]$ is suitable for a visual reproduction of the cosmic web structures.

Classification on the basis of the tidal field has its drawbacks. The application of this method requires that a spatial scale and threshold eigenvalue are chosen, we elaborate on this soon. Another disadvantage is an inescapable consequence

[25] Other approaches to web classification have been explored too, e.g. Novikov et al. (2006); Aragón-Calvo et al. (2007b,a); Sousbie et al. (2008); Bond et al. (2010a,b); Sousbie (2011); Shandarin et al. (2012); Cautun et al. (2013) but this section will be limited to the direct genealogy of the classification method used in this study.

[26] Hahn et al. do make the point that the Hubble drag term — the $\frac{\dot{a}}{a}\mathbf{v}$ that appears in the comoving Euler equation 16 — is discarded in this treatment, by considering the test element to be "frozen in time". $\ddot{\mathbf{x}}$ then indicates a derivative "with respect to a fictitious time".

[27] In the usual jargon, it is said that the volume is no longer *percolated* (Zel'dovich et al., 1982; Shandarin & Zeldovich, 1989; Shandarin et al., 2004, 2010).

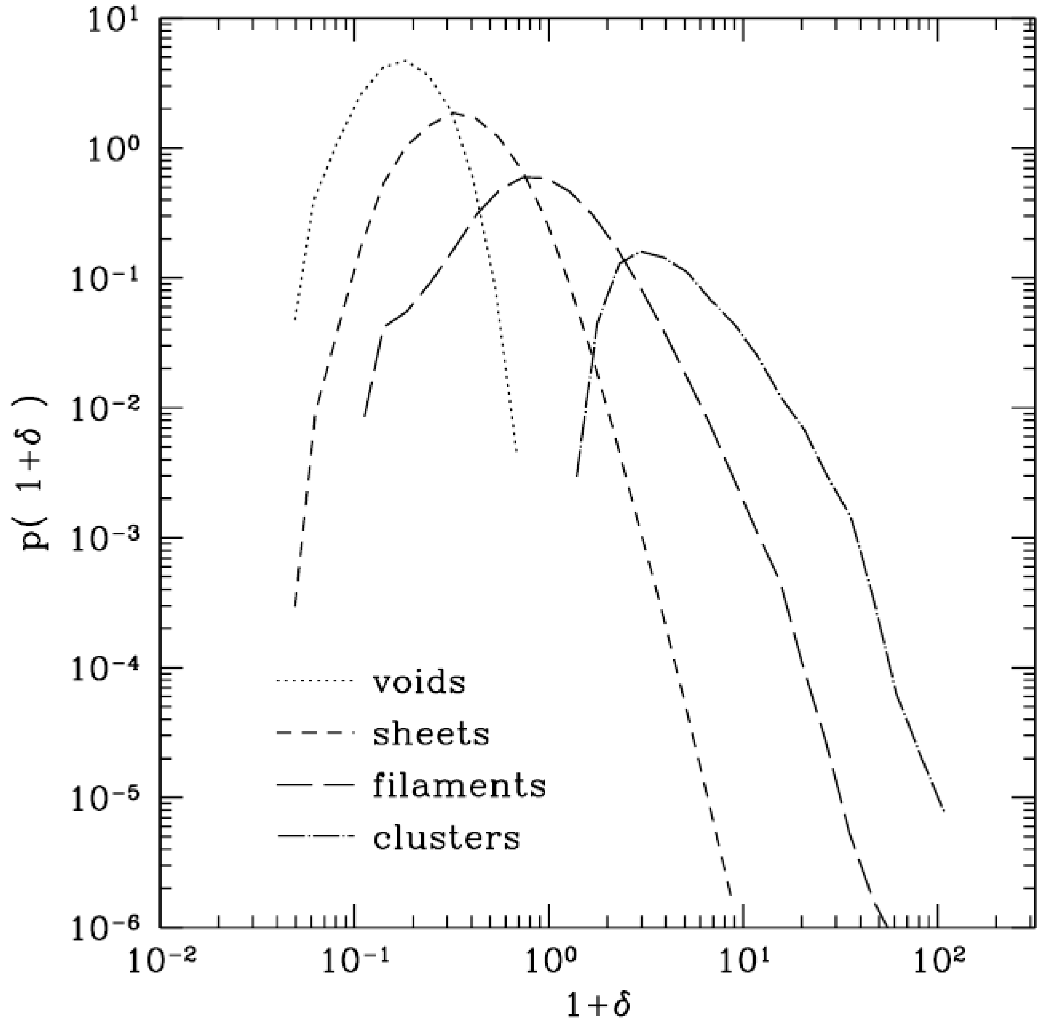


Figure 8: Statistical distributions of density taken from four different types of regions in numerical N-body simulations conducted in a study by Hahn et al. (2007). The types of regions — voids, sheets, filaments and clusters — were disentangled by the criterion of *orbit stability* — see text. Notice the large degeneracy between density and region types, resolved by this method. Image source: (Hahn et al., 2007)

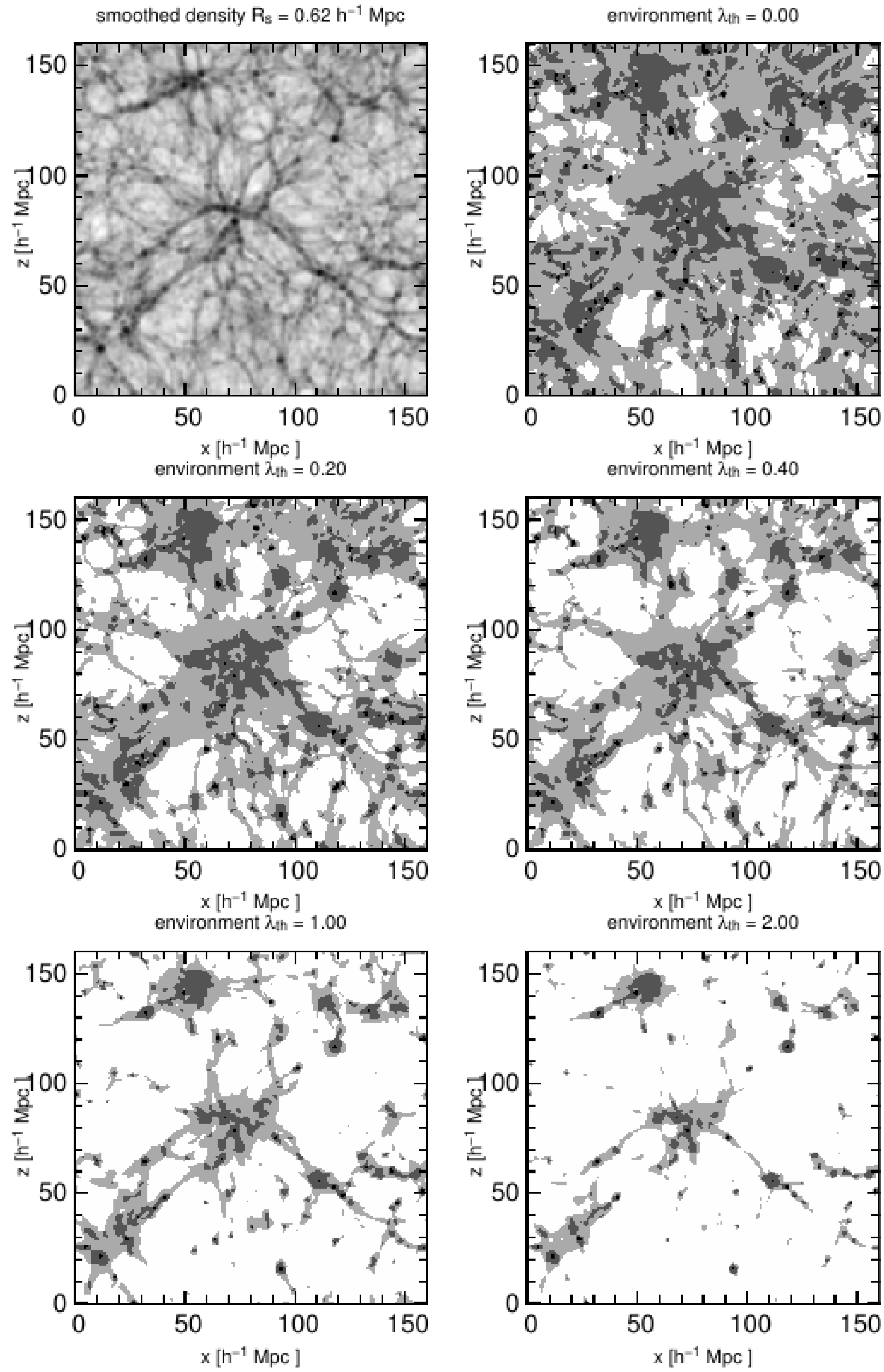


Figure 9: figure 1 from (Forero-Romero et al., 2009) shows the degree of percolation depending on the threshold eigenvalue. The top left panel shows the logarithmic density field, the other panels show the web classifications resulting from threshold eigenvalues $\lambda_{th} \in \{0.00, 0.20, 0.40, 1.00, 2.00\}$. In these plots, black regions indicate nodes, dark grey indicates filaments, pale grey indicates walls and white regions are voids. As higher threshold λ_{th} are taken, notice the increase in void volume, the decrease in nodes and filaments, and the decrease in percolation.

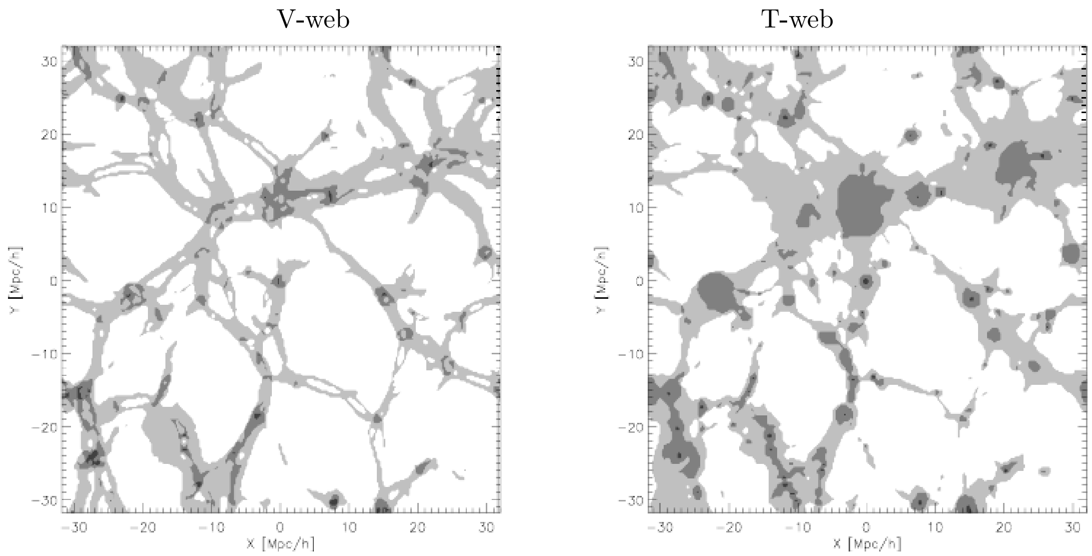


Figure 10: V-web (left) and T-web (right) classification of the same simulation volume. The same colour codes are used as in figure 9. Notice (i) that the V-web classification resolves compact features far better; and (ii) the double-spined appearance of underdense walls in the V-web.

of the nature of the potential field.

Recall from section 2.6 — equation 72 — that the gravitational potential is determined by the matter distribution on large scales. This has an important consequence for the identification of cosmic web structures based on the tidal tensor. At a given point, some distance away from a massive region — e.g. in a void, neighbouring a filament — this classification method will detect an anisotropic gravitational force, due to the long-range effect from that filament. A large part of the void will thus be classified as a filament.

Velocity flows are dominated by comparatively smaller ranges, and thus offer a finer resolution as tracer of cosmic web structures.

V-web: Web classification by velocity shear

Hoffman et al. (2012) formulated a method, based on velocity shear, rather than the tidal tensor. They coined the term "T-web" to refer to web classification based on the tidal tensor, while the term "V-web" is introduced to denote the velocity as basis for classification. Given the relationship between the tensors used, these two methods yield identical results in the linear regime, and deviate towards nonlinear structure growth. Specifically — see equation 35 —

$$\mathbf{v} = \frac{2f}{3H\Omega}\mathbf{g}, \tag{89}$$

which implies

$$\sigma_{ij} = \frac{2f}{3H\Omega}T_{ij}. \tag{90}$$

Hoffman et al. find that web classification based on the tidal tensor with a threshold λ_{th} does not resolve structure on sub-Mpc scales. Density goes nonlinear faster than velocity or gravity — see Kitaura et al. (2012) for an explicit study. This leaves the velocity field as a better descriptor of the initial conditions; and the initial conditions determine the appearance of the cosmic web (Zel'dovich, 1970; Bond et al., 1996).

This motivates a classification based on the non-traceless shear tensor Σ_{ij}:

$$\Sigma_{ij} = -\frac{1}{2H_0}\left(\frac{\partial v_i}{\partial r_j} + \frac{\partial v_j}{\partial r_i}\right). \tag{91}$$

The V-web classifier counts the number of eigenvalues Σ_{ij} possesses above a free threshold λ_{th}^{V}. Hoffman et al. find this classifier to be "much superior [...] compared with the T-web". See figure 10 for a comparison. They find walls to be generally underdense — $\delta < 0$ — regions, and filaments to be generally overdense[28]. Pogosyan et al. (1998) have found a correspondence between the density and local geometric properties of the shear field.

Interestingly, the V-web identifies some underdense sheet-like structures that appear like parallel caustics enclosing an inner planar region with positive divergence[29]. Aragón-Calvo et al. (2010); Hoffman et al. (2012); Rieder et al. (2013) find sheets to contain partially collapsed filaments, themselves containing very compact non-linear nodes. The V-web classifier boasts a nearly perfect match between spatial distributions of density and velocity divergence, which is in agreement with full nonlinear theory of gravitational instability. The occurrence of converging flows in underdense regions, then, are seen as purely nonlinear effects, and classified as walls.

Parameters

When developing a model for physical phenomena, it is preferable to rely on as few free parameters as possible. Ideally, the values adopted for all parameters used in a model can be unambiguously justified — e.g. from first principles. A model that manages to be descriptive of a wide range of phenomena without requiring that many variables are 'tweaked' to fit observations is a strong model. The issue with the structures in the cosmic web is that they are — even by eye — hardly well-delineated. Boundaries between cosmic web components are gradual, and an unambiguous framework dividing them has not been formulated. It is for this reason that various free parameters had to be incorporated in the web classifiers described above.

The *filter scale* is the length scale at which the input field is filtered. While this is a free parameter in most web classifiers, it does reflect the fact that structures occur at various scales. Rather than introducing an arbitrary choice inherent to the cosmic web model, regulating the filter scale allows us to probe ranges of the scale hierarchy.

The fact that an *eigenvalue threshold* λ_{th} appears to be necessary in order to prevent misclassification is a direct consequence of the ill-defined nature of the cosmic web. The cosmic web does contain structures that collapse only very slightly along a given direction, which should not realistically be counted towards its dimensionality. For example: our study finds many large void regions

[28] In this study, we define walls to be overdense oblate features; underdense oblate features are called *oblate collapsing void regions*.

[29] Our study reproduces this result with a somewhat different web classifier — see section 6 — we dub this artefact "double-spined" features.

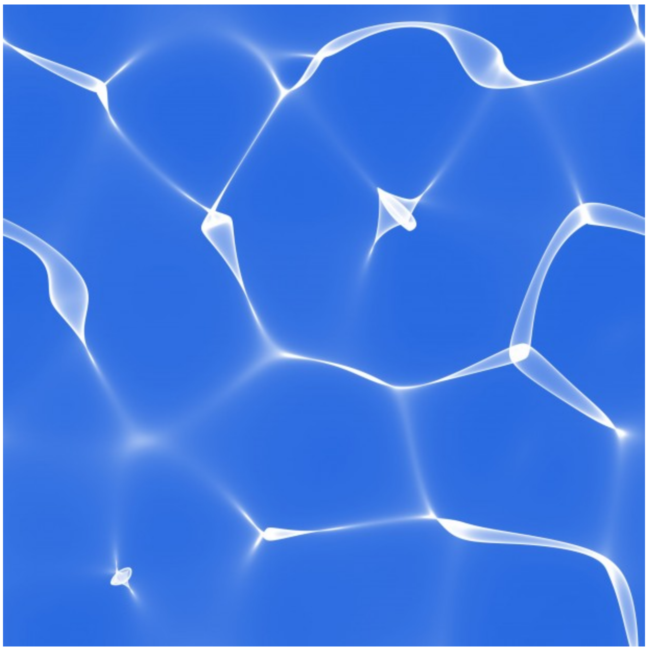


Figure 11: Illustration of white caustic lines on a blue background. Image from the public domain, created by Piotr Siedlecki.

that collapse very slightly in one direction, and are thus classified as oblate collapsing void regions. The effects of various choices for the threshold eigenvalue are illustrated in figure 9 — from Forero-Romero et al. (2009), who had to choose λ_{th} empirically.

The NEXUS(+) algorithms rely only on a lower filter scale, and a total dynamic range determined by the number of filter scales. As long as these parameters are chosen such that they span a sufficiently broad range, the results should have no dependence on those choices. The signatures for cosmic web components are determined via procedures that do not depend on any free parameters.

The web classifier used in this study is based on the non-traceless shear tensor, and makes no arbitrary assumptions to the eigenvalue threshold — it takes $\lambda_{th} = 0$. Details on this method are presented in section 5.4, and a discussion can be found in section A.

2.9 Shell crossing and the generation of vorticity

In the strongly nonlinear phase of structure evolution, it is possible for different velocity flows to overlap spatially. A mass element's *Lagrangian position* $\mathbf{q}$ is defined as its initial Eulerian position

$$\mathbf{q} \equiv \mathbf{x}(t = 0). \tag{92}$$

Throughout the linear regime, the mapping between Lagrangian and Eulerian positions is *bijective* for each mass element: the Lagrangian position can be

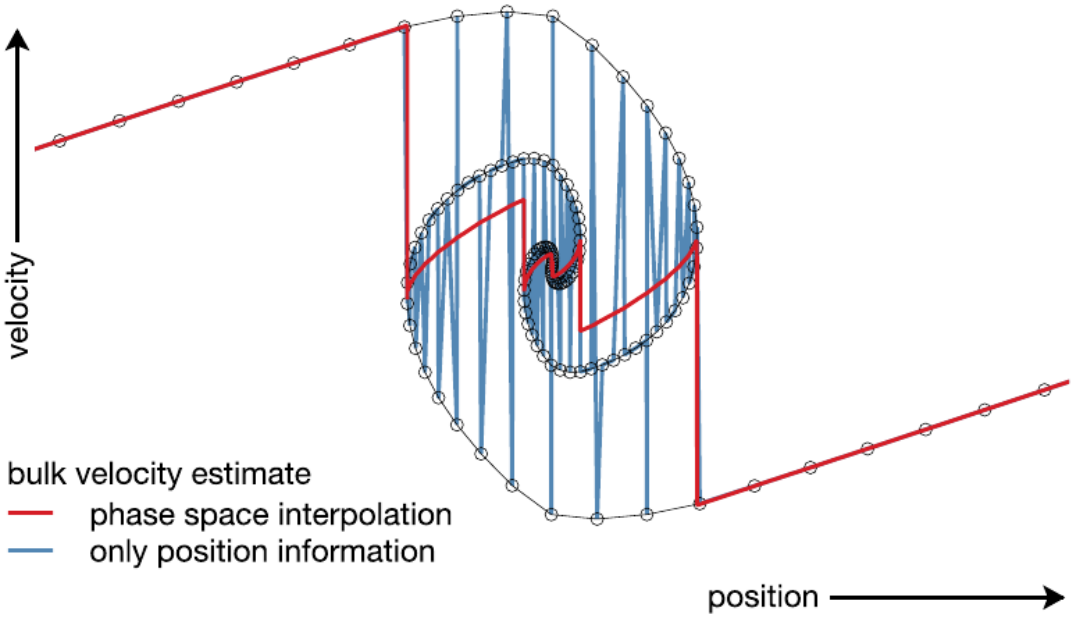


Figure 12: This illustration shows a section of the dark matter sheet in one spatial dimension (horizontal) and velocity vertically. The sheet links points that are close in Lagrangian space. The blue line shows the result of an interpolation using only physical positions and discarding all velocity information. An interpolation in phase space — red line — is far smoother. Source: (Hahn et al., 2015)

unambiguously retrieved from solely the Eulerian position. This no longer holds when velocity flows overlap, e.g. as particles i and j coincide,

$$\mathbf{x}(\mathbf{q}_i) = \mathbf{x}(\mathbf{q}_j). \tag{93}$$

The overlap of velocity flows is referred to as *shell crossing* and *multi-streaming*. This results in sharp density enhancements, spatial patterns equivalent to *caustics* formed by light rays that are bent through a wavy translucent material. See figure 11 for an illustration.

In essence, the Zel'dovich formalism (Zel'dovich, 1970) of mass element trajectories originating at an initial potential field — with some spatial correlation — already predicts the formation of three-dimensional caustics — see equation 54. The Zel'dovich approximation predicts that flattened caustics form first, followed by elongated ones. It is in these places that strong nonlinear effects on the velocity flows occur. In the general case, the approximation of the velocity field as a purely potential flow breaks down at this point, and vorticity can be generated.

Measurements in phase space

As different streams cross each other, a mere interpolation — e.g. Delaunay estimation, see section 4.5 — between spatially neighbouring particles implies a mixing of estimations from different streams. A potentially broad range of veloc-

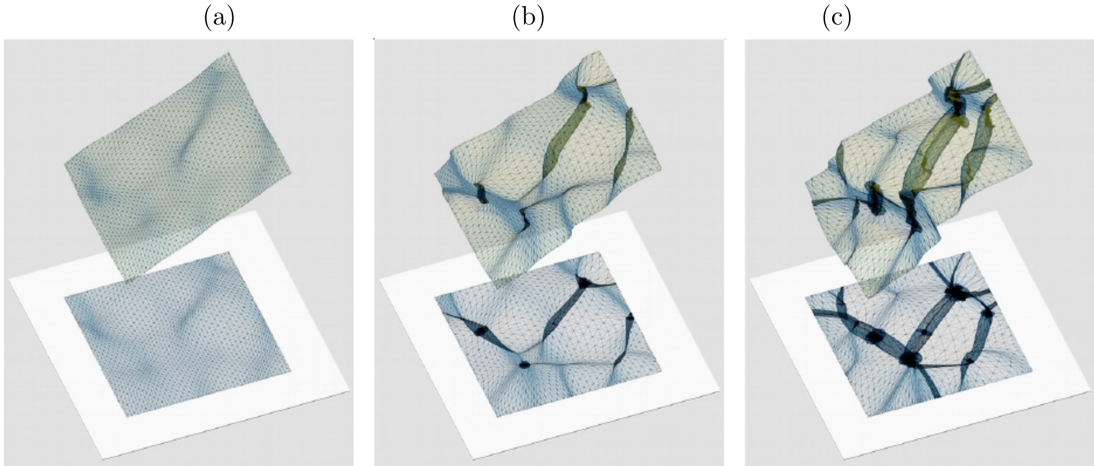


Figure 13: Projection of the dark matter sheet onto a two-dimensional Eulerian space. Panel (a) shows an early time frame, panels (b) and (c) show consecutively later ones. Notice the buckling of the dark matter sheet, and the corresponding formation of multi-stream regions in Eulerian space. Images created by Johan Hidding.

ities, with no discernible spatial structure, leads to great spurious fluctuations in the velocity measurements, this is illustrated by the blue line in figure 12 (Hahn et al., 2015). In 2012, Abel et al. developed a method to disentangle the velocities from different streams and make clean velocity estimations.

This method employs the Lagrangian coordinates of particles, to link those in any stream together, into a *dark matter sheet.* Figure 12 gives a schematic view of a dark matter sheet folded in a phase space rendered in one spatial dimension and one velocity. The crossing of streams is equivalent to the twisting of the sheet, so that streams with different velocities occur at the same spatial location — figure 13 gives a visual impression of how this forms caustics in Eulerian space. Abel et al. made use of the fact that particles linked together in the sheet are also neighbouring in Lagrangian space. A Delaunay tessellation — see section 4.5 — is made of the particles in Lagrangian space, and its connectivity is remembered throughout the evolution of the system. Interpolations between measured particle quantities are then carried out on the basis of these tetrahedra, even though they generally no longer satisfy the Delaunay condition[30]. The quantity f at position $\mathbf{x}$ is then estimated by averaging the interpolated quantities $\hat{f}^{(i)}(\mathbf{x})$ from all the tetrahedra i containing $\mathbf{x}$. This averaging is weighted by the density $\hat{\rho}^{(i)}$ at the tetrahedron:

$$\langle f\rangle(\mathbf{x}) \simeq \frac{\sum_i \hat{\rho}^{(i)}(\mathbf{x})\hat{f}^{(i)}(\mathbf{x})}{\sum_i \hat{\rho}^{(i)}(\mathbf{x})}. \tag{94}$$

This results in a much smoother estimation — one “respecting the phase-space connectivity” — illustrated by the red line in figure 12.

Now that an operator $\langle\ .\ \rangle$ for estimating — or *projecting* onto physical

[30] id est: the tetrahedra used for linear interpolation generally become stretched into less compact shapes. The cost is that this brings greater uncertainties in the interpolation — see section 4 for details.

space — is defined, any derivative of the projected quantity must also include the derivative of the projection operator. Hahn et al. (2015) provide the following expressions for the divergence:

$$\nabla \cdot \langle \mathbf{v} \rangle = \langle (\nabla \log \rho) \cdot (\mathbf{v} - \langle \mathbf{v} \rangle) \rangle + \langle \nabla \cdot \mathbf{v} \rangle, \tag{95}$$

and for the vorticity:

$$\nabla \times \langle \mathbf{v} \rangle = \langle (\nabla \log \rho) \times (\mathbf{v} - \langle \mathbf{v} \rangle) \rangle + \langle \nabla \times \mathbf{v} \rangle. \tag{96}$$

In both expressions, the first term is "purely due to the projection of a multi-stream field [...], and [...] reflects alignment of velocities with density gradients". Note that the second term in the vorticity expression vanishes for a potential flow. The vorticity "is a property of the projected velocity field alone". These derivatives of the bulk velocity field are well behaved across the discontinuities in caustics.

In the practice of analysis of the vorticity field, this projection effect has an important implication. The observable is the left-hand-side of equation 96, which is not purely the vorticity $\langle \nabla \times \mathbf{v} \rangle$, but rather the vorticity contaminated by the projection term. This requires caution to disentangle the two from any vorticity measurements. Failure to do so will result in a — partial or entire — false vorticity signal[31].

In collapsed structures in their simulations, Hahn et al. found the correlation between density and velocity divergence to reverse after shell-crossing. From linear theory it follows that $\nabla \cdot \mathbf{v} \propto -\delta$, but inside caustics, high mass predominantly induces a larger divergence. Only in the innermost regions of collapsed structures is the correlation still consistently negative.

An analysis of the effects of the stress tensor — i.e. extended perturbation theory — through the new lense of the phase space sheet tessellation method has yet to be carried out.

[31]It is with the eye on this caveat that the present study remains reserved in drawing conclusions from vorticity analysis.

3 Cosmological Simulations

N-body simulations are an inherently Lagrangian approach to simulating structure formation. In such a simulation, the trajectories of a large number of particles are approximated, by calculating the forces acting on the particles at each step in a discretised time line. The types of simulations that have been carried out are very diverse. They can contain various combinations of cosmological components, account for various combinations of physical effects, and work by various numerical methods that have been developed in recent decades.

N-body simulations have been used for the investigation of all aspects of structure formation: galaxy cluster populations, gravitational lensing, baryonic acoustic oscillations, internal properties of dark matter haloes, halo formation and evolution, statistical properties, etc. (see e.g. Evrard et al., 2002; Wambsganss et al., 2004; Springel et al., 2005; Teyssier et al., 2009; Kim et al., 2009; Crocce et al., 2010; White et al., 2010; Klypin et al., 2011; Ishiyama et al., 2013; Vogelsberger et al., 2014).

3.1 Parameters

Resolution and Dynamic range

The particles in a simulation are usually confined to a box of a fixed size, with periodic boundary conditions. The *spatial resolution* applies to grid-based methods that the simulation relies on — like field estimation and the PM method; see sections 4 and 3.2. The *mass resolution* of a simulation is a measure of the particle mass. For objects containing any given total mass, a simulation with higher mass resolution will render it as being comprised of a larger number of individual particles. Conversely, too high a particle mass does not allow the dynamical processes of low-mass objects like small haloes and dwarf galaxies to be resolved to great detail.

The *dynamic range* of a simulation is the difference between the smallest and largest scales at which structures are simulated. A large dynamic range is computationally costly, as it requires a sufficiently high resolution as well as a sufficiently large box. Simulations with a large dynamic range allow the simultaneous investigation of large and small scale processes happening in the same volume. This makes it possible, for example, to study the effects of structures across scales; and to generate a large bias-free sample of small scale objects from the same environment (Ishiyama et al., 2013). Given a total dynamic range that can be simulated with available computation power, there is still the choice as to which scales to focus on. To illustrate the possibilities: the *millennium* simulation (Springel et al., 2005) and the Cosmogrid simulation (Ishiyama et al., 2013) both have a roughly comparable dynamic range — both contain comparable numbers of particles. However, while the millennium simulation volume spans 500 Mpc in each direction, and is limited to a mass resolution of $1.2 \cdot 10^9 M_\odot$; Cosmogrid spans only 30 Mpc and consists of $1.28 \cdot 10^5 M_\odot$ particles.

The scale of homogeneity — see also the introduction — indicates the distance scale above which structure appears to be homogeneous. Estimates of this scale at the present cosmological epoch vary from 70 Mpc to several hundreds of Mpc. Boxes larger than the scale of homogeneity bring the benefit of enclosing a representative portion of the cosmology they implement — a sufficiently large box can contain a representative sample of even the rarest objects, and large-scale processes like velocity flows. A box size that is appreciably shorter than the scale of homogeneity introduce the risk of containing a volume that is not representative of the intended cosmology — for example, a small box may happen to be dominated by a large central void, causing the overall expansion to be greater than the Hubble expansion[32].

Another disadvantage of a limited box size is what we call the *fundamental mode problem*, which is discussed in section A.3. In brief, all structures need a box significantly larger than their spatial scale, in order to be represented well in Fourier space — all structures have a maximum wavelength contribution to the Fourier transform, called the *fundamental mode*. The fundamental modes for the largest structures require a large box to be reliably rendered. Particularly at low redshifts, large structures appear, and a 30 Mpc box is insufficient for rendering them accurately. Appendix A.3 presents a more detailed explanation.

As for resolutions, it stands to reason that high spatial and mass resolutions make it possible to resolve small objects. For an accurate depiction of the generation of nonlinear velocity flows, a high resolution is crucial (Pueblas & Scoccimarro, 2009) as they occur at very small scales. The application of Fourier operations[33] at low spatial resolutions introduces the risk of *aliasing* — discussed in section A.3 as well. For the simulation of small scale structures, a sufficient *temporal resolution* is of importance, too. This is because they have relatively high overdensities, which implies small dynamical time scales (Springel et al., 2005).

While box sizes are preferably large, and resolutions are preferably high, they come with a certain computational cost. Based on the amount of available computing power, one has to decide upon a trade-off between the two: between representativeness and accuracy of large scale structures on one hand, and resolution of small scale objects and nonlinear processes on the other. The box size and resolution can be therefore chosen on the basis of the goals of the simulation.

Components

While the gravitational interactions between cold dark matter particles are by far the dominant driver of structure formation, a more realistic simulation may be obtained by the inclusion of other components, as well as the implementation of other physical interactions. The Universe around us contains baryonic matter, radiation, and dark energy. Among the physical processes that are known to be of influence to structure formation are pressure, dark energy, relativistic gravity, magnetic forces, radiative processes, hydrodynamic processes, star formation, and various feedback interactions (Iliev et al., 2006; Dolag & Stasyszyn, 2009; Li & Barrow, 2011; Vazza et al., 2014; Marinacci et al., 2015; Schaye et al., 2015). When these processes are incorporated in a simulation, it is often in some simplified form, they are referred to as *sub-resolution physics* (Springel & Hernquist, 2003).

The degrees to which these processes influence structure formation vary between time scales — see the *gastrophysics* treatment in Baugh (2006), and references therein. A recent example of advanced N-body simulations including gas

[32] And this happens to be the case for the Cosmogrid simulation.

[33] e.g. convolution or differentiation via Fourier space.

physics is the *illustris* project (Vogelsberger et al., 2014). It has over $18 \cdot 10^9$ particles in a $(106.5 \text{ Mpc})^3$ box, with a mass resolution in the order of $10^6 M_\odot$, and hydrodynamics resolved down to 48 pc cells. It simulates chemical enrichment of gas, and stellar evolution, to name just a few of many aspects.

Cosmology

The currently widely accepted ΛCDM model has it that the Universe consists mainly of Cold Dark Matter and Dark Energy (Peacock, 1999). It specifies the present matter content Ω_0 to be 0.3, the Dark Energy content Ω_Λ to be 0.7, and the Hubble parameter $H_0 = 70$ km/s/Mpc. The *mass fluctuation parameter* σ_8 indicates the amount of fluctuation in density over a distance of 8 Mpc, and is determined to be $\sigma_8 = 0.8$. The power spectrum of initial density perturbations is assumed to be a power law with index $n = 1.0$ — see section 2.6 for a background.

These parameters are set when simulating a Universe. It would be the most sensible thing to match one's simulated cosmology with the parameters from observations. The most recent and precise measurements to date come from the Planck mission (see e.g. Planck Collaboration et al., 2015a). The ΛCDM model with these parameters fits well with the temperature and polarisation measurements from the Cosmic Microwave Background.

In the practice of designing simulations, the statistical properties of the initial conditions are specified by a power spectrum — see section 2.6. The power spectrum is commonly assumed to be a power law with some spectral index $n \simeq 1.0$ The integration of positions and velocities in comoving space — see the next subsection — is dependent on the values for Ω_0 and Ω_Λ.

3.2 Methodology

The core operation in any N-body simulation is calculating the evolution of every particle's position and velocity at every time step. Particles obey the following equations of motion:

$$\begin{cases} \dfrac{\partial \mathbf{x}}{\partial t} \propto \mathbf{v} \\ \dfrac{\partial \mathbf{v}}{\partial t} \propto \mathbf{F} \end{cases} \tag{97}$$

which govern the positions $\mathbf{x}$ and velocities $\mathbf{v}$ of a particle, depending on the total accelerating forces $\mathbf{F}$ acting upon it. This operation can be broken down into two steps: (i) integrating the particle positions and velocities across discrete time frames, and (ii) calculating the forces necessary for the next iteration.

Integration

For a computer to integrate the particle trajectories through a discrete set of time steps, the method of *leapfrog integration* is very suitable — see figure 14 for an illustration. The essence of this approach is that updates to the positions $\mathbf{x}$ and velocities $\mathbf{v}$ take place alternatingly in time — e.g. positions at integer time steps $\mathbf{x}_i$ and velocities at odd-half-integer time steps $\mathbf{v}_{i \pm 1/2}$. Updating these quantities follows these equations:

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{v}_{i-1/2} \cdot \Delta t; \tag{98}$$

$$\mathbf{v}_{i+1/2} = \mathbf{v}_{i-1/2} + \mathbf{F}(\mathbf{x}_i) \cdot \Delta t, \tag{99}$$

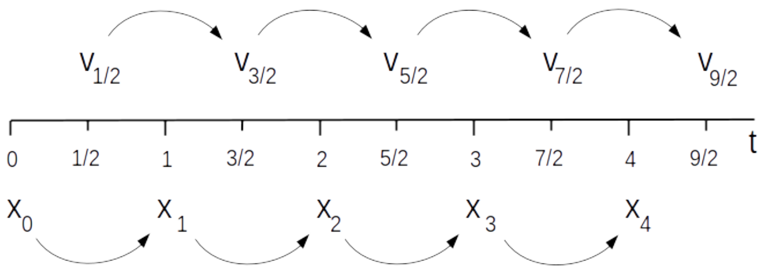


Figure 14: Visual diagram of leapfrog integration. Positions are updated at whole integer time steps, velocities at odd-half-integer steps. The fact that the two quantities are updated alternatingly lends stability to the results.

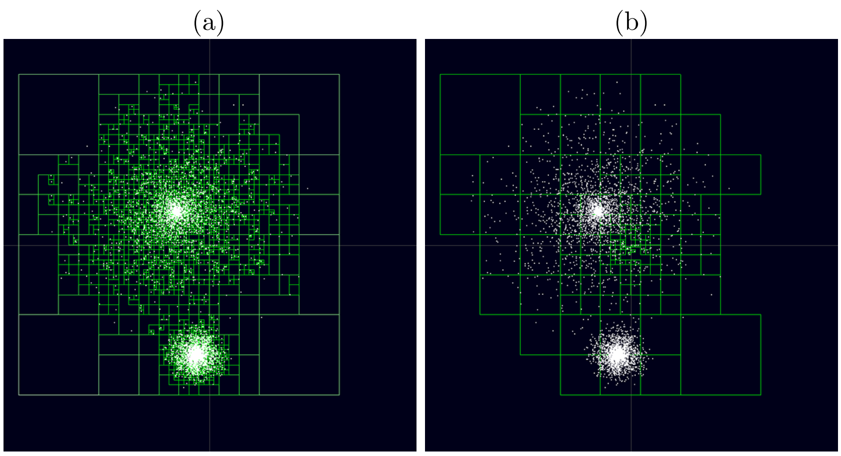


Figure 15: Two-dimensional visualisation of the tree code (Barnes & Hut, 1986). Both panels show a simulation body comprising 5000 particles. The left panel shows the entire tree with $N_{leaf} = 1$, so that each subcell contains at most one particle — note that in this case cells are divided by four, not eight as in the three-dimensional case. The right panel shows a tree of only 146 subcells, which are used for the force calculation on a particle in the middle of the diagram. Small and distant ensembles of particles can be taken together into few subcells without a substantial loss of accuracy. Diagrams are from wikimedia commons, created by eclipse.sx

where Δt is the duration of the time step, and $\mathbf{F}(\mathbf{x}_i)$ is the force acting upon a particle at position $\mathbf{x}_i$ at time step i. These equations are easily rewritten so that updates can be carried out simultaneously, once per time step. Then, the method is still founded upon positions and velocities *playing leapfrog* with one another, which brings a few crucial advantages over other integration schemes: firstly, it is a second order yet very simple integrator. Also, it prevents the unbounded increase of particle positions and energies[34]. One requirement is that the time step Δt is constant (Birdsall & Langdon, 1985). For the purposes of cosmological simulations in comoving space, integration is conducted over a discretised sequence of steps in the expansion factor Δa, rather than time. This means a nonlinear progression of redshift.

In each and any time step of a simulation, the forces acting upon all particles need to be calculated. In the case of long range forces — like gravity — all particles influence all other particles to some degree. For a discrete time step, it is possible to calculate the forces from every particle onto every other particle in the system — to this end, Aarseth (1963) has developed the Particle-Particle (PP) method — with an arbitrary number of particles N_p. The problem, though, is that the number of calculations required to do this scales with N_p^2.

Hockney & Eastwood (1981) have described a method that can replace the PP calculation on large ranges. It sacrifices some accuracy at small scales for a great acceleration of the computation. This is called the Particle-Mesh (PM) method, and it involves determining a potential — and then a force field — at a regularly distributed set of mesh points, by solving the Poisson equation on a grid, via Fourier transformation. From there, the forces acting on particles are interpolated between the mesh points. Hockney & Eastwood (1981) have suggested that the PM method be used in combination with an additional scheme acting upon individual particles at smaller scales, with forces confined to a cut-off radius r_{cut}. A combination with the PP method is referred to as the PPPM or P^{3}M method. See also Efstathiou et al. (1985); Efstathiou (1986a,b).

A technique that can be used to supplement the PM method for calculating small scale particle forces was developed by Barnes & Hut (1986). Their algorithm erects a hierarchical tree of cells: each cell containing more than a set number N_{leaf} of particles is divided into eight cubic subcells — and this is done recursively, until all cells contain at most N_{leaf} particles. Figure 15 gives a two-dimensional visualisation of this.

In principle, method for force calculation is a dicretised version of a multipole expansion; it depends on an *opening angle* θ, which determines the threshold size and distance of particle clusters to be incorporated in the calculation. The tree of subcells is built from scratch at every time step. Using these subcells for particle force calculations brings the advantages of being accurate at a large range of spatial scales, and being far less biased by grid axis orientations. Furthermore, the number of calculations required this way scale only as $N_p \log N_p$. This method was combined with the PM algorithm in the development of the

[34]because it is a *symplectic* integrator: it conserves the area spanned by the particles' velocity and momentum vectors in phase space (Ruth, 1983).

TreePM technique (Xu, 1995; Bode et al., 2000; Bagla, 2002; Bode & Ostriker, 2003; Yoshikawa & Fukushige, 2005; Khandai & Bagla, 2009).

Division of labour

The most computationally demanding simulations benefit from being run by a cooperation of several computers in parallel. In this case, each of the processes must cover a sector of the simulation volume, and communicate long range forces as well as the positions of particles crossing the sector boundaries. To this end, Ishiyama et al. (2009) have developed a parallelised TreePM algorithm called GreeM[35]. In the GreeM procedure, each process determines a density on the PM grid points by means of the TSC — see section 4.2 — estimator, and exchange the density estimates with all other processes, to interpolate the PM forces on all particles. Next, the processes construct the trees from particles in their own sectors, and calculate the PP forces on those particles; this entails the communication of tree information stemming from within the PP cutoff radius r_{cut} from particles in other sectors. When the displacements resulting from PM and PP forces send particles across sector boundaries, they are adopted by the appropriate process (Ishiyama et al., 2009).

A more recent method of parallelisation for N-body simulations is the SUSHI code (Groen et al., 2011). Like previous methods, it works by a combination of PM and tree methods, but the distribution of sectors over different processes is reevaluated at every time step, to maintain a balanced distribution of labour. At the macro level of computers at remote locations, SUSHI connects each site with only two others, resulting in a ring topology. The Cosmogrid simulation marks an important step in the spreading of N-body simulations as a globally operated endeavour.

Gastrophysics and the Gadget code

When a simulation includes gas, radiative cooling and star formation, the Gadget code — for *GAlaxy with Dark matter and Gas intEracT* — is a suitable framework. The code was introduced in 2001 (Springel et al., 2001) and a new version was published in 2005 (Springel, 2005). Currently, it is a contribution of key importance to the field of cosmological N-body simulations.

Gadget uses the Tree algorithm, but with Smooth Particle Hydrodynamics (SPH) (Monaghan, 1992; Springel, 2010) — see also section 4.2. The SPH scheme aims to solve the fluid equations on the basis of particles. A fluid is represented by smoothed, interacting particles with thermodynamic properties. This is achieved by a smoothing kernel that can adapt in size for every particle separately, depending on the sampling density in its immediate environment. This allows the use of small kernels in high-density areas and large kernels in low-density areas, resulting in a better resolution in dense regions as well as low noise levels in sparse regions.

An extension to the Gadget code has been made to incorporate radiative transfer in the simulation (Petkova & Springel, 2009). The Gadget code formulates a specific file format, where particles of various types and masses are stored. At present, the Gadget format is widely used for the storage and communication of simulation data.

[35] where the 'G' stands for GRAPE (Sugimoto et al., 1990; Makino & Taiji, 1998; Makino et al., 2003), a system for facilitating parallel computation between units with high latency, e.g. supercomputers located continents apart.

It is by virtue of the *Zel'dovich formalism* that we can construct the initial conditions for an N-body simulation in an efficient and natural way (Klypin & Shandarin, 1983). The procedure starts with the generation of a random Gaussian initial density field[36] $\delta_0(\mathbf{q})$, with the desired power spectrum $P(k)$. The Zel'dovich displacement field $\boldsymbol{\psi}(\mathbf{q})$ is determined by the density Fourier components of the field:

$$\boldsymbol{\psi}(\mathbf{q}) = \int \frac{d\mathbf{k}}{(2\pi)^2} \left(-i\frac{\mathbf{k}}{k^2}\hat{\delta}(\mathbf{k}) \right) e^{-i\mathbf{k}\cdot\mathbf{q}}. \tag{100}$$

The Fourier components $\hat{\delta}(\mathbf{k})$ are conveniently determined if the positions are placed in a regular cubic grid[37]. Next, the initial particle positions are calculated by a small Zel'dovich displacement,

$$\mathbf{x}(\mathbf{q}) = \mathbf{q} + D_\delta(t)\boldsymbol{\psi}(\mathbf{q}), \tag{101}$$

and the particles are given an initial velocity

$$\mathbf{v}(\mathbf{q}) = aDHf(\Omega)\boldsymbol{\psi}(\mathbf{q}), \tag{102}$$

from where the integration can take off.

The initial density values $\delta_0(\mathbf{q})$, are to be drawn from a Gaussian distribution with the specified power spectrum (Bardeen et al., 1986)

$$\langle \hat{\delta}(\mathbf{k})\hat{\delta}^*(\mathbf{k}) \rangle \propto P(\mathbf{k}). \tag{103}$$

The construction of a GRF with specific constraints — e.g. 'proto-voids' or density peaks with a specific height shape, and such constraints to reproduce the observed universe (van de Weygaert, 1991; van de Weygaert & van Kampen, 1993; Ganon & Hoffman, 1993) — is no trivial task (Binney & Quinn, 1991). Bertschinger (1987); Hoffman & Ribak (1991) developed an elegant algorithm that works by constructing analytically the mean constraint field, and adding a statistically independent residual GRF. van de Weygaert & Bertschinger (1996) have implemented this *Hoffman-Ribak method* by calculations in Fourier space.

Bertschinger (2001) has developed a computational package called GRAFIC — facilitating the generation of Gaussian RAndom Field Initial Conditions. A Gaussian random density field $\delta(\mathbf{x})$ is made by the convolution of *white noise* $\xi(\mathbf{x})$ with a *transfer function* $T(\mathbf{x})$ (Salmon, 1996). White noise is a very plain type of field, with power spectrum

$$\langle \hat{\xi}(\mathbf{k}_1)\hat{\xi}(\mathbf{k}_2) \rangle = \delta_D^3(\mathbf{k}_1 + \mathbf{k}_2); \tag{104}$$

the transfer function is related to the power spectrum of the initial density field. Recall from section 2.6 that the primordial power spectrum is expected to be of the Harrison-Zel'dovich type. The transfer function incorporates the evolution

[36] note that $\mathbf{q}$ denotes the initial — or *Lagrangian* — positions.

[37] For those bothered by the risk of orientation-bias due to a regular grid, there is the alternative of *glass initial conditions* (White, 1996), where a homogeneous random set of initial particle positions is determined by a number of Lloyd iterations (Lloyd, 1982) upon a random spatial sample of generating points.

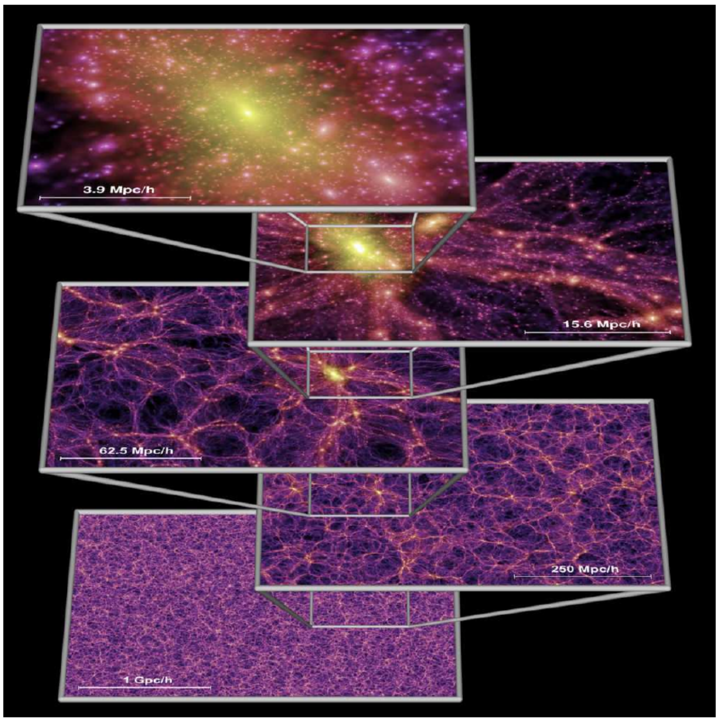


Figure 16: Dark matter density from five different slices of the Millennium simulation — at $z = 0$. Each slice is $15h^{-1}$ Mpc in thickness, and planar scales are given. The lower two images comprise several slices from the periodic box — sliced at an angle, to avoid repetition of structures. The faintest galaxies of $0.1L_\star$ consist of at least 100 particles.

of the power spectrum, reflecting the physical processes in structure formation. The convolution is discretised on an *adaptive mesh* grid with periodic boundary conditions, the range of length scales that can be represented depends on the size and resolution of the grid, see section A.3 for a description.

Prunet et al. (2008) have extended this GRAFIC package so that it can be used by various computing platforms, simultaneously generating initial conditions for one and the same simulation. This is done by a Peano-Hilbert decomposition of the simulation volume, which assigns a simply connected region of space to an arbitrary number of computing processes. Since the Message Passing Interface library was used for communication, Prunet et al. have dubbed the extended package MPGRAFIC.

3.3 Cosmogrid

Predecessor — the Millennium simulation

The Millennium simulation (Springel et al., 2005) was conducted by the Virgo Consortium in 2005, as a method for testing the inflationary Λ cold dark matter model. Springel et al. implemented new techniques to follow the evolution of structures in the simulation. For the aim of comparing simulation results to

Parameter	Value
Matter density parameter (ω_0[sic!])	0.3
Cosmological constant (Ω_Λ)	0.7
Hubble constant (H_0)	70.0 km/s/Mpc
Mass fluctuation parameter (σ_8)	0.8
Box size	(30 Mpc)^3
Softening scale	175 pc

Table 1: Cosmological parameters used in the Cosmogrid simulation. Source: (Groen et al., 2012).

redshift surveys, the simulation is required to be large enough to representatively sample rare objects; but at the same time, small scale structures must be resolved as well. As an answer to this challenge, the Millennium simulation spans a volume of (500 Mpc)^3, containing 2160^3 particles of $1.2 \cdot 10^9 M_\odot$ each. Figure 16 gives a taste of the dynamic range achieved.

The study of galaxy formation and evolution also requires a treatment of physical aspects like the interstellar medium, star formation, galactic winds, metallicity, feedback from Active Galactic Nuclei and other aspects. Springel et al. have supplemented the simulation with what they call 'post-hoc' modelling of these physics. The semi-analytic models are applied to the merger trees of haloes in the simulation, and includes a treatment of the dark matter substructure. It works by integrating a number of differential equations for all the galaxies in the merger tree. These equations describe processes like radiative gas cooling, star formation, black hole growth, various feedback processes and more.

The Millennium simulation was conducted with a version of Gadget2 (Springel et al., 2001; Springel, 2005); and force calculations were done with the TreePM method described above. Initial conditions were constructed by a random Zel'dovich displacement of particles in an initial glass configuration (White, 1996) — see also footnote 37.

Cosmogrid

This study uses data from the Cosmogrid simulation (Ishiyama et al., 2013). This simulation offers a very high spatial and mass resolution, since it was created for the main purpose of studying the statistics of small dark matter haloes. While velocity flows are inherently a large scale phenomenon — recall from section 2.6 that

$$P_v(\mathbf{k}) \propto \frac{P(\mathbf{k})}{k^2}; \qquad \text{(equation 71)},$$

— the aim of this study is to investigate the nonlinear formation of structure on small spatial and mass scales. For that reason, we appeal to the Cosmogrid simulation. Its main run contains 2048^3 particles; however, due to limited available computational power, this study is based on a 512^3 particle run.

Cosmogrid — see figure 17 for an impression — is a cosmological N-body simulation, with a periodic box of (30 Mpc)^3, and containing as many as $2048^3 \simeq 8.5 \cdot 10^9$ particles. This is comparable to the number of particles used in the famous *Millennium Run* simulation (Springel et al., 2005). The cosmological volume in that simulation spanned 500 Mpc along each axis. The Cosmogrid body is far more compact.

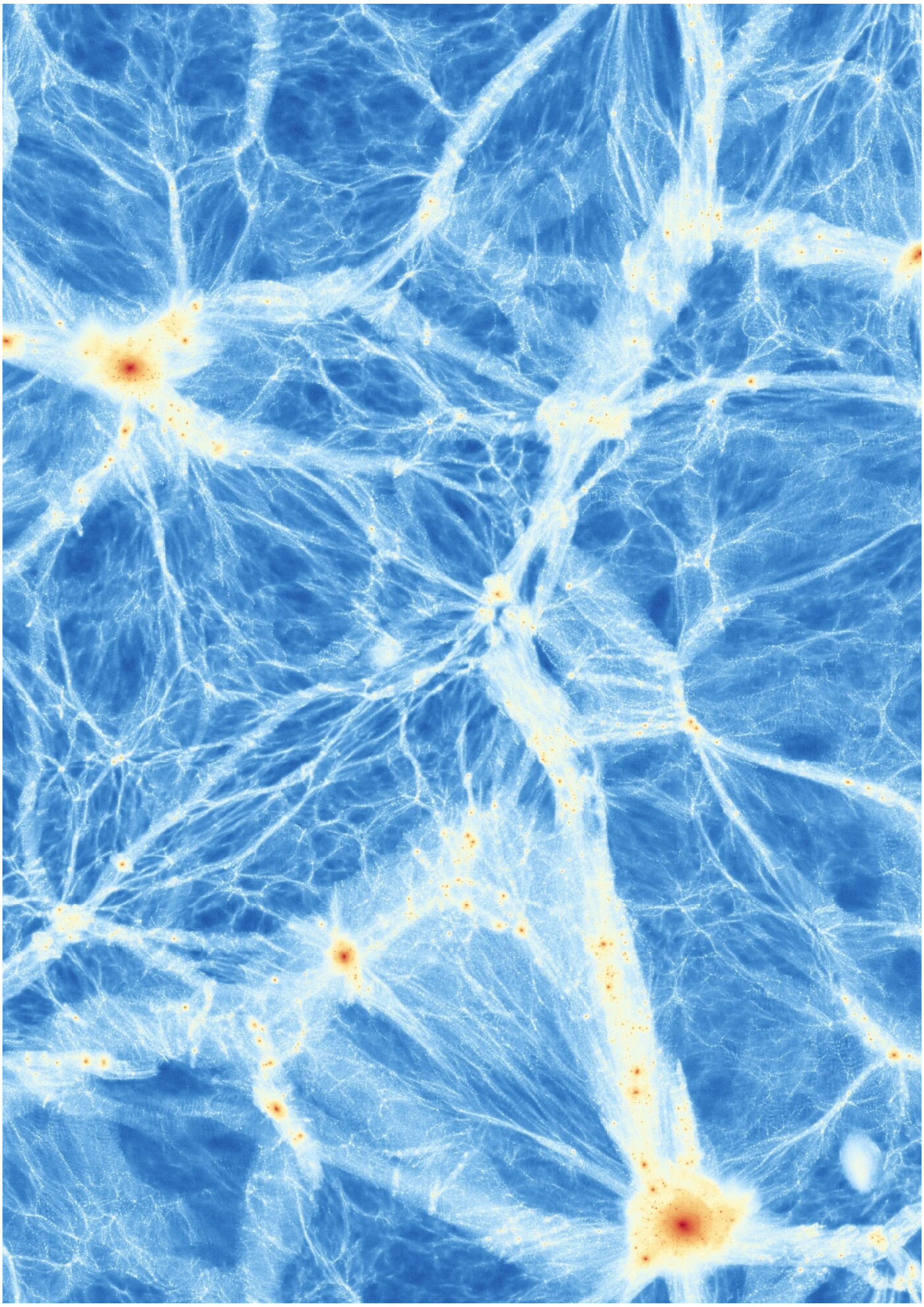


Figure 17: Slice of the Cosmogrid simulation body at $z = 0.0$, the thickness of the slice is 0.6 Mpc. Image by Steven Rieder

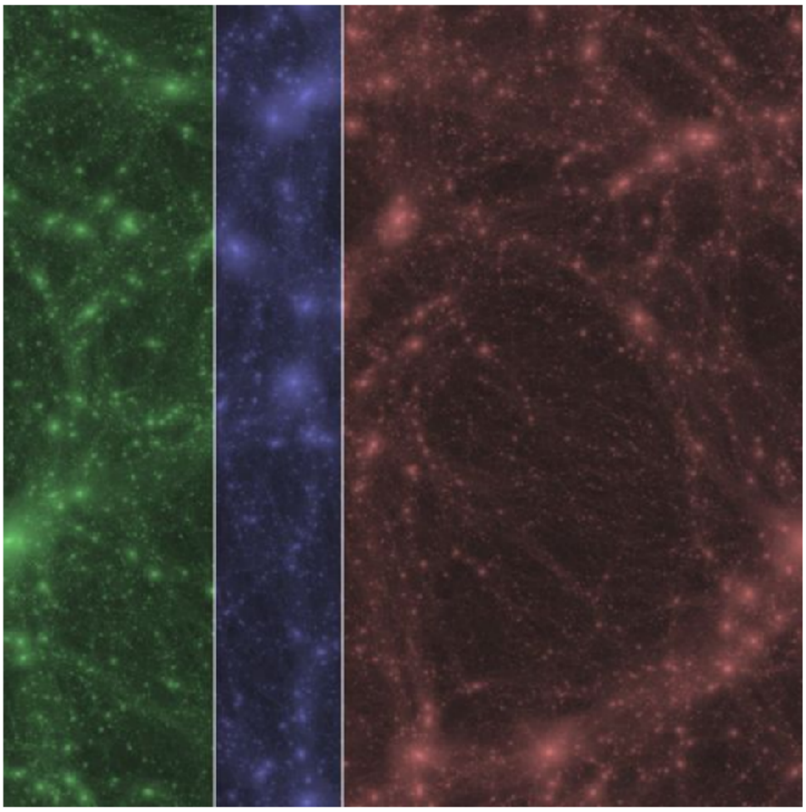


Figure 18: Division of the Cosmogrid simulation volume among the supercomputers in Espoo — green portion on the left — ; Edinburgh — blue in the centre — and Amsterdam — red portion on the right. The image shows the full (30 Mpc)^3 simulation volume at redshift 0. Source: (Groen et al., 2011).

Cosmogrid was designed with the purpose of investigating the low end of the halo mass function, in hopes of relieving the missing dwarf problem. In other studies, detailed simulation of haloes is conveniently achieved by selecting haloes from large-box, low-resolution simulations; and subsequently re-simulating those haloes at higher resolutions. This comes at a cost: namely that the haloes — all coming from a different re-simulation environment — are not statistically representative as a population. Cosmogrid aims to prevent this bias in variations between simulated haloes, by simulating them all in one and the same run. Therefore, a relatively small box size was chosen, with a resolution that allows for the reproduction of star clusters and dwarf galaxy sized haloes. Each particle weighs in at $1.28 \cdot 10^5\ M_\odot$ — though the mass resolution is only $8.12 \cdot 10^6 M_\odot$ for the 512^3 particle run.

The computational load for Cosmogrid has been divided among three supercomputers: Huygens (Amsterdam, Netherlands), Louhi (Espoo, Finland), and HECToR (Edinburgh, Scotland) (Portegies Zwart et al., 2010). The division of computational load among these sites is visualised in figure 18. Cosmogrid uses the GreeM code — introduced in section 3.2 — with 512^3 grid points for the PM calculation. For the communication between the supercomputers, the SUSHI code was used.

The cosmological parameters used in the Cosmogrid simulation are listed in table 3.3 (Groen et al., 2012). The initial redshift in the simulation is $z_{init} = 65$, but this study only uses data from $z = 3.7$ to $z = 0.0$. Within that range, the opening angle θ for the tree code force calculation remained constant $\theta = 0.5$.

Note that the box length of the Cosmogrid volume falls short of the Universe's present scale of homogeneity — estimated from 70 to several hundreds of Mpc, see the introduction and section 3.1 — so the simulation in itself fails to produce a representative portion of our Universe at low redshifts. The limitations due to the small box size are explained in section A.3. Another limitation of Cosmogrid is that it, too, is a dark matter-only simulation.

4 Field Estimators

In practice, fields of quantities resulting from an N-body simulation are represented on a regular cubic grid. When matter densities and velocity flows are sampled at regular grid points, this allows for various kinds on analysis to be performed on the simulation results. Among procedures that require a grid-representation of field quantities are: determining their statistical distributions; calculating forces acting on particles; and computing a discretised Fast Fourier Transform (FFT) of a field.

By nature, N-body simulations sample particle quantities — like masses and velocities — at highly irregular locations, namely the particle positions. Many particles will be clumped together in high-density areas, and large regions will be very sparsely populated. How is a grid representation of field quantities obtained from this? This is no trivial question, and has been addressed in various studies (e.g. Birdsall & Fuss, 1969; Hockney & Eastwood, 1981; Bernardeau & van de Weygaert, 1996; Schaap & van de Weygaert, 2001; Romano-Díaz & van de Weygaert, 2007; van de Weygaert & Schaap, 2009).

Extracting a regular grid of physical quantities from a list of particle data is called *field estimation*, and while there is no perfect way to do this, various techniques offer a wide range of trade-offs between reliability and computability. It must be noted that the representation of data on a regular grid is not physical, and induces artefacts in the analysis — e.g. Fourier aliasing and orientation biases — that must be taken into account.

The core operation in a field estimation involves the 'smearing out' of every particle over a number of grid cells. This is conceptually equivalent to filtering the sample of particles — i.e. convolving the spatial distribution with a certain kernel — and subsequently sampling the result on a set of grid points. Several strategies implement this in different ways. This section will introduce the workings of *grid-based field estimators*, as well as those operating on *tessellations* of the sampling points.

Two qualities are of particular importance to field estimation techniques: (i) resolving variations in the field values at small scales, and (ii) the ability to produce useful estimates across a broad range of sampling densities. High resolution of field estimates is of importance for the investigation of small scale objects and processes, like dwarf galaxy sized haloes and nonlinear velocity flows. This also requires a high grid resolution for the representation of estimated field quantities.

The performance of a field estimator on a broad dynamic range comes into view on larger scales. For example, a grid cell at the periphery of a void may fall in a region with no particles at all, but to estimate the density in that grid cell to be zero would be an unphysical thing, and introduce errors in the measured void density profile. Such cases of under-sampling are more likely to occur when a high grid resolution is chosen, and if there are relatively few particles. Figure 19 illustrates the trade-offs between accuracy and spatial detail, yielded by field estimations on different grid resolutions.

Grid-based and tessellation approaches

Whatever approach is used, in essence it is impossible to translate a list of particle positions and field values to a grid of field values without loss of infor-

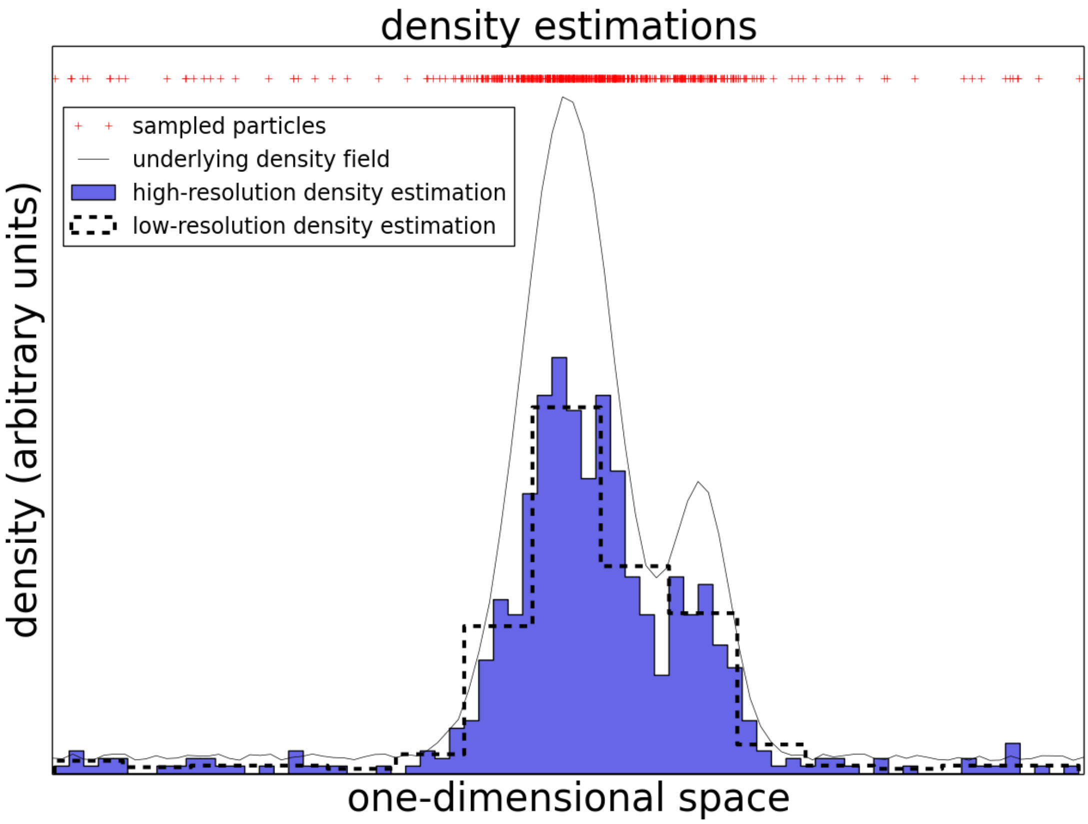


Figure 19: Two different kinds of density estimations in one dimension. The solid grey line shows the underlying density field, and the markers at the top indicate the simulated particle positions — a realisation of the density field. The filled and dashed graphs show density estimations at high and low resolutions, respectively. Notice that the low resolution estimator fails to reproduce the bimodality in the density field, while the high resolution estimator is more subject to noise and occasionally estimates a zero density due to undersampling.

mation and statistical certainty — unless those particles happen to be sitting on those grid points already. The question is how to use the information from the simulated particles optimally to this end — and different solutions to this problem exist, populating a rather broad range in effectivity and creativity.

The class of *Grid-based field estimators* is a relatively straightforward one. These methods function by overlaying a grid on the simulation body, and dividing a particle's physical quantity — like mass or velocity — over a specific selection of nearby grid points, with weights determined by the particle's position.

As a result, the high-density regions will have a rich sampling of the particle properties. Field values can then be estimated with high confidence levels, by interpolating between the nearby particle positions. A problem is that narrow density peaks require a very fine-meshed grid to be resolved. The low-density regions, au contraire, will introduce an undersampling problem, and the estimated field values will be subject to large statistical uncertainties. Additionally, a high grid resolution, while computationally costly, will hardly add any quality to the field estimation in these undersampled regions.

Tessellation algorithms divide a simulation body into irregular cells, determined by the irregular spatial distribution of the particles. They then interpolate the physical quantities within those cells. This approach has the potential to circumvent undersampling and oversampling problems. It naturally defines few and large cells in low-density regions — resulting in larger uncertainties — and makes more spatially detailed interpolations between many smaller cells in high density regions — resulting in large fluctuations.

Various existing algorithms for field estimation require various amounts of computational power and memory. Some field estimators working on large bodies of simulation data can divide the volume spatially into different partitions, to save memory usage. This will introduce a certain inefficiency[38].

4.1 Convolution and adaptivity

As mentioned, a grid-representation of field quantities is essentially achieved by the 'smearing out' of every particle over a number of grid cells. This is conceptually equivalent to filtering the sample of particles — i.e. convolving the spatial distribution with a certain kernel — and subsequently sampling the result on a set of grid points. The process is characterised by the shape and size of the kernel, and balances are struck between noise levels and the reproduction of spatial detail.

On a continuous, one-dimensional manifold x, the convolution of a function $f(x)$ with the kernel $K(x)$ is given by:

$$\hat{f}(x) = (f * K)(x) = \int f(x')K(x - x')dx'. \tag{105}$$

[38]Field estimation in any given partition of the volume will generally depend to a greater or lesser extent on the particle quantities sampled just outside the partition's boundaries. As a result, field estimation will be executed twice in the neighbourhoods of partition boundaries. The sampling density determines the amount of space needed to perform interpolations on — high sampling density means interpolation over small distances and vica versa.

A discrete sample g (a set of N points with coordinates x_i and values g_i) can be convolved similarly:

$$\hat{g}(x) = \sum_{i=1}^{N} g_i K(x - x_i) \tag{106}$$

Convolution can be done by multiplying the Fourier transform of the sample by that of the convolution kernel. If $f(x)$ and $g(x)$ Fourier transform to $F\{f\}(k)$ and $F\{g\}(k)$, their convolution satisfies the *convolution theorem*:

$$F\{f * g\}(k) = F\{f\}(k) \cdot F\{g\}(k) \tag{107}$$

Furthermore, since determining the spatial derivative of a field can be done in Fourier space, filtering in Fourier space can be carried out at the same time. These considerations make it preferable that a convolution kernel has a well-behaved Fourier transform. In practice, the most widely used kernels are the top-hat and Gaussian kernels.

- Top-hat kernel: For filter scale R_{TH}, the kernel is given by:

$$W_{TH}(\mathbf{x} - \mathbf{x}') = \begin{cases} \frac{1}{\frac{4\pi}{3} R_{TH}^3} & |\mathbf{x} - \mathbf{x}'| \leq R_{TH} \\ 0 & \text{otherwise} \end{cases} \tag{108}$$

  An advantage of this kernel is that it is conceptually simple and has a computationally low demand. More importantly, it is localised: a sample value has no influence at all outside a radius of R_{TH}. A disadvantage of the top-hat kernel is that, due to its sharp edges, its Fourier transform produces many fringes, or *shot noise*.

- Gaussian kernel: For filter scale R_G, the kernel is given by:

$$W_G(\mathbf{x} - \mathbf{x}') = \frac{1}{(2\pi R_G^2)^{3/2})} e^{(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{2R_G^2})} \tag{109}$$

  Under this kernel, each sampling point will have an influence on the rest of the domain that asymptotically approaches zero, but extends outward infinitely. Therefore, an accurate execution of this convolution requires more computation. The Fourier transform of the Gaussian kernel, though, is much better behaved than that of the top-hat kernel:

$$F\{W\}(\mathbf{k}) = e^{-\frac{k^2 R_G}{2}},$$

  another Gaussian.

Note that the 'identity operand' for convolution is the Dirac delta function $\delta_D(x)$.

$$(f * g)(x) = f(x) \Rightarrow g(x) = \delta_D(x)$$

For the purpose of cosmological simulation bodies, it is a discrete sample of particles in a three-dimensional space that is convolved. In essence, grid-based field estimators do this by erecting a kernel function of choice, centred on each particle position, and sampling the summation of all these kernels on a set of grid points[39]. This requires a kernel that is defined in three dimensions. In the process of field estimation, the kernel is represented by the symbol W, as it signifies the weight by which a physical quantity is distributed over a collection of spatial points.

Adaptivity

In regions of different sampling density, the accuracy and spatial detail have a different dependency on the kernel's shape and size. Therefore it is beneficial — though computationally more costly — to let the kernel shape and size vary according to the sampling density. An *adaptive kernel size* can be implemented in a grid-based field estimator, by varying the scale length (e.g. R_{TH} or R_G) inversely proportional to the sampling density. For example, the scale length can be chosen such that a sphere of that radius around a sample point contains a fixed number of other sampling points in its interior. High density regions will then be resolved better, while noise is kept relatively low in low density regions.

An *adaptive kernel shape* is beneficial in regions where the sampling density is distributed anisotropically in space — e.g. in filaments and walls. In order to reproduce the flattened and extended nature of such anisotropic features, a tessellation field estimator is best suited.

4.2 Grid based field estimators

A grid-based field estimator superimposes a regular grid on the simulation volume, and considers the grid cell a particle is in, and in some cases its precise within that grid cell. It then divides the field value corresponding to that particle among any number of neighbouring grid points, depending on the particle's position. This procedure can follow various schemes. In these schemes, the adopted division of field values over the grid points is usually fixed in shape and size. The weights are usually symmetric in orientation, i.e. solely dependent on the absolute distance between the particle and grid point in consideration.

In the general approach to a grid based estimator, the particle position is $\mathbf{x}_i$, and all its components which lie between 0 and the box size L. A grid point is specified by its index vector $\mathbf{n}$. Now, the spatial separation between particle i and grid point $\mathbf{n}$ becomes:

$$\mathbf{d}_{i,\mathbf{n}} = \mathbf{x}_i - \frac{\mathbf{n}}{N_G} L, \tag{110}$$

where N_G is the number of grid cells along one vertex of the box — it is thus inversely proportional to the grid resolution. Then, the general formula for any grid-based estimation of a physical quantity f is the following:

$$\hat{f}\left(\frac{\mathbf{n}}{N_G} L\right) = \frac{\sum_{i=1}^{N} f_i W(\mathbf{d}_{i,\mathbf{n}})}{\sum_{i=1}^{N} W(\mathbf{d}_{i,\mathbf{n}})}, \tag{111}$$

where $\hat{f}$ is the estimated quantity, N is the number of particles in the simulation volume, and f_i is the quantity corresponding to particle i. It is always possible to set the box size to unity, $L = 1$, and rescale all particle positions such that

[39]Subsequently, the resulting grid of field estimations can optionally be smoothed over by an FFT on the grid, and a suitable smoothing kernel.

$\mathbf{x}_i$ lies between 0 and 1. Furthermore, under the restriction that

$$\sum_{i=1}^{N} W(\mathbf{d}_{i,\mathbf{n}}) = N$$

for a normalised estimation scheme, the expression for the estimated field value simplifies to

$$\hat{f}\left(\frac{\mathbf{n}}{N_G}\right) = \frac{1}{N}\sum_{i=1}^{N} f_i W(\mathbf{d}_{i,\mathbf{n}}). \tag{112}$$

This way, any scalar quantity can be estimated. Vector quantities like velocity $\mathbf{v}$ can be obtained by combining a field estimation for each spatial component separately:

$$\hat{\mathbf{v}} = \hat{v_x}\mathbf{e}_x + \hat{v_y}\mathbf{e}_y + \hat{v_z}\mathbf{e}_z, \tag{113}$$

where $\mathbf{e}_j$ is the unit vector in the j-direction.

Density estimation

A density can be estimated by determining a grid cell mass using equation 112, and dividing it by the cell volume,

$$V_{cell} = \left(\frac{L}{N_G}\right)^3,$$

so that the estimated density $\hat{\rho}$ is given by

$$\hat{\rho}\left(\frac{\mathbf{n}}{N_G}\right) = \frac{1}{N}\left(\frac{N_G}{L}\right)^3 \sum_{i=1}^{N} m_i W(\mathbf{d}_{i,\mathbf{n}}),$$

or if we set the box volume to unity:

$$\hat{\rho}\left(\frac{\mathbf{n}}{N_G}\right) = \frac{N_G^3}{N}\sum_{i=1}^{N} m_i W(\mathbf{d}_{i,\mathbf{n}}), \tag{114}$$

where m_i is the mass of particle i.

Choice of estimation weights

What sets apart one grid-based field estimator from another is its choice of the weights $W(\mathbf{d}_{i,\mathbf{n}})$.

In a one-dimensional space, a single particle's quantity can be distributed over any number of cells along that dimension. To set this division, a field estimator can assign any number of weights $w(d_{i,n})$. In M dimensions, the total weight of a particle to any grid cell can be taken to be any combination of its weights in each dimension separately. One straightforward choice is to consider a linear product of weights w_j along each dimension j:

$$W(\mathbf{d}) = \prod_{j=1}^{M} w(d_j), \tag{115}$$

and this construction of weights is adopted by three widely used field estimators: (i) Nearest Grid Point, (ii) Cloud in Cell and (iii) Triangular Shaped Cloud. All three of them distribute a particle's field quantity in a spatially symmetric way. This is equivalent to a convolution with a spatially symmetric kernel, with a fixed size depending on the grid resolution.

- The Nearest Grid Point (NGP) method is conceptually the simplest of the three. It solely assigns all of a particle's mass to the grid point nearest to it, regardless of the particle's precise position within its cell. Effectively, the scheme uses a Dirac δ-shaped cloud. For the j-eth component of the distance $\mathbf{d}_{i,\mathbf{n}}$ between particle i and grid point $\mathbf{n}$, the weight function is given by:

$$w_{NGP}(d_j) = \begin{cases} 1, & N_G|d_j| \leq \frac{1}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{116}$$

- The Cloud in Cell (CIC) algorithm (Birdsall & Fuss, 1969) improves upon this by considering the spatial separation between a particle and its surrounding grid points. It assigns a density to these grid points as a function of that separation. The CIC method divides the weight over a particle's two closest neighbouring grid cels in each spatial direction. The number of updated grid points is then 2^M. The CIC scheme is equivalent to convolution with a top-hat kernel; the weight function is:

$$w_{CIC}(d_j) = \begin{cases} 1 - N_G|d_j|, & N_G|d_j| \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{117}$$

- The Triangular Shaped Cloud (TSC) algorithm can be seen as a refinement on the CIC method. Instead of dividing a particle mass over the 2^M closest grid points, the TSC algorithm assigns a portion of the mass to the nearest grid cell, and to both its direct neighbours in each spatial direction. This means that a small (hyper)cube of 3^M grid cells can be assigned a density. This method corresponds to a triangular kernel: the weight assigned to the closest cell is greater than that assigned to its neighbours in any direction. The one-dimensional weight function is:

$$w_{TSC}(\delta d_j) = \begin{cases} \frac{3}{4} - N_G^2 d_j^2, & N_G|d_j| \leq \frac{1}{2}, \\ \frac{1}{2}\left(\frac{3}{2} - N_G|d_j|\right)^2, & \frac{1}{2} \leq N_G|d_j| \leq \frac{3}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{118}$$

Limitations

The methods described above have a relatively low computational demand, and preserve spatial resolution well. The great limitation of these algorithms is that they perform poorly on undersampled regions — i.e. few particles and many grid cells. Cells that do not contain or neighbour any particles will simply be assigned a zero density or velocity, regardless of the populations in nearby regions. This leads to relatively high noise levels. The NGP method is most extreme in all of these considerations — it has the lowest computational cost, preserves spatial resolution as well as the grid resolution allows for, and is the most sensitive to the local sample density. The CIC and TSC methods progressively trade more computational demand and spatial detail for a higher estimation quality, based on a more detailed use of the information on a particle's position.

As a result of this sensitivity to local sample density, the grid representations of estimated field quantities can be subject to various issues. The edges in spatial features can become jagged, and artefacts can show up both in scalar fields and in their gradients.

For the fixed-shape methods holds that the value of an estimated quantity at a given position depends strongly on the chosen grid resolution. Also, there is an artificial dependence on the orientation of the grid relative to any possible anisotropic feature — e.g. filaments or walls — in the simulated structure. These limitations are a direct consequence of the fact that the chosen grids and kernels are non-adaptive.

Smooth Particle Hydrodynamics

Smooth Particle Hydrodynamics (SPH) schemes make use of available information about a particle's environment. They consider an M-dimensional weight $W(\mathbf{d}_{i,\mathbf{n}}, h_i)$ that adapts in size to the local sampling density. The shape of the kernel, just like the aforementioned non-adaptive schemes, remains symmetric in all directions — the anisotropic nature of the grid notwithstanding. However, the scaling length h_i can be chosen for each particle i separately. One example is to require that a sphere of radius h_i around the particle always contains a fixed number of neighbouring particles. This way, estimations in high density regions will have a greater resolving detail, while estimations in low density regions will be less subjected to noise.

Although these are significant advantages, certain artefacts remain, due to the fixed resolution, shape and orientation of the SPH kernel and the grid. Furthermore, determining a precise scale length h_i for a particle according to the density of its immediate environment introduces another arbitrary choice.

4.3 Weighting

We have seen that grid-based field estimators essentially count particles in predefined grid cells, and we will soon witness tessellation-based algorithms defining a space-filling network of cells according to the spatial distribution of particles. There is a fundamental difference between the outcomes of these strategies, and this difference is that between *mass-weighted* and *volume-weighted* averaging.

When estimating a physical quantity on a regular grid, the weighted sum of particle quantities is divided by the sum of weights — or, equivalently, the convoluted sample is divided by the summed kernel.

$$\hat{f}(\mathbf{x_0}) = \frac{\sum_{i=1}^{N} f_i W(\mathbf{x_i}, \mathbf{x_0})}{\sum_{i=1}^{N} W(\mathbf{x_i}, \mathbf{x_0})}. \tag{119}$$

It is important to realise that in so doing, one is in fact weighting the quantity f by mass, as equation 119 is a discretisation of

$$\hat{f}(\mathbf{x_0}) = \frac{\int d\mathbf{x} f(\mathbf{x})\rho(\mathbf{x}) W(\mathbf{x_i}, \mathbf{x_0})}{\int d\mathbf{x} \rho(\mathbf{x}) W(\mathbf{x_i}, \mathbf{x_0})}. \tag{120}$$

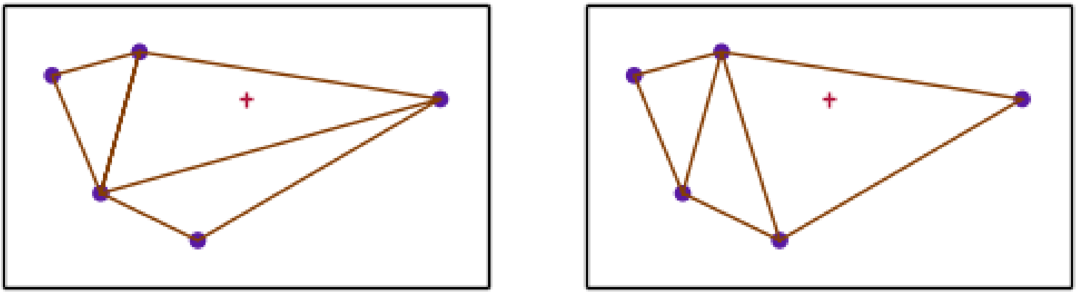


Figure 20: Two different ways to *triangulate* the same set of sampling points in two dimensions. In the two cases interpolation on the point marked with the red + will be based on two distinct sets of sample points. Note that the triangulation in the right panel will result in smaller errors, as it consists of the most *compact* triangles. The right panel satisfies the *Delaunay condition*, while the one on the left does not.

The analytically determined statistical properties of the cosmological velocity field, however, are only rarely weighted by mass. In nearly all cases, they are weighted by volume.

$$\hat{f}(\mathbf{x_0}) = \frac{\int d\mathbf{x} f(\mathbf{x}) W(\mathbf{x_i}, \mathbf{x_0})}{\int d\mathbf{x} W(\mathbf{x_i}, \mathbf{x_0})}. \tag{121}$$

Mass-weighted quantities cannot simply be compared to volume-weighted ones, and this makes it preferable to produce a volume-weighted estimation.

In one-dimensional space, the interpolation of quantities from a given set of sample points is done by one-dimensional bins. At any coordinate — short of the particle positions themselves — it is clearly defined which samples to interpolate between — namely the closest particles to the left and right of that coordinate. There is no choice to be made in the *ordering* of particles into interpolation bins.

In a higher dimensionality, though, higher-dimensional interpolation bins can be defined in various ways. This is illustrated in figure 20. At any point in space, quantities can be interpolated between various sets of neighbouring samples, and there is the choice which particles are to be used for interpolation. In other words, no unique ordering of interpolation bins exists between sampling points in higher dimensionalities.

Some choices will be more sensible than others. The Delaunay tessellation — introduced momentarily — is chosen such that its cells — triangular in two dimensions and tetrahedral in three — are as compact as possible. This way, interpolation along the edges between particles is subject to a minimal amount of uncertainty.

In contrast to an overlayed regular grid, the positions and orientations of Voronoi and Delaunay cells are highly sensitive to the spatial distribution of the sample points. It is in this capacity that these two species of tessellations are employable in producing volume-weighted field estimations.

4.4 Tessellations

A tessellation is the division of a space of any dimensionality into cells of positive definite size and equal dimensionality. In a tessellation no two cells ever overlap, and all cells taken together fill the entire space[40]. A tessellation can be defined on various geometries, and for the purposes of cosmological N-body simulations, a (hyper)toroidal geometry is usually adopted. This means that a trajectory in any direction, upon crossing the edge of the defined volume, will emerge at the same position on the opposite edge, and continue in the same direction. This is referred to as *periodic boundary conditions*, and it is a convenient way to define a finite volume without edges. Thus, it is possible to fill the volume with a finite number of tessellation cells.

Two distinct types of tessellation — the *Voronoi* and *Delaunay tessellations* — are defined on the basis of a spatial distribution of sampling points. These points, finite in number, may follow any spatial distribution.

Voronoi tessellation

Consider a finite, periodic volume of dimensionality M, containing a set p of N sampling points. In a Voronoi tessellation, one cell is erected around each of those points. A cell around a sampling point p_i is defined — i.e. the positions of the cell's vertices are determined — such that every point within that cell is closer to p_i than to any other sampling point, using a Euclidean metric.

A region with a high sampling density will have many small Voronoi cells, and a region with a low sampling density will have few large ones. Voronoi cells are always convex, and the number of vertices spanning it is always greater than or equal to $M+1$, but has no upper limit. Figure 21(a) shows the Voronoi tessellation based on a sample of 40 points, following a sampling density field consisting of a Gaussian peak superimposed on a uniform background.

Delaunay tessellation

Based on the same set p, the Delaunay tessellation is the division of space into hypertriangles, in such a way that no point p_i ever falls within the circumscribing hypersphere of any hypertriangle. On a two-dimensional plane, each Delaunay cell is a triangle, whose circumcircle has zero sampling points in its interior. In three dimensions, each Delaunay cell is a tetrahedron, whose circumsphere, likewise, contains no sampling points. For any given set of sampling points p, there may be many ways to divide the space up into hypertriangles — see figure 20 for two examples in two dimensions. As a result of the Delaunay condition, the Delaunay tessellation maximises the minimum angle at all triangle vertices. Figure 21(b) shows the Delaunay tessellation based on the same sample of points used in figure 21(a).

For the sample set p, the Delaunay tessellation is the *dual* of the Voronoi tessellation based on that same set. A Voronoi tessellation can be constructed by connecting the centres of the circumscribing hyperspheres of all Delaunay hypertriangles. By virtue of their duality, the edges in one of the two tessellations can be determined by connecting the centroids of all pairs of cells from the other tessellation if they share a face.

[40]Occasionally, one comes across a tessellation of extraordinary beauty — an experience that the cover of this thesis attempts to capture.

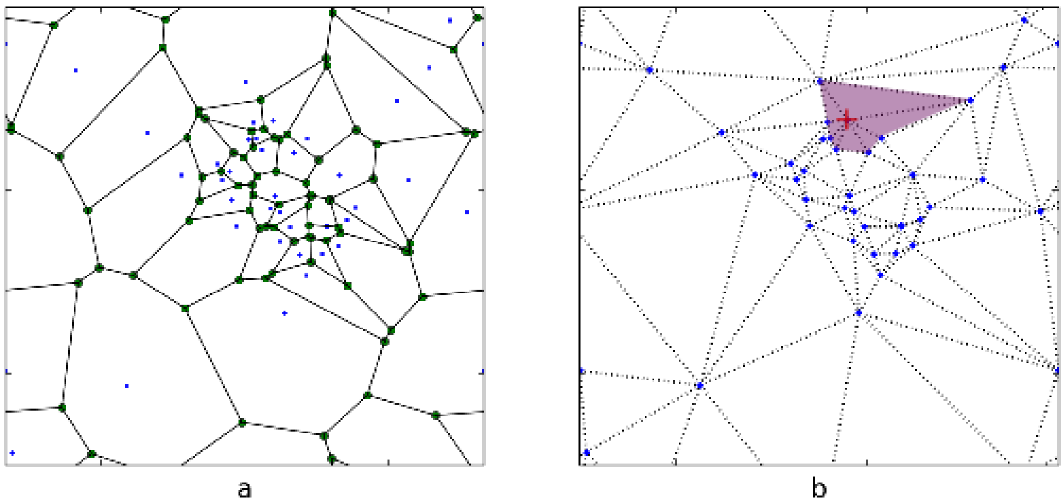


Figure 21: Two-dimensional Voronoi (panel a) and Delaunay (panel b) tessellations of the same distribution of 40 sampling points. The distribution consists of a Gaussian peak superimposed on a uniform background. Both tessellations assume periodic boundary conditions. The density peak shows the behaviour of the tiles according to the local sampling density. Notice the duality between the two tessellations. The shaded area marks the union of Delaunay cells sharing the sample point marked with the red +, it is that point's *contiguous Voronoi cell*, used to estimate the local density.

Degeneracy

In a two-dimensional distribution of sampling points, it may occur that four neighbouring points happen to lie on the same circle - with no sampling points between them. In this case, they form a *cyclic quadrilateral*[41], and as a consequence, the circle circumscribing each of the four triangles that could possibly be defined by three of these points will also pass through the fourth. This means that the triangulation can be done in two ways — either one of the two possible diagonals can be taken to be an edge. There is no preferred way, and this situation is called a *degeneracy.* The points in a square grid form a readily appreciable example of this.

Degeneracies also occur when five or more points lie on the same — otherwise empty — circle. In M-dimensional space, $M + 1$ sample points are required to define one hypertetrahedron. Thus, any empty hypersphere that passes through more than $M + 1$ will be subject to degeneracy.

4.5 Tessellation field estimators

In a space of dimension $M \geq 2$, a regular grid is only one of boundless possible ways of dividing the sample volume up into cells. There is no end to the configurations M-dimensional 'binning' regions can be defined in. As an alternative to a regular grid, a very natural and logical choice would be to define one region for each simulation particle, with a size that corresponds to the sampling density in its environment. This is done by computing the Voronoi tessellation generated by the particles[42] (Icke & van de Weygaert, 1987; van de Weygaert, 1991, 1994; Schaap & van de Weygaert, 2000; van de Weygaert & Schaap, 2009). This description of tessellation field estimators follows broadly Bernardeau & van de Weygaert (1996) and van de Weygaert & Schaap (2009).

The Voronoi estimator

The Voronoi method can be seen as a very natural extension into M- dimensionality of the approximation of a one-dimensional field as a piecewise constant function in a set of bins. The one-dimensional analogue is illustrated in figure 22(a), the upper panel in figure 23 shows an example of Voronoi interpolation on a two-dimensional body of sample points. In each Voronoi cell, the estimated scalar field quantity simply equals that of the generating particle. Only at the boundaries separating the Voronoi cells, there is a non-zero gradient. The total area of cell boundaries within a certain region of space then forms a measure of the volume-averaged quantity. By filtering the field estimation with a radius that is sufficiently large, this volume-averaged quantity is obtained. Evidently, a kernel that does not exceed the confines of a single Voronoi cell will not measure any gradient in the field value. For a volume spanning a length L in each dimension, the characteristic length scale at which this effect starts occurring is of the order $L/N^{1/M}$.

Since the Voronoi cell boundaries are the locus of all points with a non-zero gradient, the gradient of a scalar field — like one individual velocity component — at point $\mathbf{r}$ can be determined from a Voronoi tessellation directly. We assume

[41] Cyclic quadrilaterals are generally asymmetric, but special cases exist: Squares, rectangles, isosceles trapezia, antiparallellograms and some kites.

[42] Note that it is not necessary to use all of the particle positions to create a space filling tessellation. However, using all particles as generating points will result in a greater accuracy.

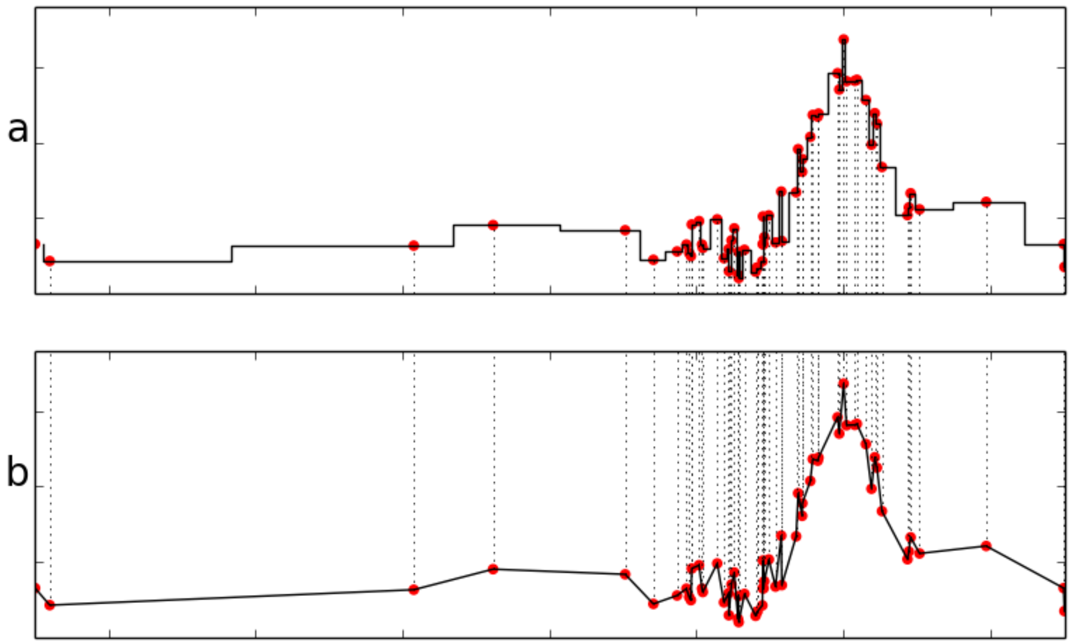


Figure 22: Piecewise constant (panel a) and piecewise linear (panel b) approximations of a one-dimensional field as sampled by a set of 62 randomly located points. These approximations are one-dimensional equivalents of the Voronoi tessellation (panel a) and Delaunay tessellation (panel b) as generated by the circular markers. The behaviour of the tessellations across a range of sampling densities is visible. The generating points generally do not form the geometrical central points of their corresponding Voronoi cells. Note that degeneracies in the Delaunay tessellation never occur in one dimension.

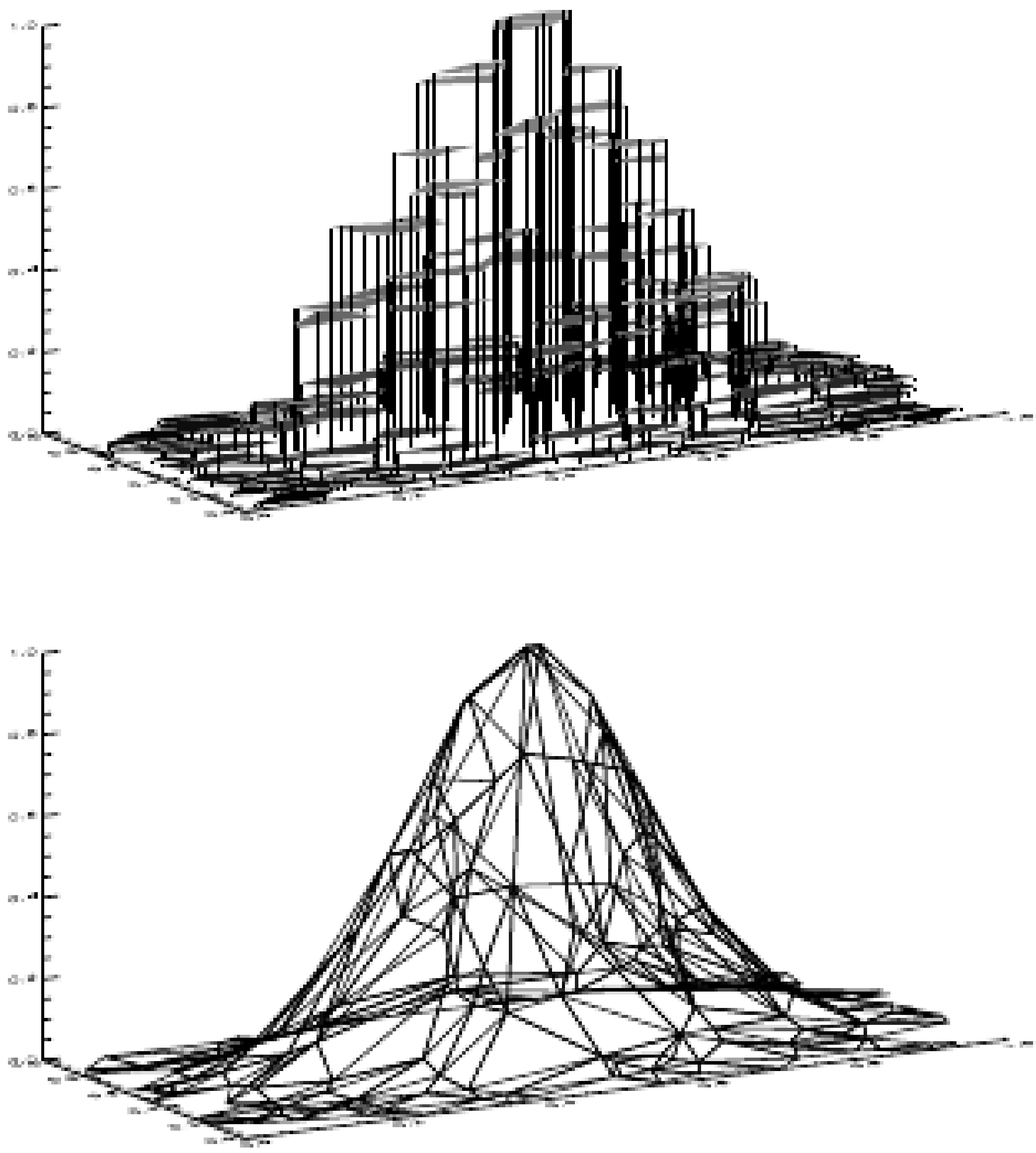


Figure 23: Interpolation on a two-dimensional body of sample points. The upper panel shows an interpolation using Voronoi cells, in which each cell is given a constant field value — a *piecewise constant* interpolation. The lower panel uses the Delaunay triangulation of the sample points, allowing for a *piecewise linear* interpolation. Source: (Schaap & van de Weygaert, 2001).

each boundary k, separating cells k_1 and k_2, to be of an infinitesimal thickness. A volume-averaged velocity gradient can be estimated at point $\mathbf{r}$ by integrating over all the — partial — cell boundaries within a radius R from $\mathbf{r}$.

The Delaunay estimator

To improve upon the approximation of a field value as piecewise constant in a set of Voronoi cells, we might perform a linear interpolation between the generating sample points. To construct a *piecewise linear* approximation — see figure 22(b) for a one-dimensional example and the lower panel in figure 23 for a two-dimensional one — we appeal to the Delaunay tessellation (Delone, 1934). Since each edge runs linearly from one particle to one of its closest neighbours, a constant gradient can be defined along it. The same analogy applies here: the Delaunay method can be seen as a very natural extension into M-dimensionality of the first-order interpolation of a field between a finite number of sample points. This is illustrated in figure 22(b).

For a Delaunay simplex — see figure 24 for an example of a Delaunay tetrahedron in three dimensions — each point in its interior is a linear combination of the cell's vertices. Therefore, the field value at any desired point can be estimated as the same linear combination of the field values sampled at each of the vertices. Let a point $\mathbf{r}$ fall within a certain Delaunay simplex with $M+1$ vertices $\mathbf{r}_k$. Its representation as a linear combination of the vertex locations looks like:

$$\mathbf{r} = \sum_{k=1}^{M+1} \alpha_k \mathbf{r}_k. \tag{122}$$

In a linear interpolation, the field estimation at $\mathbf{r}$ then uses the same coefficients α_k:

$$\hat{f}(\mathbf{r}) = \sum_{k=1}^{M+1} \alpha_k f(\mathbf{r}_k). \tag{123}$$

As with the Voronoi method, the spatially adaptive nature of the tessellation allows us to estimate a field gradient directly from the sample values at the simplex vertices. In three dimensions, a Delaunay tetrahedron's shape — i.e. the locations of r_1, r_2 and r_3 relatively to r_0 — can be characterised by a matrix $\mathbf{A}$, defined as:

$$\mathbf{A} = \begin{pmatrix} \Delta x_1 & \Delta y_1 & \Delta z_1 \\ \Delta x_2 & \Delta y_2 & \Delta z_2 \\ \Delta x_3 & \Delta y_3 & \Delta z_3 \end{pmatrix}, \tag{124}$$

where $\Delta x_k \equiv x_k - x_0$, and similarly for Δy_k and Δz_k. In the same way, along each vertex, the gradient in the l-velocity is the constant $\Delta v_{l,k} = v_l(\mathbf{r}_k) - v_l(\mathbf{r}_0)$. From here, the gradient components $\partial_j v_l$ of the velocity field quantity v_l throughout the tetrahedron can be determined by:

$$\begin{pmatrix} \partial_x v_l \\ \partial_y v_l \\ \partial_z v_l \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} \Delta v_{l,1} \\ \Delta v_{l,2} \\ \Delta v_{l,3} \end{pmatrix}. \tag{125}$$

In order to determine a volume-averaged velocity gradient at point $\mathbf{r}$, the piecewise constant environment around $\mathbf{r}$ is filtered, again with a kernel of sufficient size. When a top-hat kernel is used, this involves finding the intersection

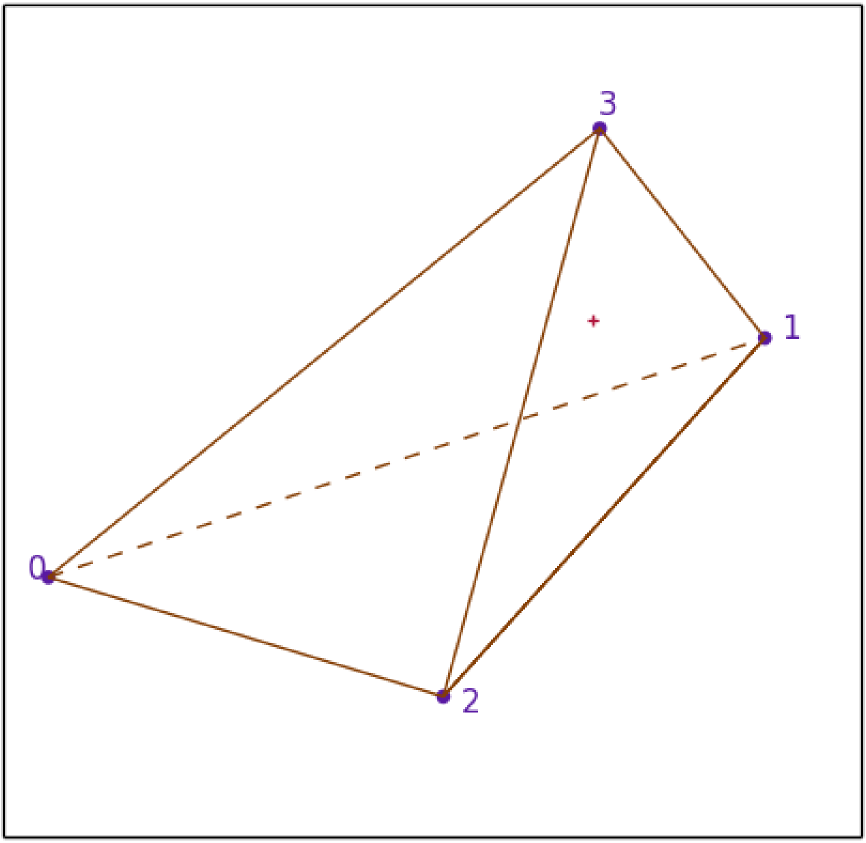


Figure 24: Delaunay tetrahedron in a three-dimensional scenario. In three dimensions, every Delaunay simplex is defined by four points, {0-3}, and every point in its interior can be described as a linear combination of the four points' coordinates. This is how linear interpolations are made at an arbitrary point inside the tetrahedron. See the text leading up to equation 125.

between the spherical volume and all tetrahedra it encompasses and passes through. For any tetrahedron k, let $V_k(\mathbf{r}, R)$ be its total volume of intersection with the sphere of radius R around $\mathbf{r}$. The velocity gradient is then estimated by summing over all tetrahedra k with non-zero intersection:

$$\hat{\partial_j v_l}(\mathbf{r}, R) = \frac{3}{4\pi R^3} \sum_k (\partial_j v_l)_k V_k(\mathbf{r}, R). \qquad (126)$$

Density estimates can be made from the Delaunay tessellation, by using the fact that the local matter density and the size of Delaunay simplices are inversely correlated — provided all particles are equal in mass. The density at a sampling point is determined by the size of all the cells it is a part of (van de Weygaert & Schaap, 2009). The union of all Delaunay cells sharing a point $\mathbf{x}_i$ is known as that point's *contiguous Voronoi cell*, and designated by the symbol $\mathbf{W}_i$ — see the shaded area in figure 21(b) for a two-dimensional example. In M dimensions, the density estimate at $\mathbf{x}_i$ is proportional to the inverse volume of the contiguous cell $V(\mathbf{W}_i)$:

$$\hat{\rho}(\mathbf{x}_i) = (1 + M)\frac{w_i}{V(\mathbf{W}_i)}, \qquad (127)$$

where w_i is the weight of sample point i. The density estimate must be normalised and weighted by the particle masses, if all particles are equal in mass, w_i will be equal as well. Note that $1 + M$ is the number of points making up each Delaunay simplex in M dimensions.

The superiority of the DTFE method — particularly in resolving small scale structures at high sampling densities — can be clearly seem from a comparison with the TSC technique applied to the same data set. See figure 25.

Note that this method can be used in any tessellation that defines hypertriangles with the particle positions as their vertices. However, in order to optimise the accuracy of the interpolation, it is of crucial importance to minimise the distances between the estimation points and the cell vertices. A Delaunay tessellation yields these optimal results, as it produces the most compact simplices.

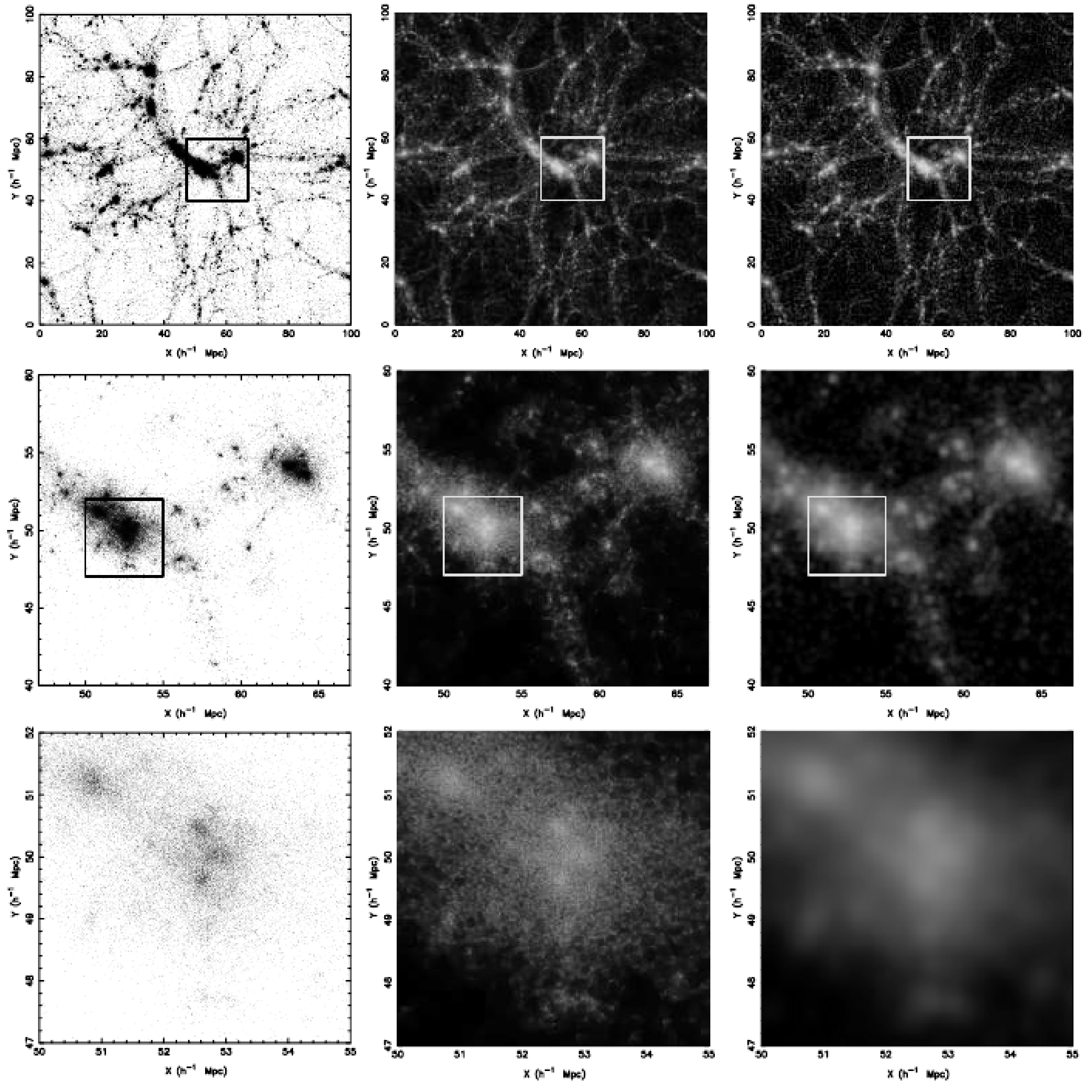


Figure 25: Comparison between the DTFE and TSC techniques. The left column shows the distribution of separate particles from a simulation, at three different scales. The central column shows the corresponding DTFE density estimations, and the right column shows the TSC estimated densities. Source: (Schaap & van de Weygaert, 2002)

z	a	z	a
14	0.07	0.47	0.68
3.7	0.22	0.30	0.77
2.1	0.32	0.16	0.86
1.4	0.42	0.04	0.96
0.97	0.51	0	1
0.68	0.60		

Table 2: Redshifts and corresponding scale parameters for all the Cosmogrid snapshots used in the analysis.

res	cellw	cellpop
512	0.059	1
256	0.117	8
128	0.234	64
64	0.469	512

Table 3: Intrinsic smoothing scales associated with various grid resolutions on a box sized 30 Mpc in each direction, containing 512^3 particles. The column labelled 'res' displays the number of grid cells along each edge of the box. The column labelled 'cellw' displays the width of each sell in Mpc. The column labelled 'cellpop' displays the average number of particles per cell.

5 Methods and definitions

5.1 Simulation data

The analysis in this project concerns eleven snapshots from the Cosmogrid simulation — see section 3.3. These are spaced evenly in the scale parameter a, at the redshifts listed in table 2. Each snapshot contains the positions and velocities of 512^3 dark matter particles — all equal in mass. For the lower resolution analyses, downsamples were made of 256^3 particles and fewer, by a uniform random selection from the original list of particles.

The velocities stored in the `gadget` files[43] are internal velocities. In order to convert those to comoving velocities — which are needed for the analysis — they are multiplied by the square root of the scale factor, $\sqrt{a}$, corresponding to the epoch of each snapshot.

To motivate this, we insert $H(t) \propto a^{-3/2}$ and $f(\Omega) = 1$ into equation 37. These specifications correspond to the the Einstein-de Sitter Universe, which is in a close agreement with the predominant epochs of structure formation. This results in the proportionality

$$v(t) \propto a^{1/2}. \tag{128}$$

5.2 Field estimation

From the simulation data, we determined the density and velocity values on regular grid points, once using the TSC estimator and once with DTFE — see

[43]Gadget (Springel et al., 2001) is a file format used to store particle data for cosmological N-body simulations. See section 3.2 for details.

sections 4.2 and 4.5). Here, the grid resolution determines the intrinsic smoothing of the resulting density and velocity fields. For 512^3 particles in a 30 Mpc box, these scales are listed in table 3.

The TSC estimation was made by an algorithm that stacks the TSC weights of particles upon a grid that is initially zero everywhere. The density is determined in correspondence with equation 114. Each velocity component separately is determined conform equation 112.

The DTFE estimation was done by the *DTFE-1.2.0* programme — developed by Marius Cautun — which divides the computational load of the procedure described in section 4.5 over various available processors. By virtue of the periodicity of the Cosmogrid Universe, the Delaunay triangulation was constructed under periodic boundary conditions without loss of accuracy at the edges of the box.

The DTFE programme determined the volume-averaged densities and velocities — while the TSC estimator, by nature, yields mass-weighted quantities. The internal workings of the DTFE algorithm allow for a direct calculation of each of the velocity gradient components. The TSC estimator returns velocities, but the velocity derivatives were calculated separately.

5.3 Velocity differentiation

In the case of TSC-estimation, the velocity gradient was calculated subsequently. Departing from a regular grid of samples for each velocity component, determining the nine gradient components can be done in various ways.

Real space derivatives

One straightforward method is to calculate the difference between the field values at neighbouring grid points, and combining these differences into an approximate derivative quantity at each position. This scheme is described in Råde & Westergren (2004), and produces a first-order approximation of the velocity derivative. The velocity derivative $\partial_j v_i$ in cell j_n will be

$$\partial_j v_i|_n \simeq \frac{v_i(j_{n+1}) - v_i(j_{n-1})}{2d}, \tag{129}$$

where d is the cell width. If the volume has periodic boundary conditions, this property can be exploited to approximate the derivatives at the edges — e.g. for a 128 cell grid,

$$\partial_j v_i|_{127} \simeq \frac{v_i(j_0) - v_i(j_{126})}{2d}.$$

Otherwise an extrapolation from the interior grid points can provide a solution:

$$\partial_j v_i|_{127} \simeq \frac{v_i(j_{125}) - v_i(j_{126})}{2d}. \tag{130}$$

The fact that this method can be used in non-periodic volumes is one of its strengths. However, this first-order approximation does not make use of the data on any grid points further than one unit away from the point of interest. The derivatives determined in this way are also particularly sensitive to noise in the original samples.

Fourier space derivatives

An alternative method makes use of the Fourier transform of the derivative operator:

$$\hat{f}'(k) = -ik\hat{f}(k) \tag{131}$$

Once the discrete Fourier transform of the sample velocities is determined, differentiation with respect to the j-direction is done by a simple multiplication by $-ik_j$, and subsequent inverse transformation back to real space. This makes the derivative of the l-component of velocity with respect to spatial dimension j:

$$\partial_j v_l(\mathbf{x}) = -i \int \frac{d\mathbf{k}}{(2\pi)^3} k_j \cdot \hat{v}_l(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}}, \tag{132}$$

where[44]

$$\hat{v}_l(\mathbf{k}) = \int d\mathbf{x} v_l(\mathbf{x}) e^{i\mathbf{x}\cdot\mathbf{k}}. \tag{133}$$

As the discrete Fourier transformation of the whole grid is evaluated, periodicity is a requirement. Aperiodic boundary conditions will result in artefacts rippling inward from the edges — although this can be prevented by *tapering* the data.

Contrary to the aforementioned real space differentiation scheme — where the derivatives are only dependent on the field values in immediately neighbouring cells — the Fourier method assigns derivatives to grid points within the context of the whole volume: the derivative at any location generally depends on the field values at all other locations. While this introduces the risk of certain Fourier artefacts — see section A.3 — it is mostly considered an advantage. Differentiation via Fourier space uses all available data in a field and produces more accurate derivatives. On top of that, the computational cost is limited.

When carried out numerically, it is possible that the result is not purely real — some small imaginary residuals may be left after the inverse transformation of the field values, despite the original velocity field being purely real. This can be prevented by setting the condition that the complex conjugate $\hat{v}_l^*(\mathbf{k}) = \hat{v}_l(-\mathbf{k})$. This may be necessary in cases of noisy data.

All TSC estimated velocities are differentiated using this Fourier method — the real space method provides a reality check for the Fourier results.

5.4 Web classification

A core element in this study is the decomposition of the cosmic volume into the different components that make up the cosmic web. Identifying and separating cosmic web components is known as *web classification*, and section 2.8 provides a background into this. For our study, we have formulated a unique web classification method. Previously developed classifiers are based solely upon the geometry of a tracer tensor — the individual eigenvalues. Ours makes an additional distinction on the basis of the local overall expansion or contraction — determined by the sum of eigenvalues. Thus, we identify six different cosmic web components, they are listed in table 4. The rest of this section provides a

[44] Equations 132 and 133 are written under the *Kaiser convention*, in which the latter is the standard formulation of the Fourier transformation. The former is the inverse transformation of $-ik_j\hat{v}_l(\mathbf{k})$. Note that i is the imaginary unit, and l is the velocity component index.

ID	λ_1	λ_2	λ_3	$\sum_i \lambda_i$	feature:
0	-	-	-	-	Node
1	+	-	-	-	Filament
3	+	-	-	+	Prolate collapsing void
2	+	+	-	-	Wall
4	+	+	-	+	Oblate collapsing void
5	+	+	+	+	Expanding void

Table 4: Classification of spatial regions according to their eigenvalues — individually as well as summed. The + symbol indicates positive definite values, while the − symbol indicates negative ones. Note that the presupposed condition that $\lambda_1 > \lambda_2 > \lambda_3$ excludes several combinations of signs. Note, also, that the cases labelled either 1 or 3 and those labelled 2 or 4 share the same eigenvalue signature, but are distinguished solely by the sign of the summed eigenvalues. The six types of regions identify the components of the cosmic web. They are mutually exclusive and span all of space: No single element of space falls within zero or more than one of these categories. See text for details.

detailed explanation of our method.

Deformation tensor eigenvalues

Following in the footsteps of Pogosyan et al. (1998); Hahn et al. (2007), we identified different regions of space as specific components of the cosmic web, by means of investigating the eigenvalues of the deformation tensor. The deformation tensor D_{ij} is defined in terms of the shear and divergence:

$$D_{ij} = \sigma_{ij} + \frac{1}{3}(\nabla \cdot \mathbf{v})\delta_{ij}. \tag{134}$$

Like the shear, this tensor is symmetric. It is not traceless, however, and thus it has three linearly independent eigenvalues.

Symmetric matrices — and other Hermitian matrices — are guaranteed to be diagonalisable, and have a number of independent real eigenvalues equal to their dimensionality. The eigendecomposition of the deformation tensor was conducted with the help of a linear algebra package in a `python` interface: `scipy.linalg`. This package contains an algorithm that makes use of the Hermitian symmetry. At each grid point, we determine an *ordered* set of eigenvalues λ_i:

$$\lambda_1 > \lambda_2 > \lambda_3.$$

Based on the signs of these eigenvalues, without loss of generality, six unique and mutually exclusive cases can be separated, see table 4. These cases are based on the sign of the individual eigenvalues, as well as the sign of the summed eigenvalues. A positive eigenvalue indicates expansion along the direction of the corresponding eigenvector; a negative eigenvalue indicates compression.

Identification

Where there is collapse along each spatial direction, this unambiguously points towards the collapse of a region, which characterises the nodes of the cosmic web. Expansion in one direction occurring alongside compression in the other two indicates a prolate feature. If this prolate feature collapses on the whole, it forms a filamentary feature, otherwise it is classified as a void. Similarly, expansion along two directions and compression along one indicates an oblate feature. This is classified either as a wall or as another type of void, depending on the sum of the eigenvectors — which indicates the total expansion or contraction. Lastly, expansion along all directions characterises the expanding voids. These regions are expected to dominate the cosmic web spatially.

While the sum of eigenvalues in the regions labelled as prolate or oblate collapsing voids is positive, and these voids do expand overall, they still collapse along one or two axes (see also Sheth & van de Weygaert, 2004; Lavaux & Wandelt, 2010). They are flattened, as they get squeezed between other regions — typically between expanding voids on one hand and anisotropic massive features on the other (e.g. Dubinski et al., 1993). For a void to keep expanding, it is a necessary condition that it expands in all directions — although it is possible for these voids to increase in anisotropy (van de Weygaert & Babul, 1997). These voids have been shown to be rare in number, but dominant in physical volume (Hidding et al., 2012).

Discussion

There is a consequence to the method of detecting collapse and compression purely based on the *sign* of eigenvalues. Forero-Romero et al. (2009) were the first to point this out, see section 2.8 for a more detailed description of this. In brief, when a given structure collapses only slightly along a given direction, our web classifier treats it the same way as it would a full collapse. It makes no distinction based on the measure of collapse and expansion occurring in any direction, and may therefore misclassify some structures. Forero-Romero et al. establish some control over this by introducing a threshold eigenvalue λ_{th}, below which no collapse is ascertained. This strategy introduces an arbitrary choice for the value of λ_{th}. Our study attempts no such assumptions — de facto adopting $\lambda_{th} = 0$. For comparison, section A presents a visual example of a web classification with a non-zero threshold eigenvalue.

Web classification in a Gaussian random field

A simple additional numerical experiment has been conducted, to compare the web classification in near-Gaussian initial conditions to mathematical expectations. In this experiment, three Gaussian random eigenvalues are drawn at each cell in a grid of equal resolution to the data set — 128^3 cells. These are ordered within each cell, and the eigenvalue signatures are determined. A study of the relative amounts of volume occupied by different signatures provides a reality check for the web classification applied to the data set at high redshifts.

In this experiment, the stochastically generated divergence was defined as the sum of eigenvalues. The statistical distribution of this divergence per web component also provides a background to compare the distributions of divergence measured from the data set too. See section 6.5.

5.5 Determining velocity gradients

Other than the density perturbation, all physical quantities that are analysed in this study are combinations of spatial velocity derivatives. Recall the definitions of shear:

$$\sigma_{ij} = \frac{1}{2}(\partial_j v_i + \partial_i v_j) - \frac{1}{3}(\nabla \cdot \mathbf{v})\delta_{ij}, \tag{135}$$

vorticity:

$$\omega_{ij} = \frac{1}{2}(\partial_j v_i - \partial_i v_j), \tag{136}$$

and the deformation tensor:

$$D_{ij} = \sigma_{ij} + \frac{1}{3}(\nabla \cdot \mathbf{v})\delta_{ij}. \tag{137}$$

The quantities analysed in this study are then the velocity divergence:

$$\nabla \cdot \mathbf{v} = \partial_x v_x + \partial_y v_y + \partial_z v_z; \tag{138}$$

deformation eigenvalues:

$$\lambda_i \cdot \mathbf{a} = D \cdot \mathbf{a}; \tag{139}$$

and the shear and vorticity *magnitudes* are defined as

$$|\sigma| \equiv \sqrt{\sum_i \sum_j \sigma_{ij}^2}; \tag{140}$$

and

$$|\omega| \equiv \sqrt{\sum_i \sum_j \omega_{ij}^2}. \tag{141}$$

Discarding all information about any specific directionality, both $|\sigma|$ and $|\omega|$ solely measure the amount of local vorticity and shear in a region. It must be noted — see section 2.9 and particularly equation 96 for a detailed discussion — that the numerically measured vorticity largely consists of a *projection effect*. In other words, the measured values in general do not indicate real vorticity[45].

Normalisation

In order to make a comparison between the density and velocity gradient modes — i.e. divergence, shear and vorticity — each of these quantities needs to be normalised. The density perturbation $\delta + 1$ — see equation **??** — is obtained by dividing the density grid by its mean value. We add unity to the density perturbation for the sole purpose of simplifying the analysis, as $\delta + 1$ is guaranteed to be positive.

Using linear theory, the procedure for scaling the divergence with time follows from the continuity equation. The normalisation for any linear combination of velocity derivatives — e.g. any individual component of the shear and vorticity tensors — is done in exactly the same way. Considering the derivation leading towards equation 38, we get the proportionalities:

$$\nabla \cdot \mathbf{v} \propto Haf(\Omega),$$

and more generally

$$\partial_i v_j \propto Haf(\Omega).$$

[45]As can be seen in section 6.9, in our case the 'vorticity' measurement consists almost entirely out of noise.

Therefore, any velocity derivative quantity $\partial_i v_j$ is to be normalised as

$$\partial_i v_j \rightarrow 1 - \frac{\partial_i v_j}{Haf(\Omega)}, \tag{142}$$

in order to make the comparison with $\delta + 1$ at any given cosmic time.

This normalisation applies only to the linear regime of velocity growth. As soon as any velocity gradient becomes nonlinear, the linearised continuity equation is no longer an adequate basis for normalisation. The nonlinear time evolutions of velocity gradients is not currently described by any known analytical theories. Note, however, that the linear value gives an adequate point of reference.

In this case, the most logical tentative approach is to maintain the same normalisation as with the divergence. Then, it is not expected that the velocities have the same time redshift evolution as the density. Consequently, it is not expected that their statistical distributions match with the density distributions at low redshifts. However, a comparison of their shapes may lend insights. This comparison is made by fitting lognormal curves to the statistical distributions — see section 5.7 — after the data have been smoothed and different cosmic web components have been disentangled. See section 5.7.

5.6 Smoothing

Noise in the spatial distribution of the physical quantities can be smoothed out by convolving the grids with a Gaussian kernel. The workings of this procedure is described in section 4.1. Without smoothing, the data in the grid will still have an *intrinsic* smoothing scale, depending on the grid resolution. Table 3 lists the cell widths corresponding to different grid resolutions for a 30 Mpc box, these cell widths mark the lowest scale of detail that can be represented in the grid. Any kernel size exceeding the cell width can be chosen.

Other than smoothing out noise, convolving the grid with a kernel of a specific size allows us to follow the spatial distribution of quantities at that specific scale. In this capacity, the same process is also referred to as *filtering*, in the language of signal processing. For example, in order to follow the evolution of structures formed at very high redshift — which occur on very small scales — it is suitable to filter at a small kernel size. A large kernel, conversely, can provide an outlook on the assembly of structures emerging at lower redshifts — $\sim 8Mpc$ is a suitable size for structures at $z = 0$.

Other than a specific outlook on spatial distributions of various physical quantities, smoothing at a particular scale can also offer a specialised view on the statistical distributions. As such, smoothing at well chosen scales is a tool to follow e.g. the correlation between density and any velocity derivative mode across the evolution of structures of specific spatial scales.

In our study — in part bounded by the spatial resolution of the estimated fields — we found most structures when filtering at scales of 0.1 Mpc, 0.25 Mpc and 0.5 Mpc.

Where operations in Fourier space are concerned — e.g. smoothing / filtering, other forms of convolution, or differentiating via Fourier transformations — one warning particularly applies to the Cosmogrid dataset. For any spatial distribution of density in a box, there is a certain range of wavelengths required to represent it accurately in Fourier space. For individual structures that occur at a scale of several Mpc, the wavelengths in the Mpc region are very important — there is a lot of *Fourier power* in that part of the spectrum. When the largest structures occur at a scale comparable to the box size, these wavelengths can no longer be accurately represented in a Fourier transform. The Cosmogrid volume spans only 30 Mpc in each direction, and at lower redshifts, the largest structures occur at scales approaching this limit. The results of this effect can clearly be seen at the lower redshifts, and section A.3 further discusses this effect.

5.7 Lognormal distribution fitting

The expected statistical distributions of density and divergence — refer to section 2.6 — are of lognormal nature, the logarithm of a Gaussian distribution:

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{\log{(x-\mu)}^2}{2\sigma^2}\right\}, \tag{143}$$

where μ and σ are the mean and standard deviation of the Gaussian distribution that this lognormal distribution is based upon.

No theory has been formulated to describe fully the nonlinear evolutions of these distributions. However, some insight may be harvested from studying the shapes of these distributions, and how they vary between redshifts, spatial scales, and cosmic web components. In this study, the shapes of distributions are quantified by fitting lognormal curves, and studying the parameters of those fits. A lognormal distribution can be fully determined by two parameters: the median e^{μ} — where μ is the mean of the underlying Gaussian distribution — and the standard deviation *std*[46]. Some examples of lognormal curves are given in figure 46, which visualises how these parameters influence the distribution's shape.

A study of the redshift evolution of these fit parameters for a distribution of various quantities from various web components provides a way to follow the basic evolution of cosmic structures. It must be mentioned that any deviations from lognormal distributions do not show up in this analysis. Reducing a distribution to two parameters, we discard the finer details of these distributions. The reduced chi squared goodness of fits provides a way to gauge how great these deviations are.

This analysis is applied to the density and divergence distributions — section 6.5 — to the shear magnitude — section 6.7 — and the vorticity magnitude — section 6.9.

[46] we avoid denoting the standard deviation with the symbol σ, to prevent confusion with the velocity shear.

massive features	voids	
nodes rgb={.85,.22,.33}	**expanding voids** rgb={.25,.75,.40}	
filaments rgb={.21,.35,.79}	**prolate collapsing voids** rgb={.10,.07,.03}	prolate features
walls rgb={.42,.52,.49}	**oblate collapsing voids** rgb={.90,.50,.18}	oblate features

Table 5: Colours used in this thesis for plotting data from the six cosmic web components.

6 Results

Overview

The analysis techniques explained in the previous section were applied to various physical quantities measured from the Cosmogrid simulation: density, divergence, shear, deformation tensor eigenvalues and vorticity. This section exhibits the results, one physical quantity at a time. First, an outlook on the raw simulation data is provided in section 6.1. Then the DTFE estimated density is explored in section 6.2, and the divergence in section 6.3. These sections present the spatial and statistical distributions of the fields of interest.

Next, section 6.4 follows the classification of web components. The method explained in section 5.4 is used to identify six types of regions in the cosmic web.

Spatial and statistical distributions of the shear magnitude $|\sigma|$ are presented in section 6.7. The shear magnitude measurements are decomposed into the six different cosmic web components, and their statistical distributions are presented separately, and so are the correlations with density. An investigation is also made into the parameters of the lognormal fits to these separate distributions.

Section 6.8 presents a face-on view of a wall, displaying the density, divergence and shear magnitude. A substructure of filaments embedded in this wall is discovered.

Analogous to section 6.7, section 6.9 presents findings related to the vorticity magnitude $|\omega|$. Note, however, that these measurements are of a poor quality, and mainly consist of a projection effect, see section 2.9.

This section closes with a visual comparison between the TSC and DTFE estimators. While all other results are obtained by means of the DTFE technique, it is useful to note how this technique relates to an alternative estimator.

Within the scope of this section, two things are useful to note. Firstly, unless specified otherwise, all spatial plots show the distribution of the quantity of interest in the same slice of the Cosmogrid volume. A slice of $30 \times 30 \times 0.234$ Mpc^3.

Secondly, for convenience in interpreting all plots that decompose data from different cosmic web components, the same colour convention is used everywhere. Six different colours are assigned to the six different types of regions in the cosmic web, according to table 5.

6.1 Simulation: particle and velocity distributions

Particle positions

The positions of the particles from the simulation already exhibit clear structures. Particles from thin slices of the Cosmogrid volume are plotted in figures 26 and 27 — for redshifts 3.7 and 0.0, respectively. In both figures, the large upper panel shows a $30 \times 30 \times 0.469$ Mpc^3 slice; the lower panels show enlargements of the areas marked in black perimeters. The colour of each particle indicates the magnitude of its velocity, see figures 28 and 29 and related text for details.

At both redshifts, the particle maps show a hierarchical structure of underdensities, clusters and filamentary features. Note that, since these figures show thin slices of the cosmic volume, filaments appear as compact regions, and walls appear like filamentary features. At $z = 3.7$ the dominant structures appear to occur on a smaller scale than those at $z = 0.0$. The small scale structures at $z = 3.7$ — e.g. the anisotropic features in the zoom-ins on the right — are no longer found at $z = 0.0$. This is evidence of the expected merging of structures over the course of cosmic time.

The zoom-ins on the left show a void region, becoming sparser over time. A comparison between the two also reveals the formation of a thin filamentary structure.

Particle velocities

The raw particle velocities are shown in figures 28 — for redshift 3.7 — and 29 — for redshift 0.0. The velocities are sampled at a uniform random selection of particle locations, the vectors are projections of particle velocities onto the displayed plane. Vector sizes and colours indicate the velocity magnitudes, dark arrows indicate low velocity, red arrows indicate high velocities. The figures show 0.234 and 0.469 Mpc thick slices of the Cosmogrid volume in the large upper panels. The lower panels present enlargements of the highlighted areas.

These velocity maps clearly show the velocity streams pointing away from the void interiors, along filaments, and into the nodes of the cosmic web. Examples of the quadrupolar gravity field generated by two overdensities and two underdense regions are found everywhere — particularly in the right zoom-in at $z = 3.7$. Such shearing motions — compare these velocity flows to the those in figures 1 and **??** — are important aspects of structure formation.

A clear example of a sub-filament embedded in a void is found in the left zoom-in at $z = 0.0$. This is a filament[47] running upwards through a large void region. The velocities in this filament point outward, along with the expanding velocity stream of the void it is embedded in.

A very relevant observation of shell crossing can be made from these velocity maps. Generally, shell crossing first occurs at small scales, and spreads to larger scales later. In these maps, at $z = 3.7$ no shell crossing can be discerned

[47] A *tomography* of the cosmic volume — see section 6.8 — shows the three-dimensional appearance of this feature. This method allows us to exclude the possibility that the filamentary feature is some cross section of a wall.

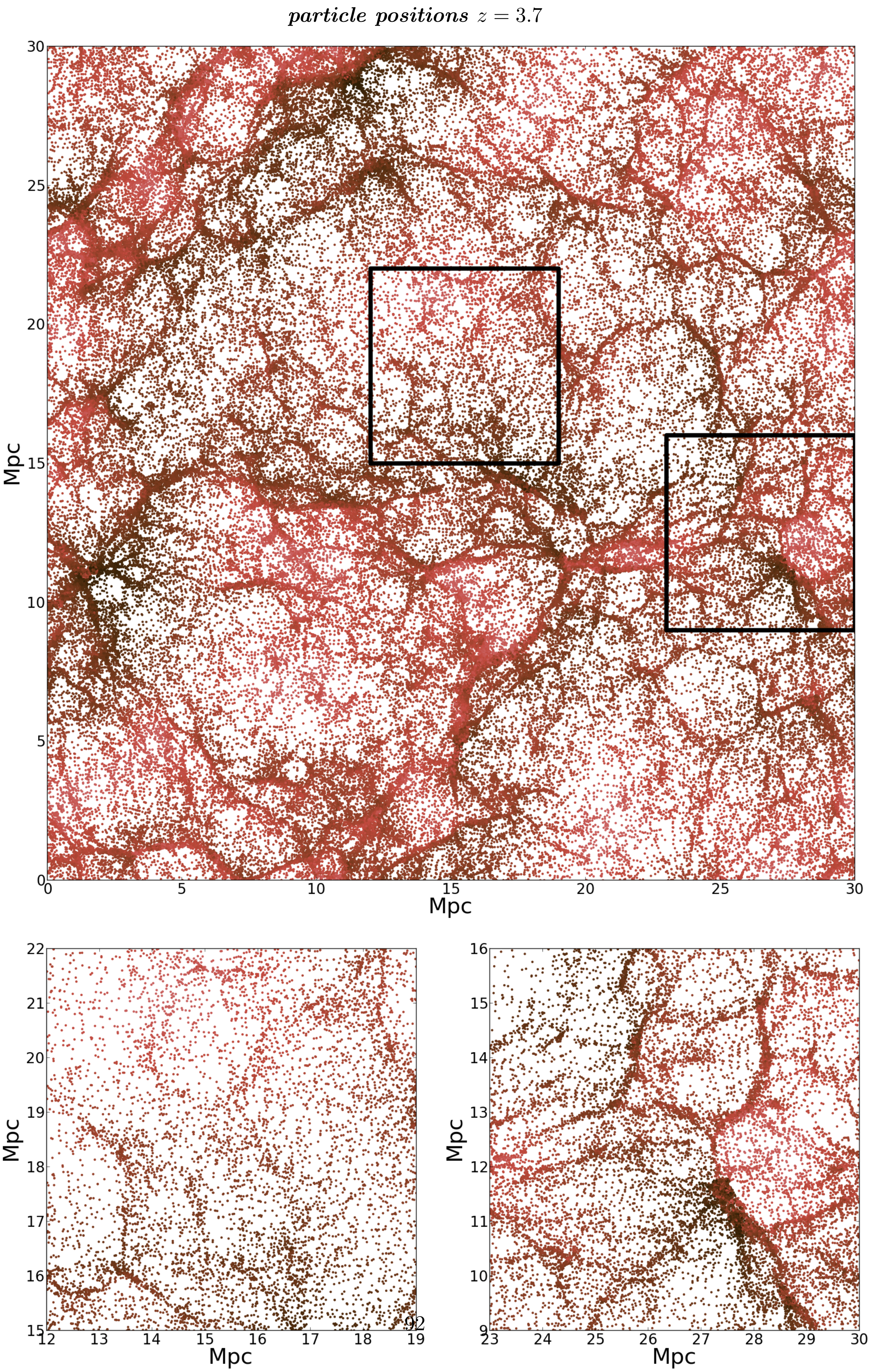


Figure 26: Spatial distribution of particles at redshift $z = 3.7$. Red: high velocity particles; black: low velocity particles. The large upper panel contains data from a slice of 0.469 Mpc in thickness. The lower panels are zoom-ins of the areas marked in the black borders, both depict a 0.703 Mpc thick slice.

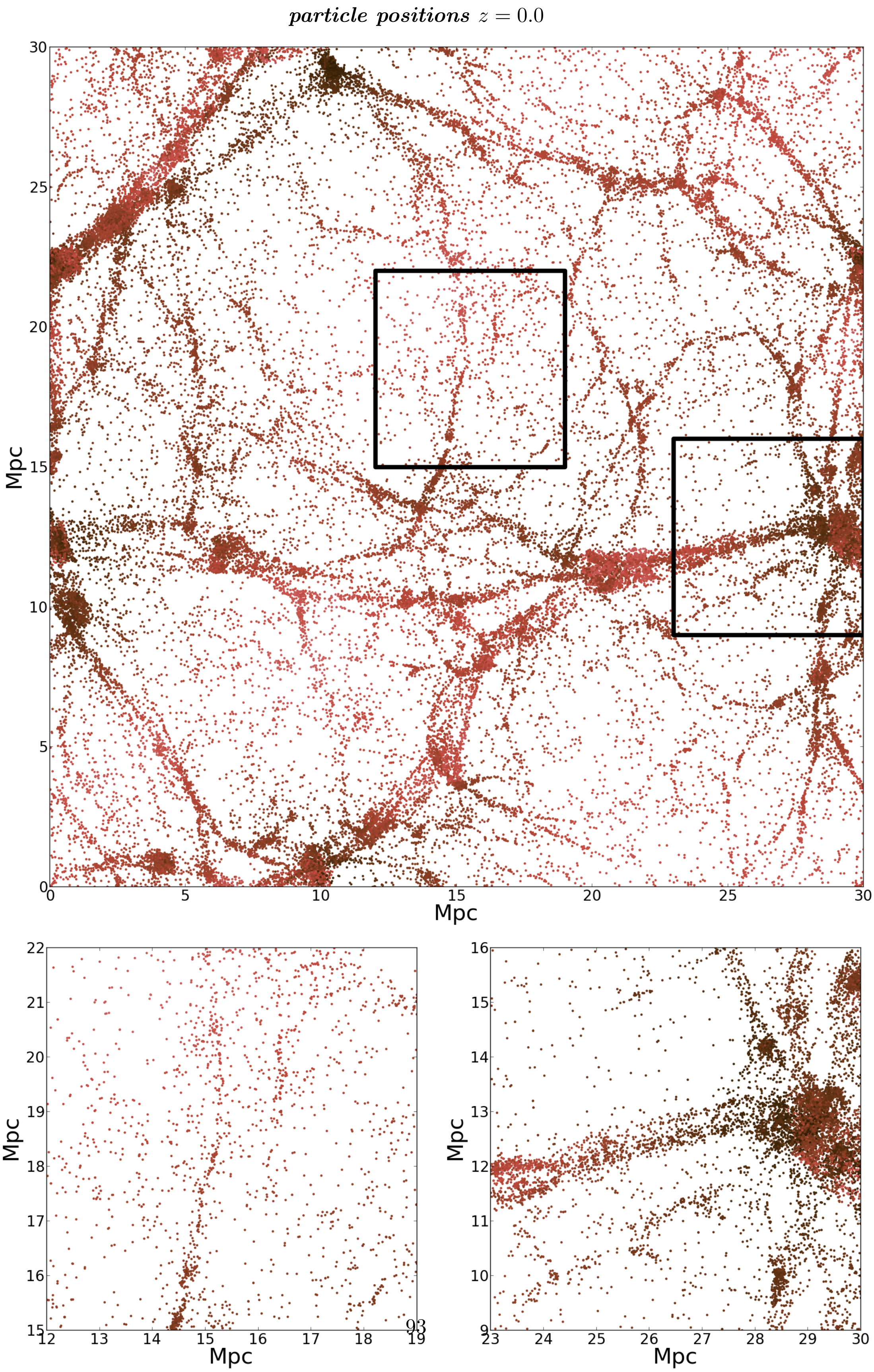


Figure 27: Spatial distribution of particles at redshift $z = 0.0$. Red: high velocity particles; black: low velocity particles. The large upper panel contains data from a slice of 0.469 Mpc in thickness. The lower panels are zoom-ins of the areas marked in the black borders, both depict a 0.703 Mpc thick slice.

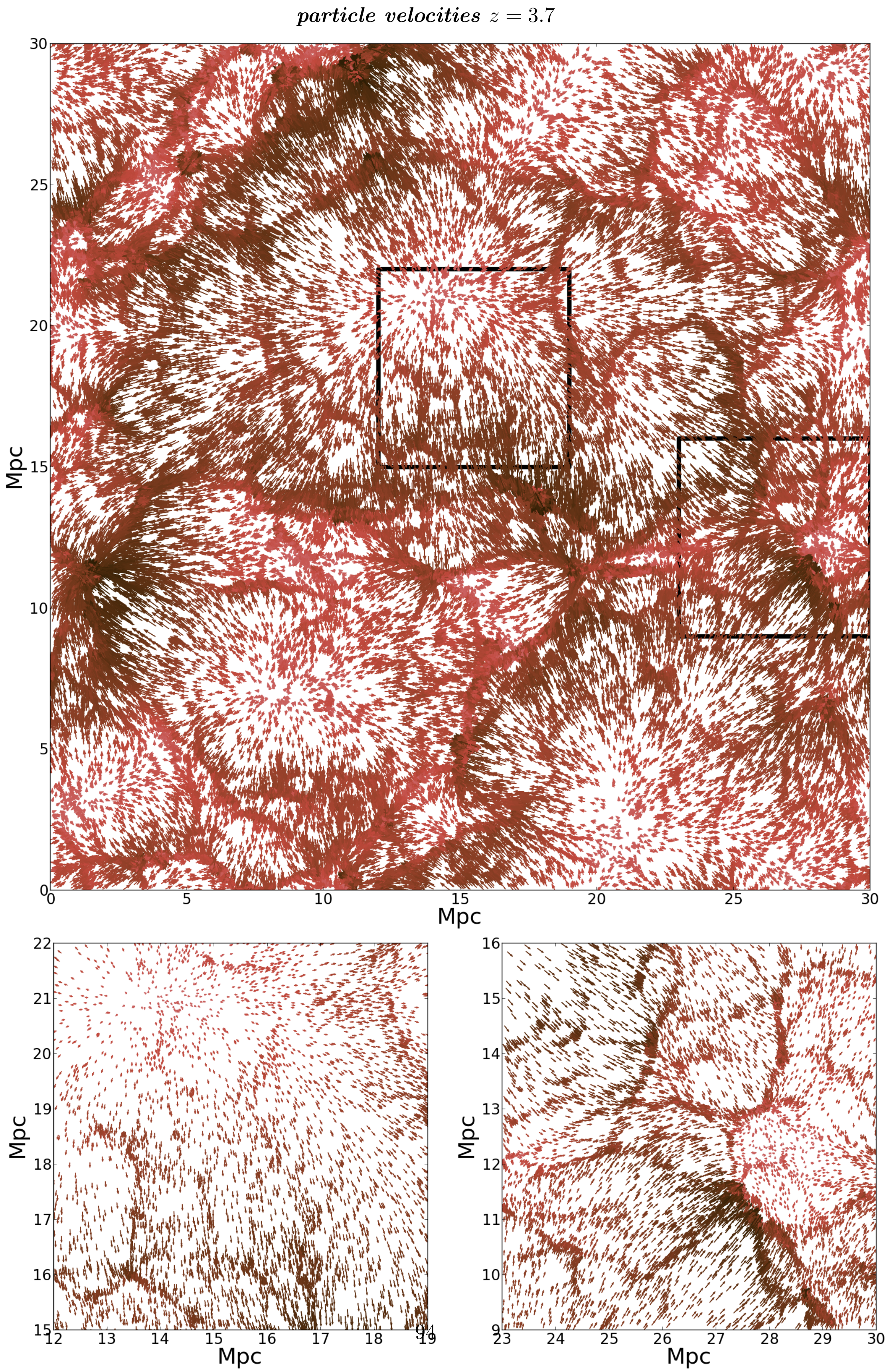


Figure 28: Particle velocities at $z = 3.7$. Red arrows: high velocity particles; black arrows: low velocity particles. The large upper panel contains data from a slice of 0.234 Mpc in thickness. The lower panels are zoom-ins of the areas marked in the black borders, but 0.469 Mpc thick.

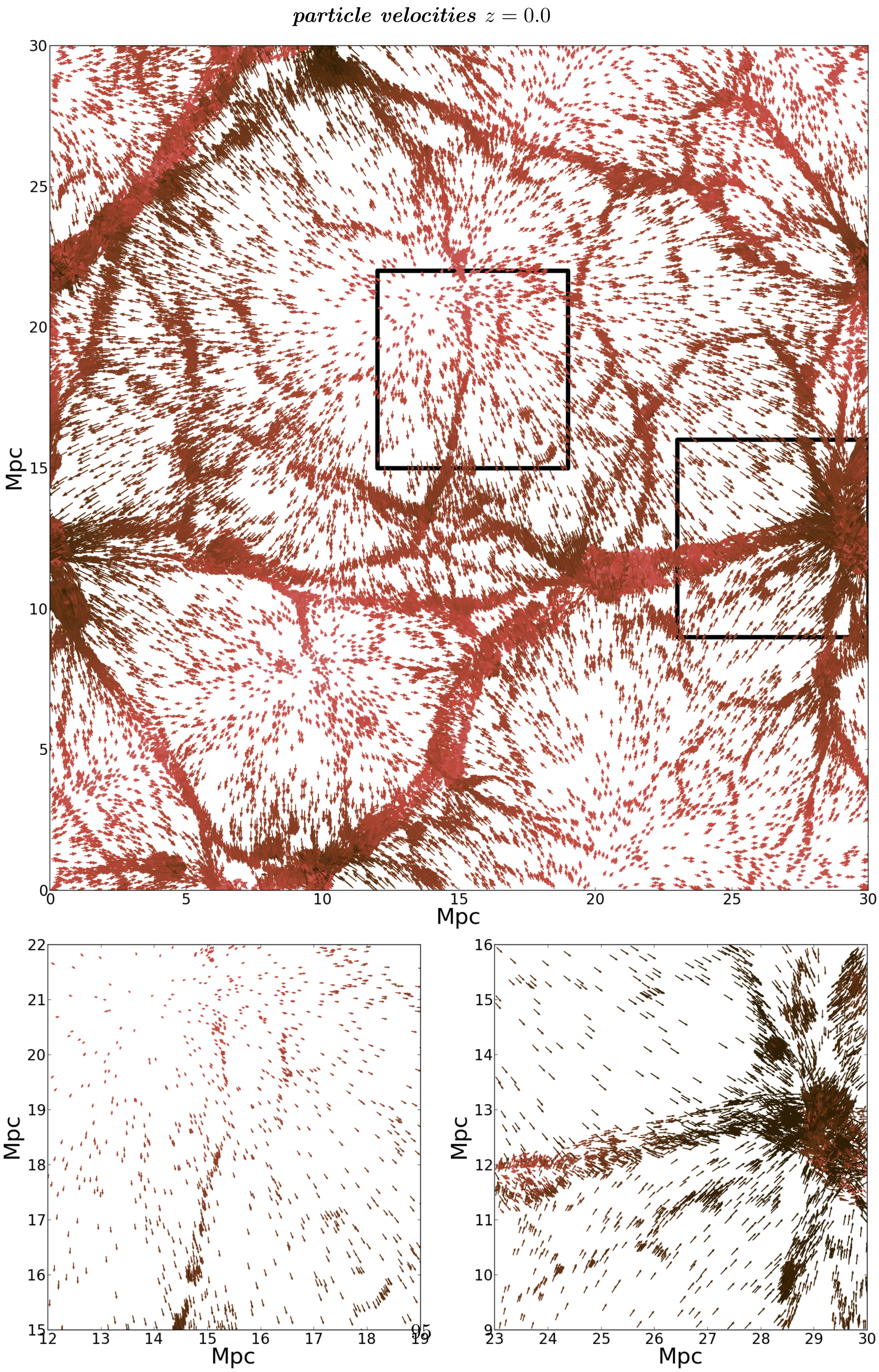


Figure 29: Particle velocities at $z = 0.0$. Red arrows: high velocity particles; black arrows: low velocity particles. All panels contains data from a slice of 0.469 Mpc in thickness. The lower panels are zoom-ins of the areas marked in the black borders.

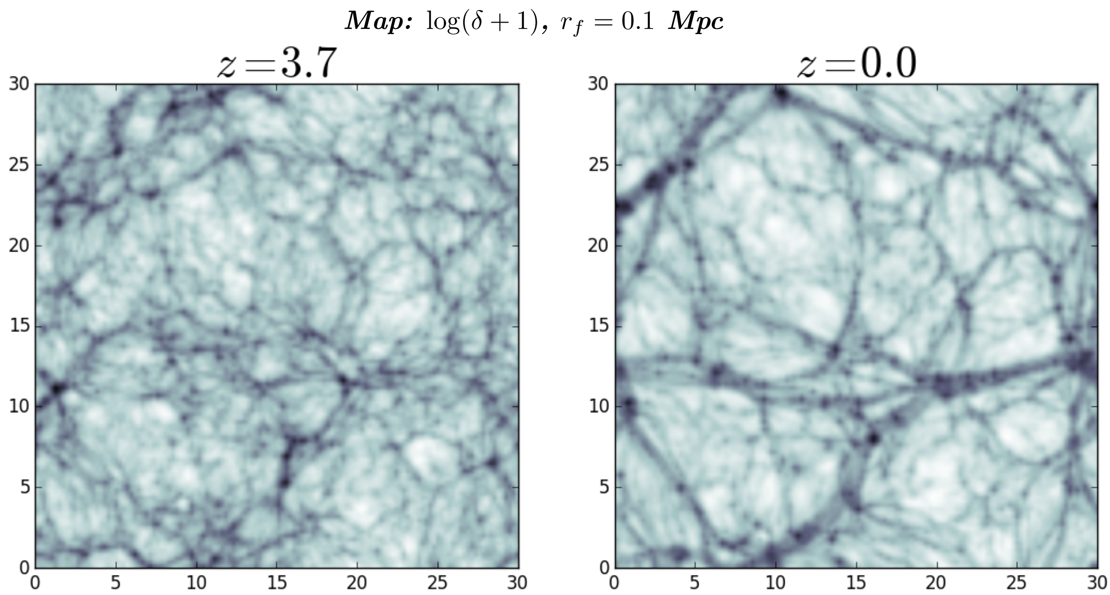


Figure 30: Spatial distribution of logarithmic overdensity at redshifts 3.7 and 0.0. Dark areas indicate high density, light areas are sparser. Data shown come from a slice of the Cosmogrid volume 0.234 Mpc in thickness. Horizontal and vertical coordinates are given in Mpc.

at all, but at $z = 0.0$, the overdense areas appear to be very active multi-stream regions. For example: the large horizontal structure across the right zoom-in in figure 29 appears to contain multiple velocity streams at the same place: flows pointing towards the node on the right, as well as the streams of matter flowing out from each of the two neighbouring voids. The cluster on the right side is an even more active place, where velocity streams from all adjacent filaments and voids come together.

Lastly, notice the substructure of this same large horizontal feature in the right zoom-in at $z = 0.0$. This feature appears to consist of two chains running parallel to each other — a *double-spined feature*. The velocity vectors — simply sampled at particle positions — appear to be more concentrated towards the edges of this structure, and sparser in the middle. Caution is required when making such observations by eye,[48]; these regions do not consist of separate parallel overdensities. Their double spined appearance is an artefact due to shell crossing, see section A for a discussion. Double-spined features like these occur often in the spatial distributions of velocity-related quantities — see figures 36, 32, 57, and 51. Section 6.4 will reveal that these features are walls, and section 6.8 will explore these objects further.

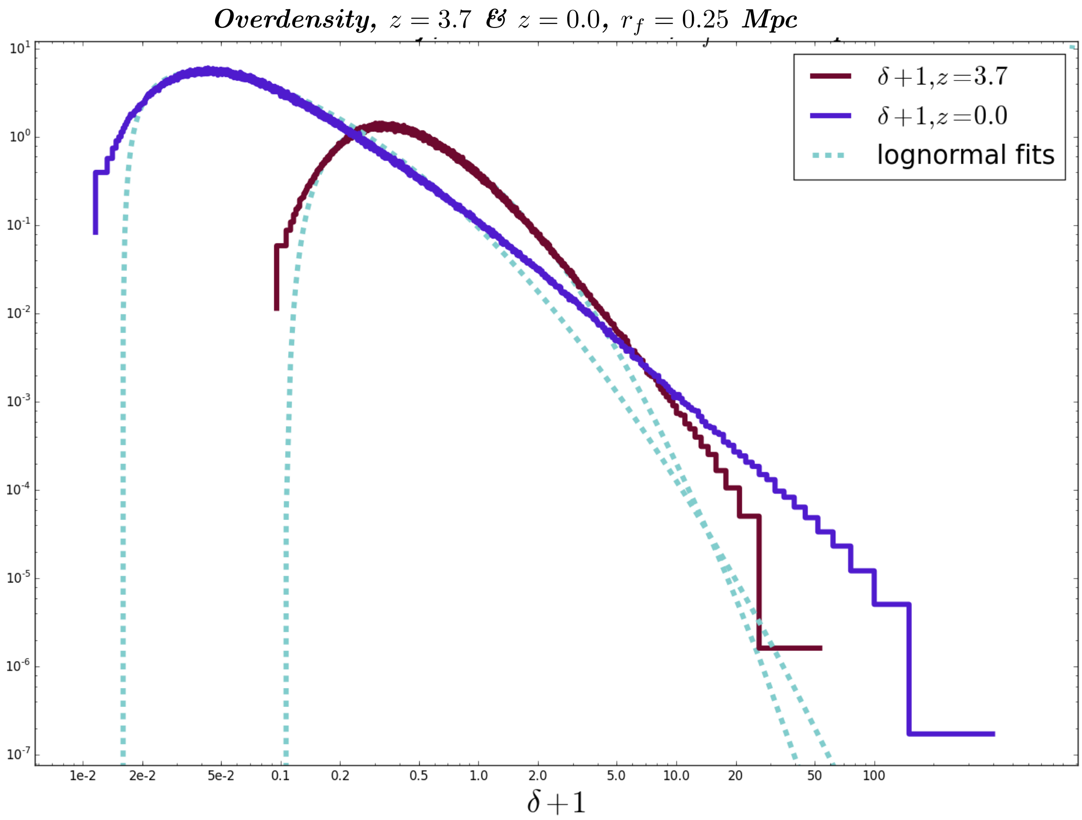


Figure 31: Statistical distributions of density at $z = 3.7$ (red) and $z = 0.0$ (purple). The dashed blue lines are lognormal fits to the distributions. Notice the elevated high density tails.

6.2 Density

Maps

The DTFE-estimated density fields of two snapshots are mapped out in figure 30. Again, the evolution and hierarchical nature of structures are clearly visible, as is the merging of structures over time. Complex strands of small scale structures can be found everywhere at $z = 3.7$; but at $z = 0.0$ they form into fewer larger structures. Similarly, groups of small underdensities at $z = 3.7$ appear to form larger and deeper voids over time.

Notice the sharp localised density peaks at the nodes, and the smaller spikes in density along the anisotropic features. These compact features form the high-density tail of the statistical distributions.

Statistical distribution

Figure 31 shows the statistical distribution of the density at redshifts 3.7 and

[48] especially since the arrows are predominantly horizontal, enhancing our tendency to notice horizontal correlations visually.

0.0. The quantity $\delta + 1$ is shown in a fixed-weight histogram, in which each bin holds a fixed number of data points. From the bin widths it can be seen that the sampling density is higher in the low density end, and lower in the high density tail. This means smaller errors in the low density regions, and greater uncertainty towards the high density end of the distribution.

It can be seen that the densities span a far greater range at $z = 0.0$. The spreading of the low density end of the distribution towards the $\delta + 1 = 1 \cdot 10^{-2}$ mark indicates the deepening of void underdensities over time.

It is plausible that the long tail in the lognormal distribution already accounts for the nonlinear growth of density in the bound regions (Coles & Jones, 1991). This suggests that the elevated high-density tail in the presently measured distribution is in part a spurious enhancement. However, Uhlemann et al. (2015) have produced analytical density distributions[49] that match the measured distributions in figure 31 in appearance. The analytical distributions by (Uhlemann et al., 2015) reproduce the straight downward slopes; and a high-density knee. This knee is not clearly present in our results, but at high densities the statistical certainty is not enough to verify this.

The light blue dashed lines in figure 31 show the lognormal fits to the density distributions — see section 5.7. In this study, lognormal fits are made to all the density, shear and vorticity distributions in this study — including those from separate cosmic web components. These fits are compared in sections 6.5, 6.7 and 6.9.

Notice the large deviation between the distribution at $z = 0.0$ and the lognormal fits in the high-density tail. The high-density measurements are less significant compared to low densities, but the deviation is very clear. We conclude that the density distributions are not lognormal, which is in line with analytical work by Uhlemann et al. (2015). The other distributions in this study — i.e. velocity related quantities — are far better approximated by lognormal curves.

6.3 Divergence

Maps

Figure 32 shows the spatial distribution of divergence at two redshifts. It plots the positive logarithm of positive divergence $\log |\nabla \cdot \mathbf{v}|$ in orange tones — the darkest reddish shades indicating the most extreme divergence. The negative logarithm of absolute divergence $-\log |\nabla \cdot \mathbf{v}|$ is shown in dark colours — the darkest colours indicate the greatest convergence of velocity. The white areas indicate where the divergence is precisely zero. This figure shows the evolution of structures at various spatial scales in parallel.

The divergence maps bear a great similarity to the density maps, reflecting the infall of matter towards high-density regions, and the outflow from empty regions. Two interesting dissimilarities may be observed.

Firstly, recall the sharp spikes that populate the density maps, they are lacking in the divergence. For as far as they indicate realistic density enhancements, the velocity streams do not seem to converge on these points proportionally.

Secondly, notice in the $z = 0.0$ map some elongated features consisting of two convergent strips enclosing a highly divergent region. In section 6.4 we will

[49] Those were produced using the Large Deviation Principle (Bernardeau & Reimberg, 2015).

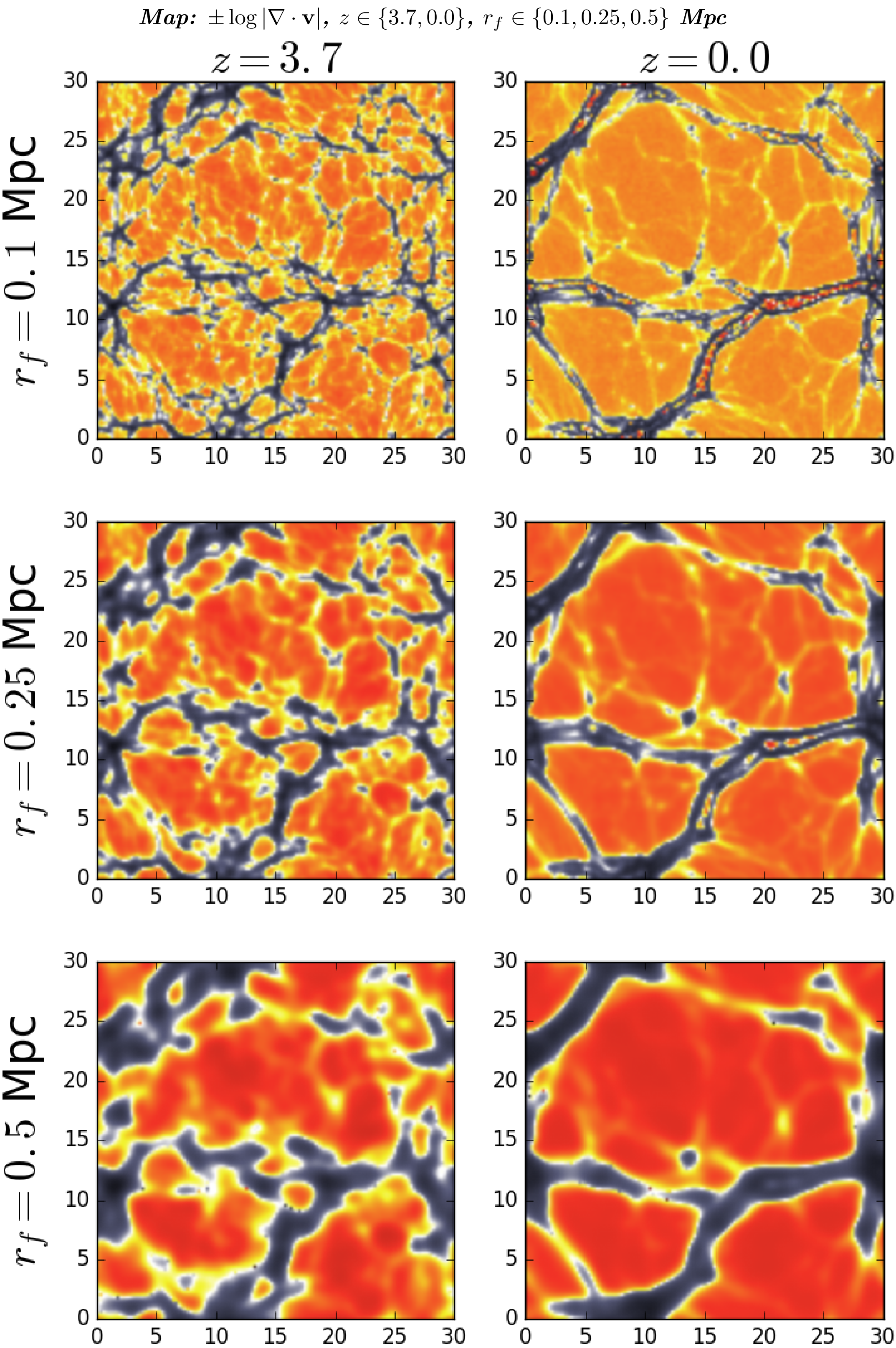


Figure 32: Spatial distribution of logarithmic velocity divergence at $z = 3.7$ (left column) and $z = 0.0$ (right column). Maps in the top row are filtered at 0.1 Mpc, those in the middle at 0.25 Mpc, and those at the bottom at 0.5 Mpc. Orange tones indicate positive divergence, dark regions have a negative divergence — i.e. convergence. White regions have $\nabla \cdot \mathbf{v} = 0.0$.

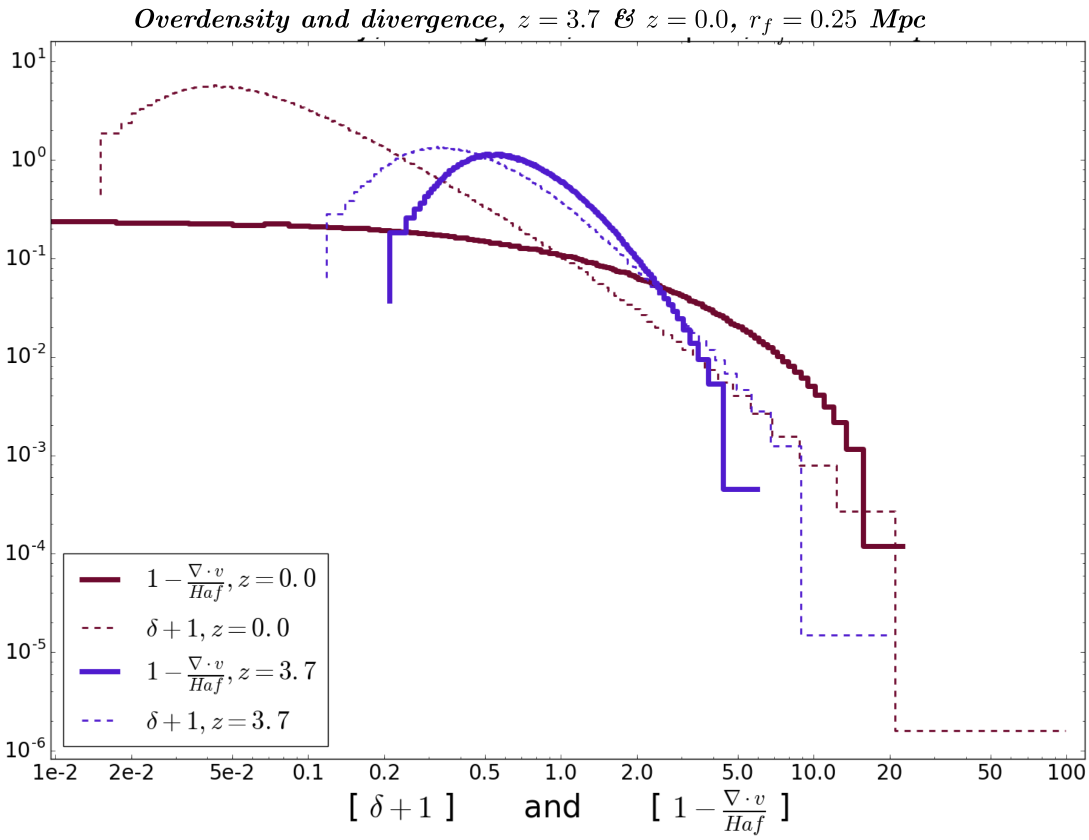


Figure 33: Statistical distribution of velocity divergence (solid lines) and density (dashed lines) at $z = 3.7$ (purple) and $z = 0.0$ (red). The divergence has been normalised to allow a comparison with the density distributions.

find that these are walls. They do not consist of parallel high-density sheets — the density maps indicate no such structure. Rather, the *double spined* appearance of these walls is an artefact in the DTFE estimator. Briefly put, shell crossing occurs the interior regions, and the DTFE estimator performs poorly there, see also (Hahn et al., 2015). The overlapping of various velocity flows in these areas results in the spurious measurement of a very large net divergence. Section A explains this in more detail.

The high convergence measurements at either side of a wall is attributed to the velocity flows in the immediate environment. They cause the velocity-based web classifier to identify anisotropic structures regardless of the local matter density. These double spined appearances are also recovered by Hoffman et al. (2012), who used a velocity-based web classifier. See figure 10.

Even though figure 32 shows data filtered at a low spatial scale, it shows little structure on small scales in divergence, compared to the density maps in figure 30. Velocity flows — when not subjected to nonlinear effects — are a large scale phenomenon, as explained in section 2.6.

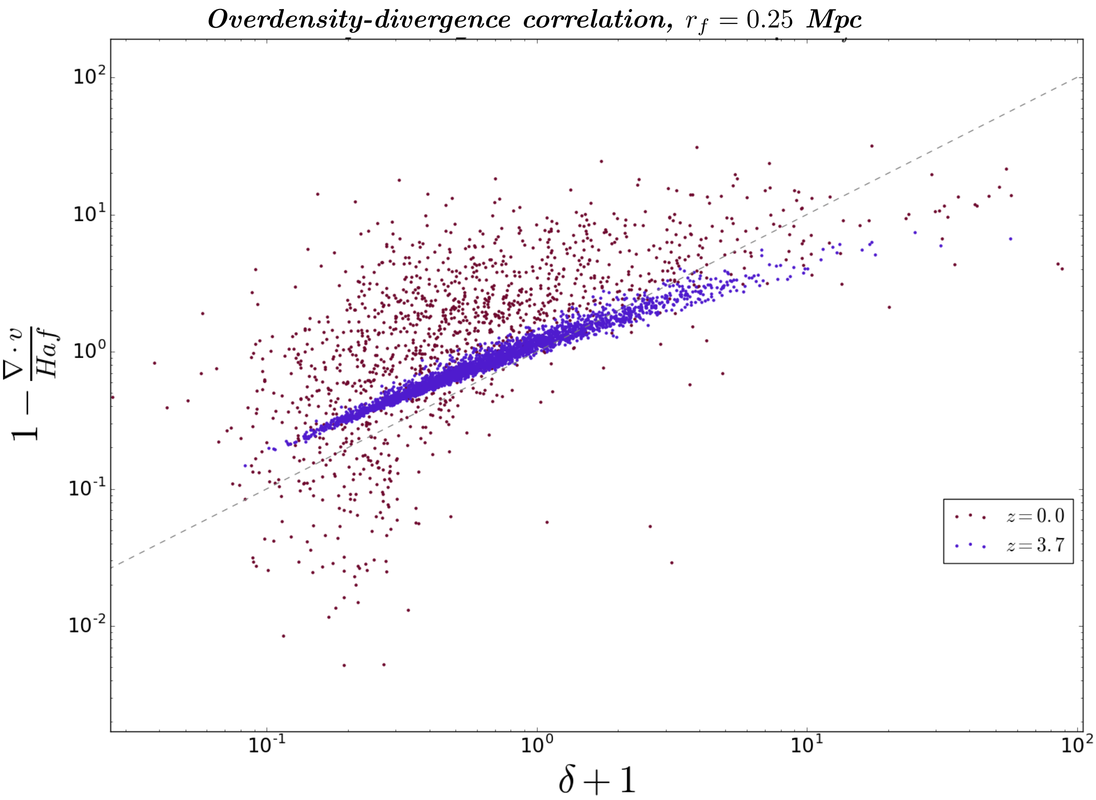


Figure 34: Correlation between velocity divergence and overdensity at $z = 3.7$ (purple) and $z = 0.0$ (red). The divergence has been normalised to allow a comparison with the density distributions.

Statistical distribution

Figure 33 displays the statistical distributions of divergence at $z = 3.7$ and $z = 0.0$. It also shows the density distributions on dashed lines. The divergence has been normalised so that it can be compared to the density distribution, see section 5. It is clear that the distributions from $z = 3.7$ form a far better match than those from $z = 0.0$.

Even so, the divergence distributions do not show the same elevated high end tail like the density distributions do. The divergence distributions are much better described by lognormal distributions. Do note that the $z = 0.0$ divergence distribution spreads into the negative domain, where $\nabla \cdot \mathbf{v} > -Haf$.

Divergence-density correlation

Figure 34 shows the correlations between divergence and density at redshifts 3.7 and 0.0. Corresponding densities and divergences from a uniform random selection of grid cells are plotted as dots. The dashed grey line indicates where $1 - \frac{\nabla \cdot \mathbf{v}}{haf} = \delta + 1$: a perfectly correlated body of data points would lie precisely on this bisector. The overall distribution of data points indicates that the divergence is well correlated with density.

The divergence and density from $z = 3.7$ are clearly more narrowly correlated, compared to those from redshift 0.0. The narrow body of data points is also slightly tilted with respect to the $y = x$ line: the correlation is somewhat shallower than unity. This reflects the previously made observation that the density distribution continues to high values, while the divergence does not. The correlation at $z = 0.0$ has spread mostly to higher convergence values; only at the lower left end do we see the measurements dropping towards negative values.

6.4 Cosmic web classification

Component maps

The cosmic web components were identified by the method described in section 5.4. For each grid cell, we determined the signature of the deformation tensor eigenvalues. The deformation tensor is defined as:

$$D_{ij} = \sigma_{ij} + \frac{1}{3}(\nabla \cdot \mathbf{v})\delta_{ij}, \tag{144}$$

and it has three eigenvalues, ordered as

$$\lambda_1 > \lambda_2 > \lambda_3.$$

A positive eigenvalue indicates local expansion along the direction of the corresponding eigenvector — and a negative eigenvalue indicates contraction. The sum of eigenvalues is therefore related to the local density. This has been used as a basis for the classification of six unique cosmic web components. Table 4 presents the definitions of these types of regions.

Section 2.8 provides background information on different web classifiers that have been used in previous studies. It is relevant to restate that the classification algorithm used in this study does not make assumptions on the threshold λ_{th} eigenvalue to determine whether a structure is collapsing or not along a given eigenvector direction — in effect it takes $\lambda_{th} = 0$. This decision simultaneously relieves us of arbitrary choices for the free parameter λ_{th}, but also introduces the risk of mistaking a slight compression for a full collapse of a structure. See also section A for a discussion.

The resulting classification of web components for $z = 3.7$ and $z = 0.0$ is displayed in figures 35 and 36. The top left panels show filaments in purple and nodes in yellow. Prolate void regions appear to be very rare, and limited in size — see figure 39 — and have not been demonstrated to be of interest in structure formation. As such they are not plotted. Note that, since the figures show a thin slice of the Cosmogrid volume, filaments appear as compact regions, and walls appear like filamentary features.

The features in the $z = 3.7$ plot were determined at a spatial scale of 0.10 Mpc, which appears to form a lower bound on the smallest structures that can be seen in the simulation at that redshift. For the $z = 0.0$, the dominant structures appear on larger scales. The classification was done on 0.25 Mpc.

It is clearly visible on both figures that the oblate void regions form the peripheral parts of the voids, surrounding the expanding void regions. This is a very adequate confirmation of the expectation that the outer void regions get squeezed[50] between expanding voids and massive surrounding regions. These

[50]Vastly more likely along one direction — Oblate — than along two — Prolate.

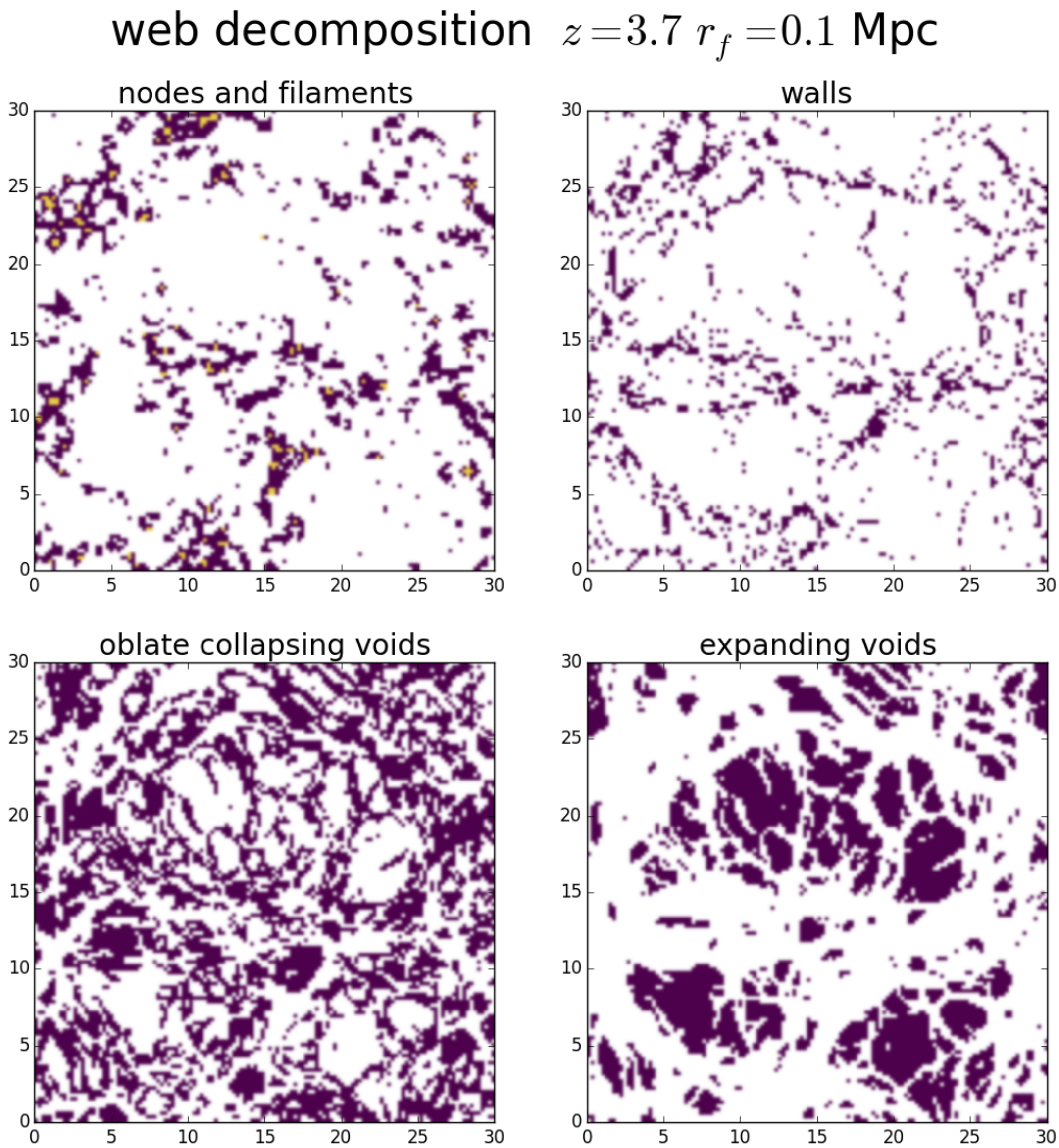


Figure 35: Results of the cosmic web classification from a slice of the Cosmogrid volume at $z = 3.7$. Data are filtered at a scale of 0.1 Mpc. The top left panel shows regions classified as nodes in orange, and filaments in purple. The other panels show the mentioned web components in purple. Note that there is no overlap between these regions. Prolate collapsing void regions are not shown. Since this figure shows a thin slice from the simulation volume, filaments tend to appear as compact structures, and walls appear as filamentary features.

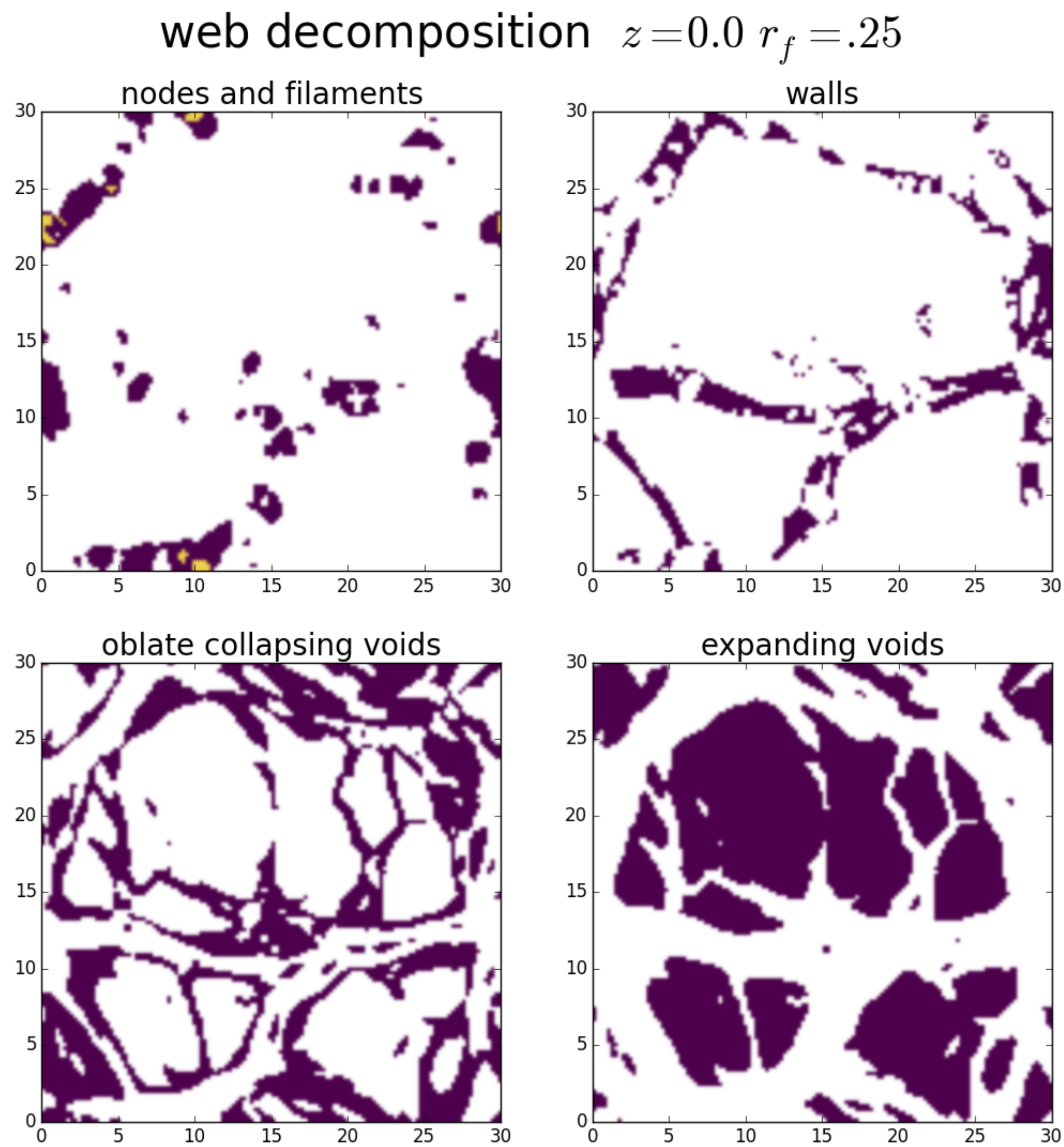


Figure 36: Results of the cosmic web classification from a slice of the Cosmogrid volume at $z = 0.0$. Data are filtered at a scale of 0.25 Mpc. The top left panel shows regions classified as nodes in orange, and filaments in purple. The other panels show the mentioned web components in purple. Note that the expanding void regions are surrounded by large oblate collapsing void regions. Notice the double-spined appearance of the walls.

oblate void regions enclose the expanding regions systematically throughout redshift and distance scales. What is not unsurprising is the sheer consistence and prominence of these oblate void regions: most expanding void regions are almost completely enclosed in oblate void regions. Moreover, oblate void regions occupy a far larger proportion of total space than any other type of region, at any redshift and on any scale — see figure 39.

At $z = 0.0$, the vertical sub-filament crossing the large void, described in the previous subsection, does not appear as a filament. The left zoom-ins in figures 28 and 29 show that the velocity flow in this region is dominated by the outflow of the surrounding void — no velocity flow along the structure's principal direction is seen. This is reflected in the fact that it is identified as an oblate collapsing void region, being compressed between two expanding voids — see figure 36.

Dubinski et al. (1993) model a simplified model of void mergers, resulting in the same oblate structures we see. In their simulated scenarios, spherical empty regions expand and merge together, trapping any matter sitting between them. This then forms an underdense structure that is sheet-like in dimensionality, and diminishes in density. There is no fundamentally defined boundary separating underdense walls from oblate collapsing void regions. Our web classifier draws that line at $\sum_i \lambda_i = 0$, and thus identifies oblate void regions mostly in underdense regions.

The squeezing of oblate regions between merging voids also manifests in velocity shear. This can clearly be seen in our spatial distribution of shear magnitude, figure 57. Notice from figures 35 and 36 that the identified oblate collapsing void regions occupy a very large proportion of total space. The next subsection will address this. In brief, this is a consequence of the fact that our classifier assumes no non-zero threshold eigenvalue λ_{th}.

It is in the spatial distribution of the walls that great differences between redshifts and distance scales arise. Figures 35 and 36 show the emergence of sheet-like structures on very large scales, and at $z = 0.0$ they appear as double spined features. Do note that this double spined appearance is an artefact due to the DTFE estimator performing poorly on multi-stream regions. These figures only provide edge-on views of walls, but section 6.8 will explore these features further.

Our web classification strategy is based on velocities, which are a large scale phenomenon — see section 2.6. One consequence of this is that velocity flows cause the web classifier to identify anisotropic structures regardless of the local matter density. This way, a low-density area, some distance away from a wall, may be identified as a wall region due to the velocity shear occurring there. These double spined appearances are also recovered by Hoffman et al. (2012), who used a velocity-based web classifier. See figure 10.

Oblate void regions

The fact that expanding void regions are nearly completely enclosed by oblate void regions is illustrated in figure 37. The two upper panels show voids with large bulk motions, and the lower left panel shows the oblate regions in an intersection of massive elongated structures. Most oblate void regions appear to have a somewhat higher sampling density than the expanding void regions — and this is confirmed in section 6.5.

Most of the detected oblate void regions have a slim, elongated appearance

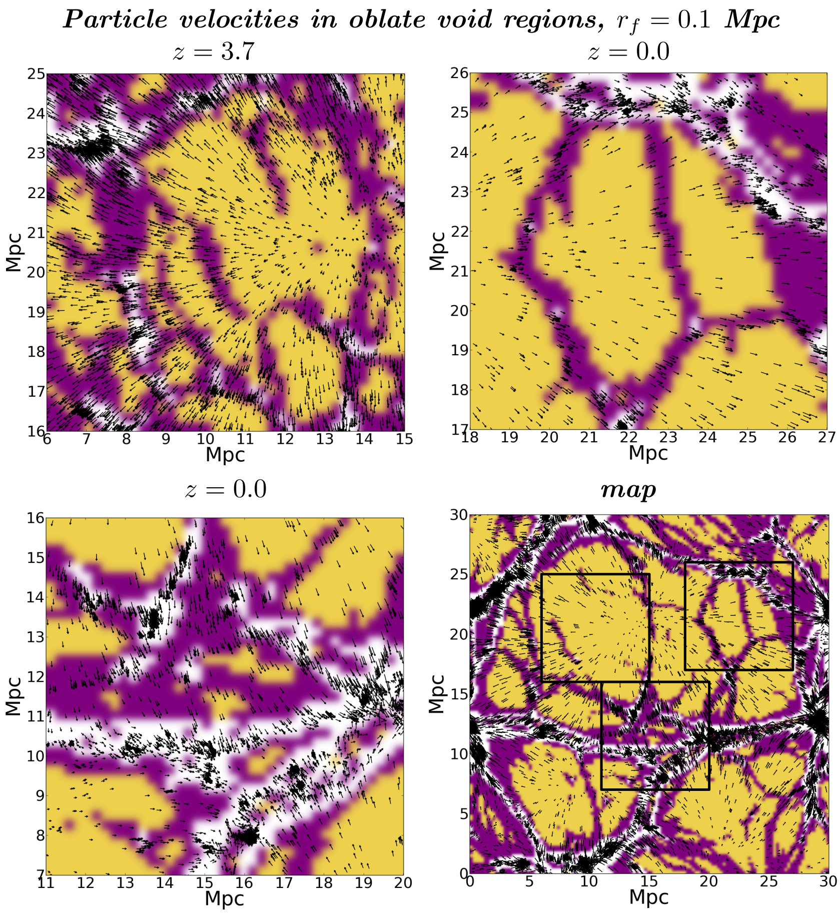


Figure 37: Decomposition of several regions in the Cosmogrid volume. Purple areas mark oblate void regions, yellow marks expanding void regions. Vectors indicate particle velocities. All data are filtered at 0.1 Mpc. The top left panel corresponds to the region marked at the top left in the map, but data are from redshift 3.7. The other panels correspond to the other areas marked in the map, both at $z = 0.0$

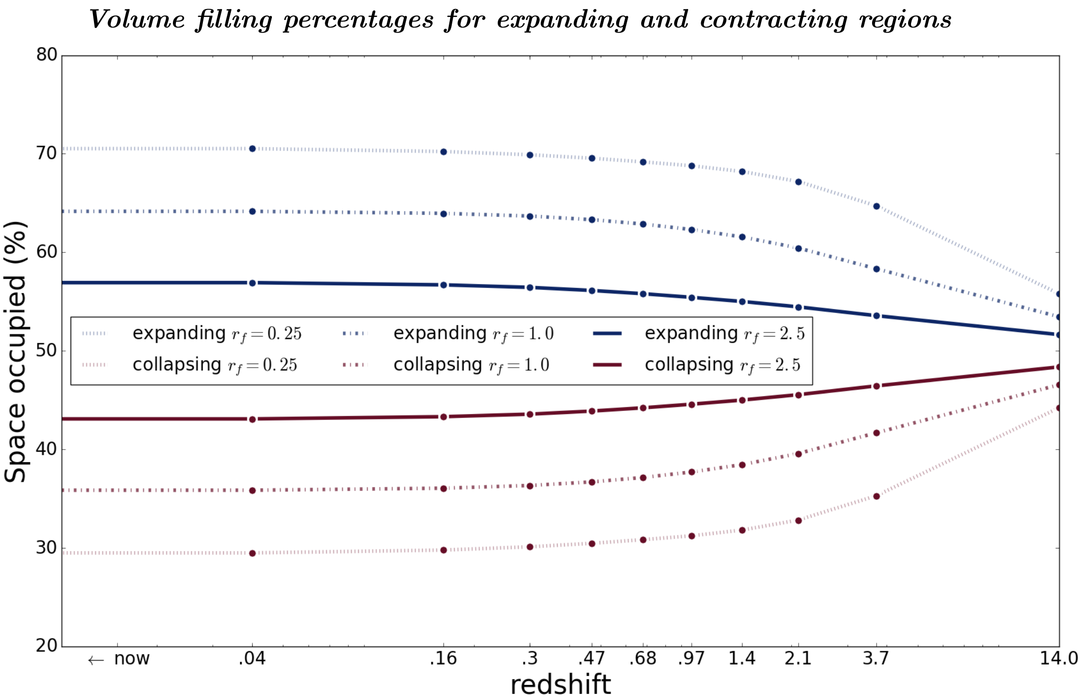


Figure 38: Relative amounts of space occupied by regions that are: on the whole expanding (blue) and on the whole contracting (red). Solid lines: data from $r_f = 2.5$ Mpc, dash-dotted: $r_f = 1.0$ Mpc, dotted: $r_f = 0.25$ Mpc. The separation of the two region types is made on the basis of the sign of summed eigenvalues, see text for details.

on these maps[51]. In most of the cases, velocities are generally oriented perpendicular to the principal direction of the oblate region. This — together with a low particle density — suggests the presence of actual oblate void regions, being squeezed between expanding void regions and massive components.

Occasionally, velocities are more aligned with the oblate regions. This suggests that matter is trapped in underdense wall-like structures, and flowing outward in the features principal plane[52].

Volume occupation

Figure 38 shows a breakdown of the cosmic web into expanding and collapsing regions. The expanding regions are determined by the criterion that $\sum_i \lambda_i > 0$, and collapsing regions by $\sum_i \lambda_i < 0$. Note that, in Gaussian initial conditions, these regions are expected to be precisely balanced — i.e. 50% of space expands, and 50% contracts. An extrapolation of the highest redshift measurements suggest that this balance occurs at redshifts upwards of $z \sim 34$. Note that the Cosmogrid simulation starts at $z_{init} = 65$.

The relative amounts of space occupied by the six distinguished cosmic web components are compared in figure 39. It follows the percentages of occupied

[51] Note that this is the likely appearance of any cross section of any oblate shape.

[52] A reliable distinction between these cases would require a bulk-correction, where the particles velocities are corrected by an estimated velocity field from some larger spatial scale.

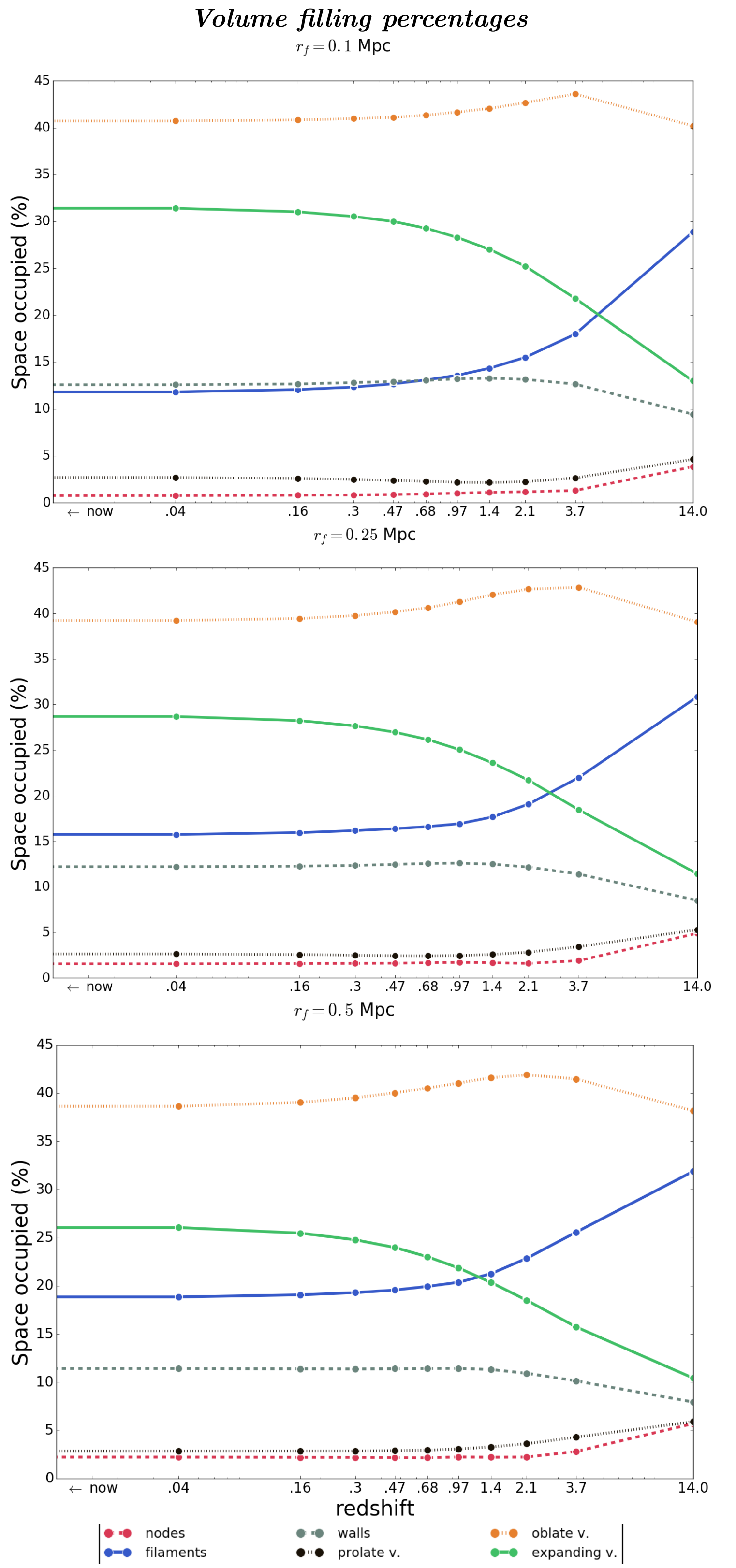


Figure 39: Relative amounts of space occupied by the six different cosmic web components, as a function of redshift, at filter scales 0.1 Mpc, 0.25 Mpc and 0.5 Mpc. A comparison between these plots allows us to probe the differences between the appearance of the cosmic web at different scales.

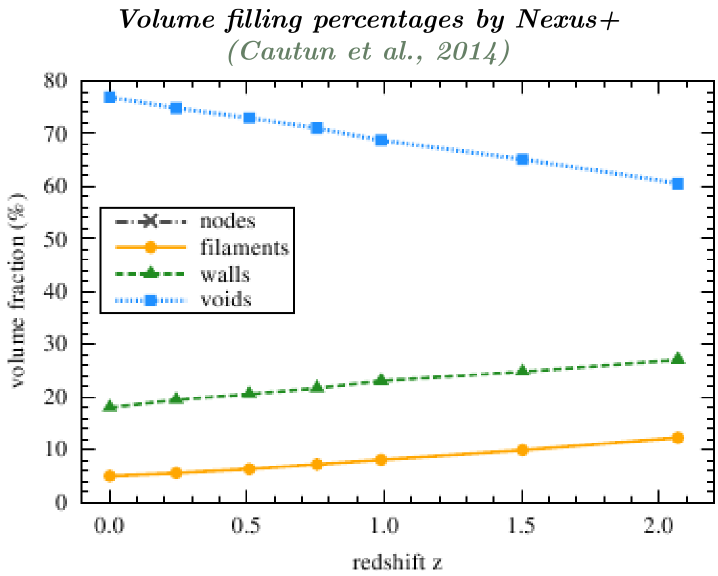


Figure 40: Relative amounts of space occupied by the web components identified in Cautun et al. (2014). They used the Nexus+ method — see section 2.8. See their paper for details.

space as a function of redshift. Note that the horizontal axis indicates logarithmic redshift, resulting in a roughly logarithmic time line running from right to left. The top panel displays data from the web components determined at spatial scales around 0.1 Mpc, while the middle and lower panels show structures at 0.25 and 0.5 Mpc, respectively.

A comparison between the occupation levels of web structures in these three panels shows how sensitive the data are to the spatial scales considered. Particularly the prominence of filamentary regions undergoes a great transformation when the focus is shifted from low to high scales: while they play a relatively modest role in the small scale structures, filaments are the dominant type of massive features in the large scale structure of the Universe.

Note that an assessment of structures on a given scale requires a sufficient spatial resolution. This study is based on a velocity estimation at a grid resolution of 128 cells spanning a 30 Mpc box length, resulting in a cell width of 0.234 Mpc. A finer *intrinsic smoothing* would be preferable for a proper view on the 0.1 Mpc structures.

These plots also offer a general overview of the time evolution of the structures. At each scale, there seems to be a break-even point between filaments and expanding voids, and this happens at a progressively later time as higher scale structures are considered. This is a clear manifestation of the hierarchical development of linear and nonlinear structure formation, where small scale structures mature before large scale structures do. On the whole, a large part of the 'shifts in balance' involved in structure formation appears to be completed at a relatively early stage in cosmic history — most of the change in volume occupation occurs at higher redshifts.

For a comparison, figure 40 shows the volume filling percentages obtained by the Nexus+ technique[53] applied to the Millennium simulations (Cautun et al., 2014). They find voids to keep expanding linearly with redshift, at the expense of walls and voids — this is roughly in agreement with our results. In contrast to our study, Cautun et al. find walls to dominate over filaments throughout the redshift range. The difference between the measured volume occupation percentages for walls can be attributed to the difference between the web classifiers used. Our web classifier identifies underdense walls as oblate collapsing void regions, which are found to be surprisingly dominant.

The oblate void regions detected by our web classifier dominate the volume at every scale and every redshift studied. They increase from the near-Gaussian initial conditions, peak around $z = 3.7 - z = 2.1$, and decrease somewhat in prominence at later times and at higher scales. Note that the majority of structure formation occurs at high redshifts, which may explain the peak in oblate void volume fractions. The fact that it peaks later at larger scales is in line with the expected hierarchical structure formation, too.

A separate numerical experiment was conducted in order to gauge the statistically expected volume filling percentages in Gaussian initial conditions — See section 5.4. Since there are 8 possible unordered eigenvalue signatures each is expected to occur in 12.5% of a Gaussian random volume (Doroshkevich, 1970). While, for example, only one of these signatures — i.e. (+ + +) — corresponds to expanding void regions. The condition of oblate regions, however, may be satisfied by three signatures — i.e. (+ + -); (+ - +); (- + +) — so that 37.5% of space is expected to be oblate. Our experiment separates the generally contracting and expanding regions by the sum of eigenvalues — e.g. it separates oblate void regions from walls. The following volume filling fractions are thus recovered:

Volume filling percentages for independent random eigenvalues

feature	signature $(\lambda_1; \lambda_2; \lambda_3; \sum_i \lambda_i)$	Gaussian λ_i relative volume	
nodes	(- ; - ; - ; -)	$\sim$ 12.5%	
filaments	(+ ; - ; - ; -)	$\sim$ 29.4%	total prolate:
prolate	(+ ; - ; - ; +)	$\sim$ 8.1%	37.5%
walls	(+ ; + ; - ; -)	$\sim$ 8.1%	total oblate:
oblate	(+ ; + ; - ; +)	$\sim$ 29.4%	37.5%
expanding	(+ ; + ; + ; +)	$\sim$ 12.5%	

The measured volume filling percentage of oblate void regions can be brought down significantly by adopting a suitable threshold eigenvalue λ_{th} for the web classification. In this method, a region is only judged to be collapsing along a given principal axis if the corresponding eigenvalue exceeds λ_{th}. See 2.8 for detailed information and figure 9 for a visual explanation for the effects of a λ_{th}. Section A presents the effects of a λ_{th} for our web classification.

A very sharp contrast can be seen between the occupation levels of oblate and prolate voids. The likelihood of a void being squeezed in two directions simultaneously is dwarfed by the likelihood of collapse in only one direction. The increase in total space occupied by expanding void regions is in agreement with expectations. Since voids tend to merge, the absolute number of individual voids decreases, but this has no net effect on the volume occupation.

[53]The Nexus+ web classifier is based on the density — it uses a lognormal filter kernel — and is scale independent.

6.5 Density-divergence relations per web component

Density per web component

The individual density distributions from the separated cosmic web components are shown in figure 41. The region types are coded with the colours conform table 5. Firstly, notice the horizontal ranges the distributions occupy at either end of the redshift evolution. The node and filament distributions broaden by great amounts, getting more extreme in both high and low density. The other four regions spread to lower densities, but their upper density bounds change only little. Nodes are invariably overdense at $z = 3.7$, but not so at $z = 0.0$.

These results bear a strong similarity to those in a previous study by Hahn et al. (2007). Their classification method was based on the gravitational potential. Their study defined only four types of regions — it did not distinguish between different types of underdense regions. Their density distributions are plotted in figure 8 — section 2.8.

At $z = 3.7$, while the massive features have long high-density tails, we see a sharper cutoff at the right hand sides of the void distributions. This is of course a result of the conditions under which these void regions were defined. For regions of progressively higher density, the likelihood that a positive sum of eigenvalues is found falls off steeply. The sum of eigenvalues is, after all, closely related to the density — this can be derived from the continuity equation.

Conversely, at $z = 0.0$, all distributions have longer high-density tails, and there is far more overlap between density levels of separate types of regions. To the point that a narrow range of densities around $\delta + 1 \simeq 0.2$ falls within the ranges of all types — even distributions from nodes and expanding void regions overlap. The fact that such low density areas are still classified as nodes on the basis of the eigenvalue signature indicates a limitation of the validity of the classification method. Velocity based web classifiers may unduly identify empty regions as nodes, filaments or walls, if the appropriate velocity flows occur. This means that the massive regions are overrepresented, at the cost of underdensities, which is reflected in the amount of massive features occurring at densities below $\delta + 1 \simeq 1.0$.

The fact that this misrepresentation occurs to a larger degree at low redshift than at high redshift is likely in part due to an actual difference in spatial balance between overdensities and underdensities. The density distribution gradually deviates from its initial Gaussian nature.

An analytical assessment of how likely different eigenvalue signatures are to be found at different densities is provided by Pogosyan et al. (1998). They, too, used the deformation tensor as basis, but considered only the eigenvalues, not their sum. For Gaussian random initial conditions — and the Cosmogrid body at $z = 3.7$ is close to Gaussian — Pogosyan et al. calculate that any region at $\delta = 0$ has a probability 0.5 of having a signature $(- + +)$; characterising a filamentary structure. In the other half of cases, the signature is $(- - +)$, characterising a sheet-like structure. The upper panel of figure 41 reveals that the filaments enter a break-even point with oblate void regions around $\delta \sim 0$. So do the walls and prolate void regions, at redshift 3.7. The measured density distributions for nodes, filaments and walls[54] agree to some extent with Pogosyan

[54]particularly when taking into account the walls contained in the oblate void region measurements

Statistical distribution $\delta + 1$

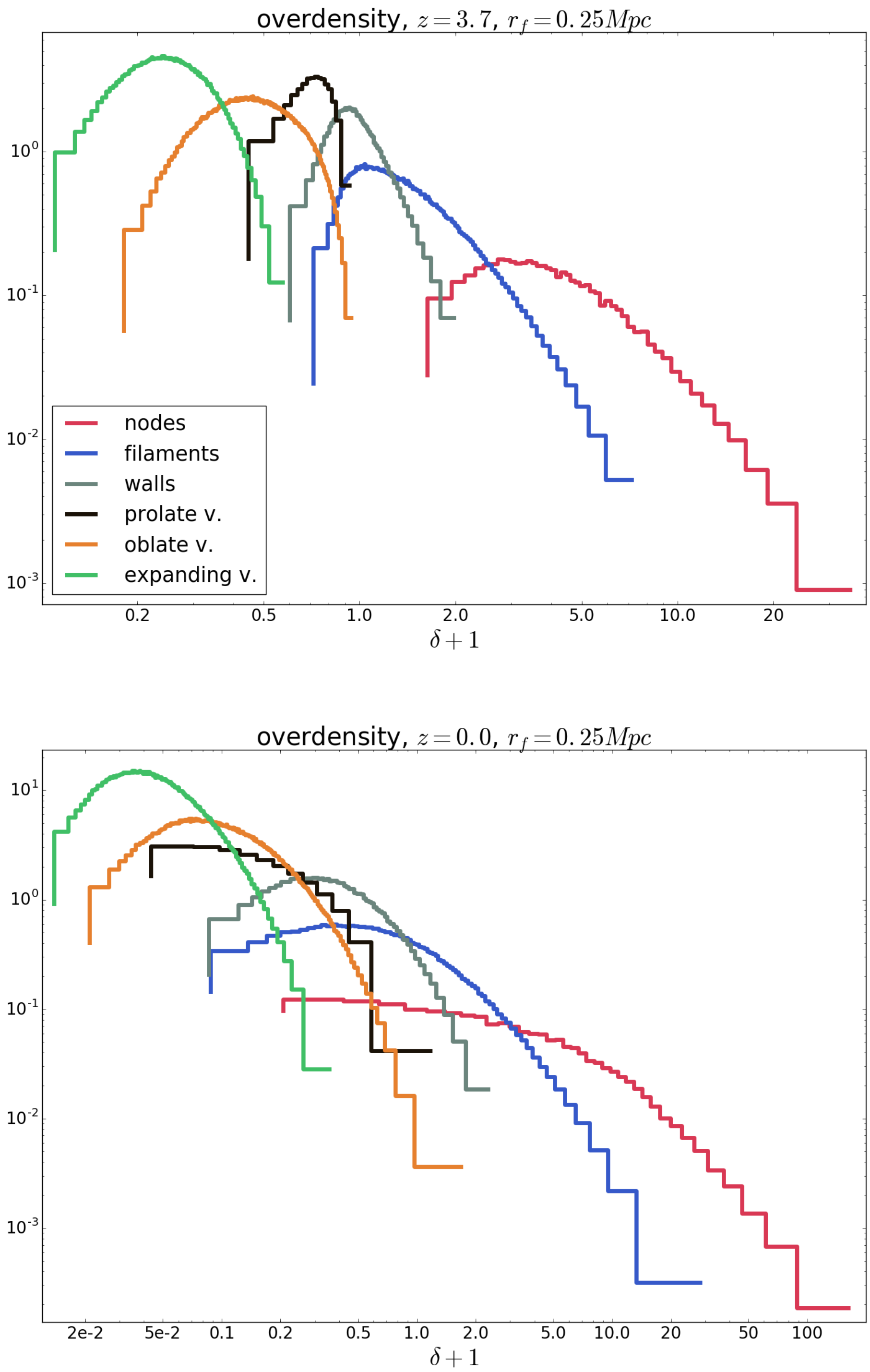


Figure 41: Statistical distributions of overdensity $\delta + 1$ at $z = 3.7$ (top panel) and $z = 0.0$ (lower panel). Different distributions come from regions identified as different cosmic web components, see legend. Notice that the overdensities range between $\sim [0.1; 20]$ at $z = 3.7$ but at $z = 0.0$ they span $\sim [0.01; 100]$.

GRF shear eigenvalue - density probabilities
(Pogosyan et al., 1998)

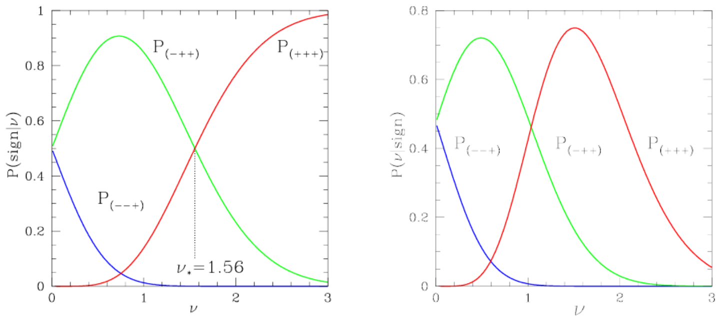


Figure 42: Statistical distributions for eigenvalue signatures, given a density (left); and densities per eigenvalue signature (right). Like in our study, eigenvalues are based on the non-traceless shear tensor. Signatures $(+++)$ characterise nodes, $(-++)$ filaments and $(--+)$ walls. ν is defined as $\nu = \delta/\sigma$, where σ is the standard deviation of the δ distribution. The left panel shows the likelihood of signatures at each density; the right panel displays density distributions. These calculations are done by Pogosyan et al. (1998), they apply to Gaussian random conditions.

et al.'s predictions. See figure 42.

Divergence per web component

The discrepancy between the long tails of the density and divergence distributions is illustrated in figure 43(a). The convergence in the nodes falls off far more steeply than the density. This is mirrored in the expanding void regions, where densities are low in proportion to the divergence. Both density distributions span a broader range than the divergences in corresponding regions. Note that, in absolute terms, the discrepancy between the node distributions is far greater than that between the expanding void distributions.

In other web components, the corresponding density and divergence distributions match more closely. Exampli gratia, both distributions for the filaments are shown in figure 43(b). This panel lifts out the oblate and prolate features. Most striking is the sharp divide at $1 - \frac{\nabla \cdot \mathbf{v}}{Haf} = 1.0$. There is a certain overlap around $\delta + 1 = 1.0$ between the densities of these four regions — see figure 41. But in contrast, no net convergence was found in the voids, nor any net divergence in the massive regions. The absoluteness of this divide is not a physical phenomenon, but a consequence of the method of classification: The web components on either side are distinguished by the sign of the sum of their deformation tensor eigenvalues, which is determined by the divergence.

Figure 43(c) shows the divergence distributions for the oblate and prolate regions at $z = 0.0$. At this stage of structure formation, the greatest divergences in the void regions have grown to the point that the quantity $1 - \frac{\nabla \cdot \mathbf{v}}{Haf}$ spreads into the negative domain. This is a sign that the velocity distribution has become nonlinear. This plot, too, shows a divide between the overdense and underdense regions at unity.

Given the definitions of these four regions — table 4 — certain continuities between the four distributions may be expected: the filaments and prolate voids

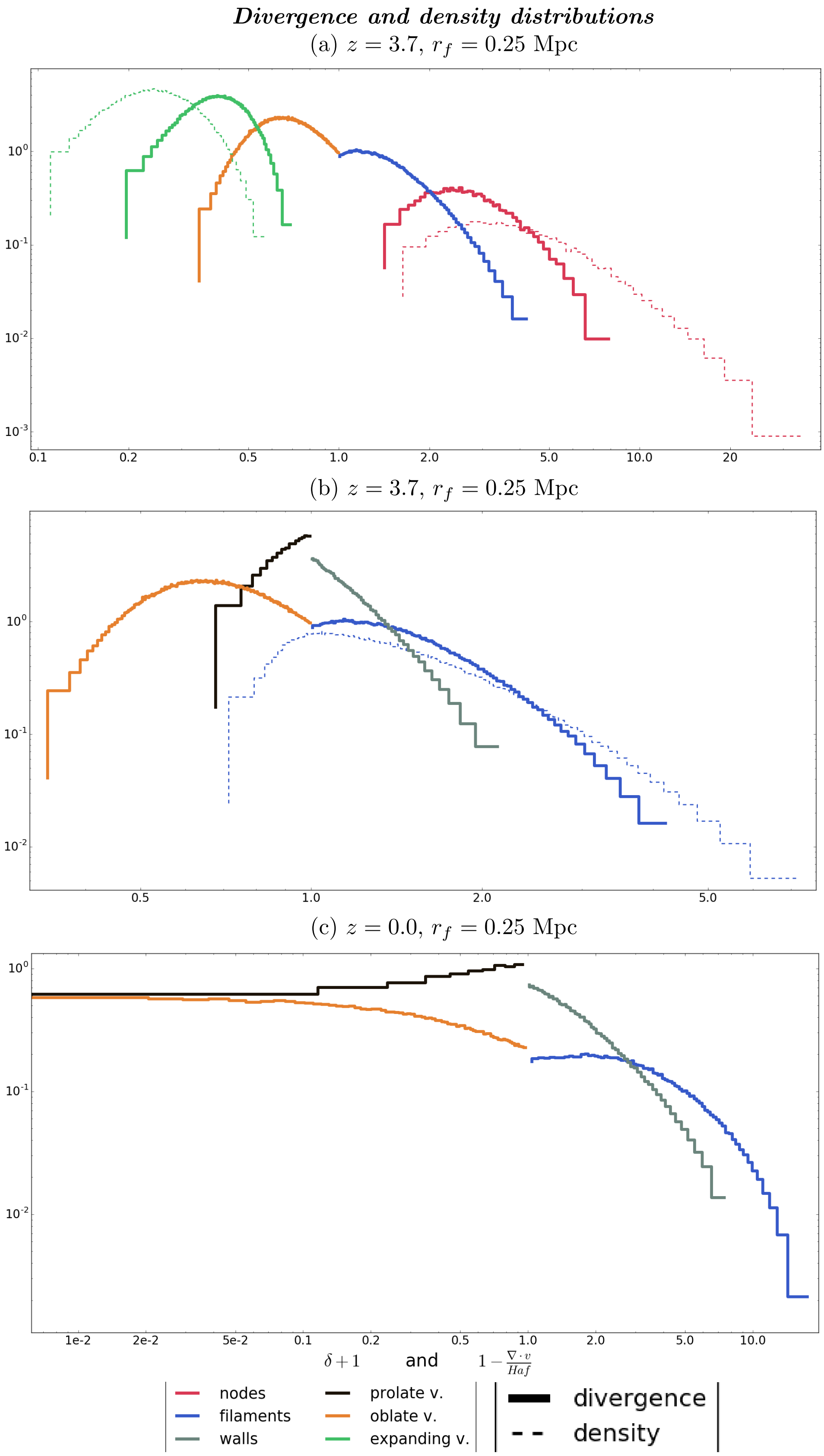


Figure 43: Statistical distributions of density velocity divergence. Solid lines: divergence, dashed lines: density. Colours indicate cosmic web regions corresponding to the legend. Redshifts and filter scales are displayed above panels. Horizontal axes show densities and normalised divergence values, note that horizontal ranges differ between panels.

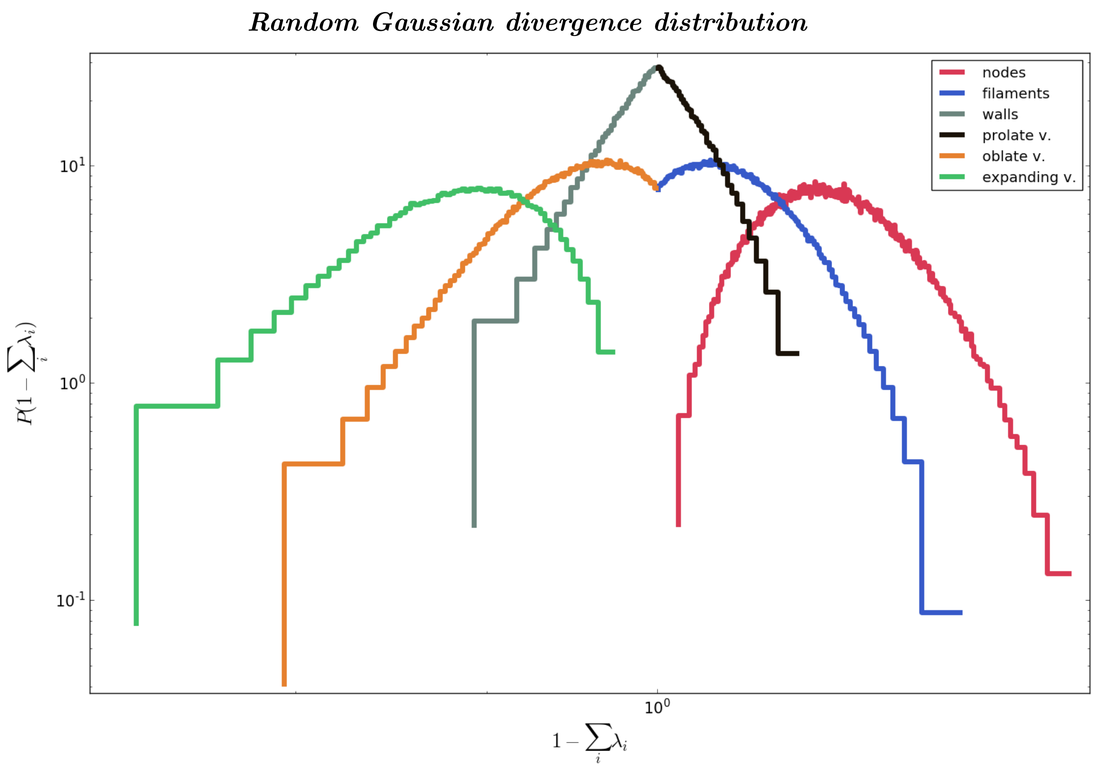


Figure 44: Divergence distributions from a separate numerical experiment. Independent random Gaussian eigenvalues were drawn as a basis for web classification, and divergence $\equiv \sum_i \lambda_i$. The horizontal axis is logarithmic as well.

share the same eigenvalue signature. They differ only in the sign of the *sum*, $\sum_i \lambda_i$, and the same holds for the walls and oblate voids. Interestingly, it turns out to be the distributions for oblate void regions and filaments that lie closely together at $1 - \frac{\nabla\cdot\mathbf{v}}{Haf} = 1.0$; and the walls and prolate void regions meet relatively closely as well. Statistically, walls and prolate voids are among the rarest types of regions, and they span only small ranges in density and divergence. Figures 43(b,c) reveal that they are concentrated around unity.

These results are in agreement with a numerical experiment where independent random Gaussian eigenvalues are drawn for each grid cell. The divergence is defined as the sum of eigenvalues, and the signatures determine cosmic web components conform our web classifier. Figure 44 shows the resulting divergence distributions — compare these to the results in figure 43. The same discontinuity at unity occurs in both cases. As we have seen in the previous exploration of volume filling percentages, the filaments and oblate void regions are far dominant over walls and prolate void regions in Gaussian random conditions. The latter two appear at only narrow ranges of divergence.

Another interesting continuity at unity can be seen in figure 43. Note that the *slopes* for oblate void regions and walls match at unity, and the same holds for prolate void regions and filaments. This can be seen both at $z = 3.7$ and $z = 0.0$. Note that, disregarding the sign of $\sum_i \lambda_i$, oblate voids and walls share the same eigenvalue signature — they are both oblate. Prolate void regions and filaments are both characterised by a prolate eigenvalue signature. Therefore, this continuity makes intuitive sense: considering simply the statistical distribution of a region defined by a specific eigenvalue signature, there would be no grounds to expect a discontinuity in slope at any point. Note that the continuity in slopes is also reproduced by the aforementioned numerical experiment, see figure 44.

Density - divergence relation per web component

Figure 45 shows the correlations between between divergence and density at $z = 3.7$. Corresponding densities and divergences from a uniform random selection of grid cells are plotted as dots, with colours indicating the cosmic web components. The dashed grey line indicates where $1 - \frac{\nabla\cdot\mathbf{v}}{Haf} = \delta + 1$: a perfectly correlated body of data points would lie precisely on this bisector. The overall distribution of the data points indicates that the divergence is well correlated with density.

This plot reflects some previously made observations: the sharp divide at $1 - \frac{\nabla\cdot\mathbf{v}}{Haf} = 1.0$ appears as a sharp horizontal boundary in the density-divergence plane. Also, the highest node densities continue while the convergence in corresponding regions levels off. The correlation in the underdense regions mirrors that in the nodes: the correlation is somewhat shallower than unity, as the density spans a somewhat broader range than the divergence.

Note the boundary between the expanding and oblate collapsing void regions — between the green and yellow loci. The skewness of this line is indicative of the differences between density and divergence distributions for the two types of voids. In the expanding void regions, the two distributions are shifted, while the distributions from oblate void regions match relatively well. The same holds for the boundary between nodes and filaments.

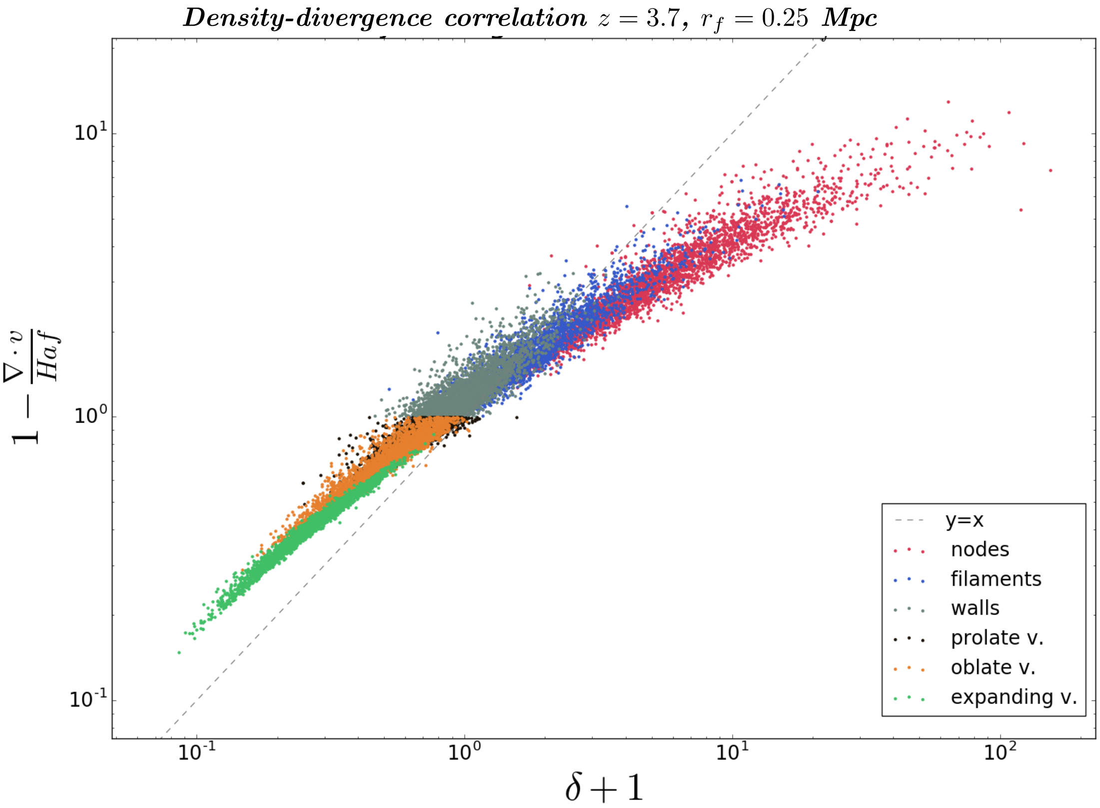


Figure 45: Correlation between overdensity and velocity divergence. Each point is a density-divergence measurement form a uniform-randomly selected grid cell from the simulation at $z = 3.7$. Colours show the identified region types. The dashed grey line indicates where $1 - \frac{\nabla \cdot \mathbf{v}}{Haf} = \delta + 1$, i.e. a perfect unity correlation.

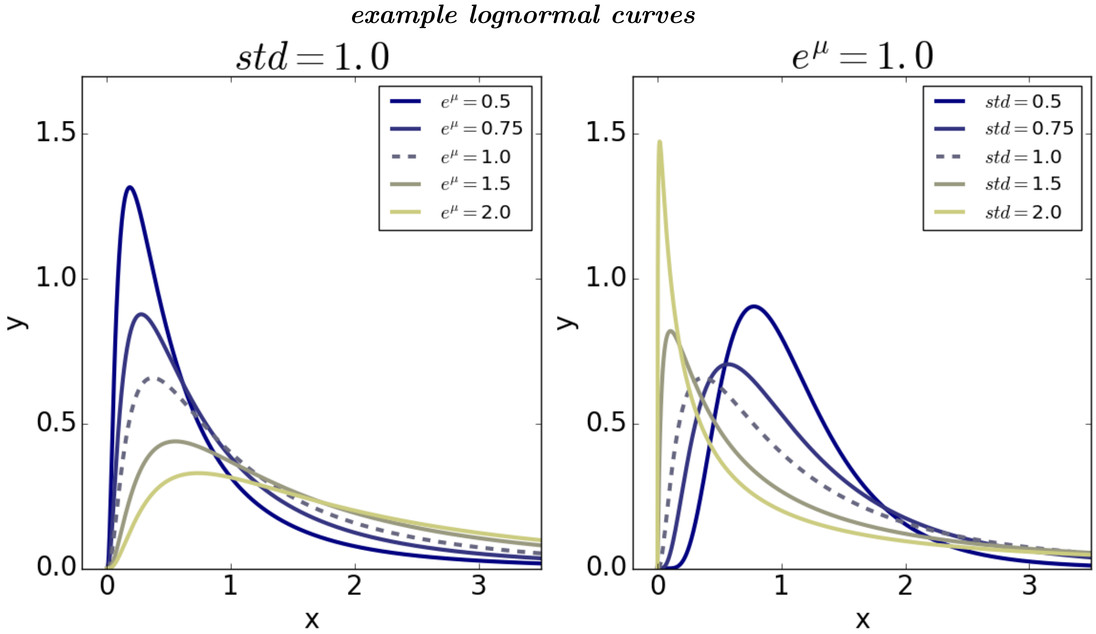


Figure 46: Examples of lognormal curves. In the left panel, all curves have unity standard deviation, scales vary according to the legend. On the right, all curves have unity median, while the standard deviations are varied.

Lognormal fit parameters

As explained in more detail in section 5.7, lognormal curves were fitted to the density and shear magnitude distributions from all cosmic web components, and from all redshifts in the data set. For each distribution, two lognormal fit parameters were recovered: the median e^{μ} and standard deviation std[55]. Figure 46 shows a number of examples of lognormal curves with varying medians and standard deviations.

Figure 47 shows the redshift evolution of the determined fit parameters. The horizontal axis indicates redshift, so that the right portions of the graphs constitute the linear regime, deviations from linearity increase towards the left. This figure shows that density distributions — in all web components — clearly increase in the std parameter over the entire course of structure evolution. This is mirrored by a decrease in medians. For reference, figure 46 provides examples of the consequences for a distribution. Higher standard deviations and lower medians result in a great deviation from a Gaussian appearance.

These deviations from Gaussianity reflect the unbounded growth of the density contrast. While density peaks keep increasing, ever larger volumes only get sparser. Notice that these deviations appear to be more pronounced in voids than in the massive region types. The densities in nodes exhibit a surprisingly

[55]Note that this is not the statistical standard deviation of the distributions, but rather that of the underlying Gaussian distribution that the lognormal curve is defined upon. Similarly, e^{μ} is the median of the lognormal distribution, where μ is the mean of the underlying Gaussian. See section 2.6 for details. Rather than the conventional symbol σ, we denote the standard deviation by std, to prevent confusion with velocity shear.

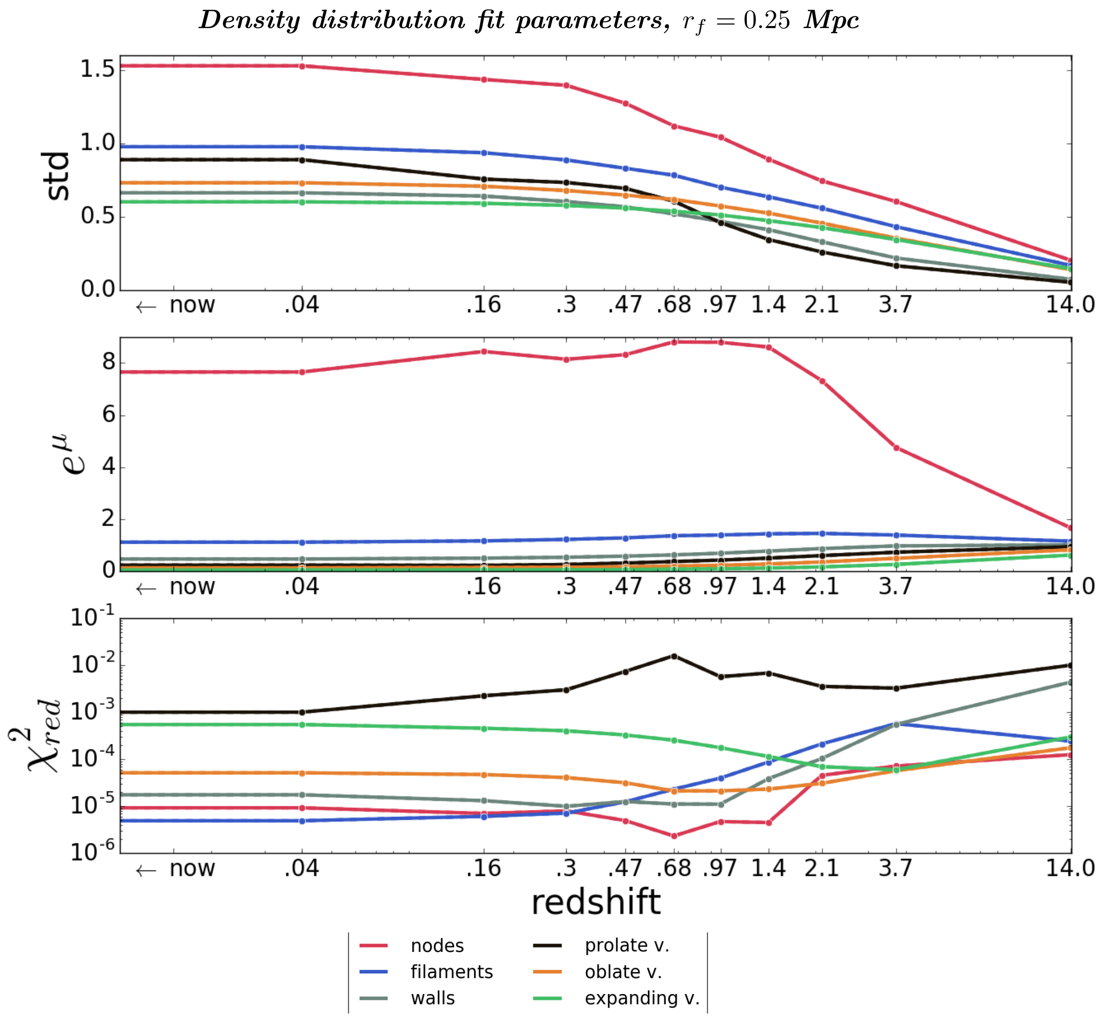


Figure 47: Redshift-evolution of parameters for lognormal fits to density distributions. Top panel: standard deviations, middle panel: medians, lower panel: reduced chi squares. Colours indicate the web components.

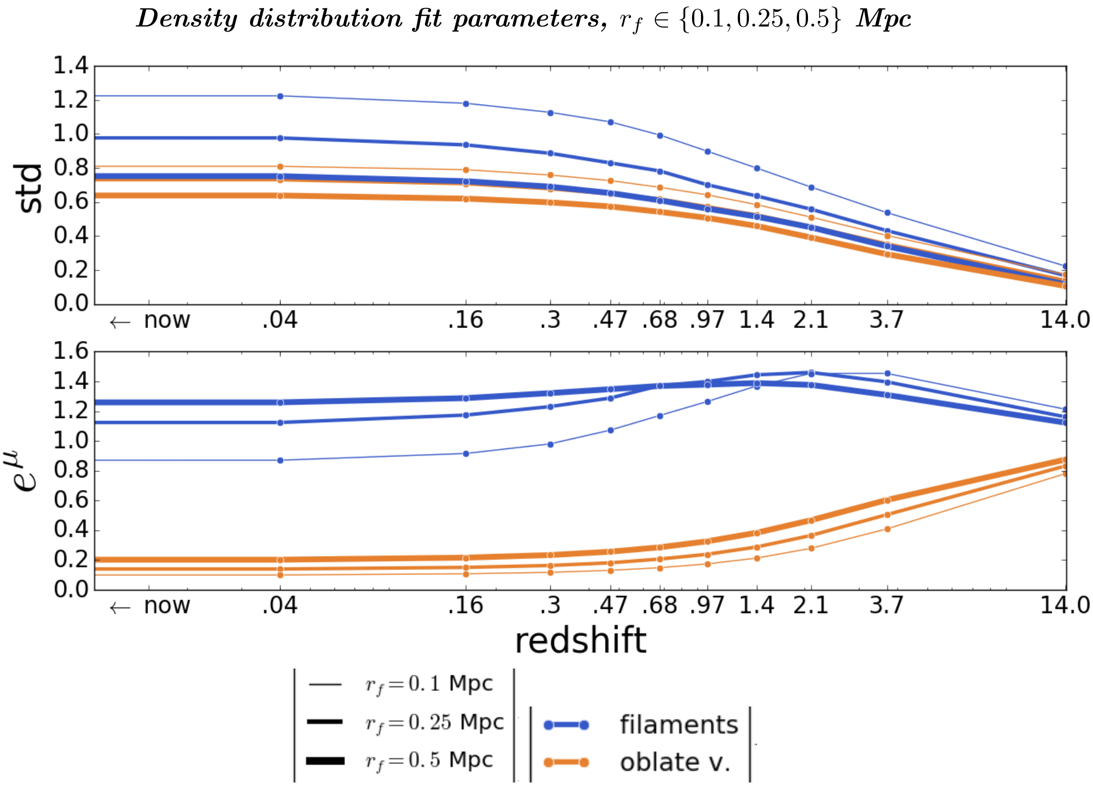


Figure 48: Redshift-evolution of parameters for lognormal fits to density distributions. Only data from filaments and oblate void regions are shown. Line widths indicate the spatial scale at which distributions were taken.

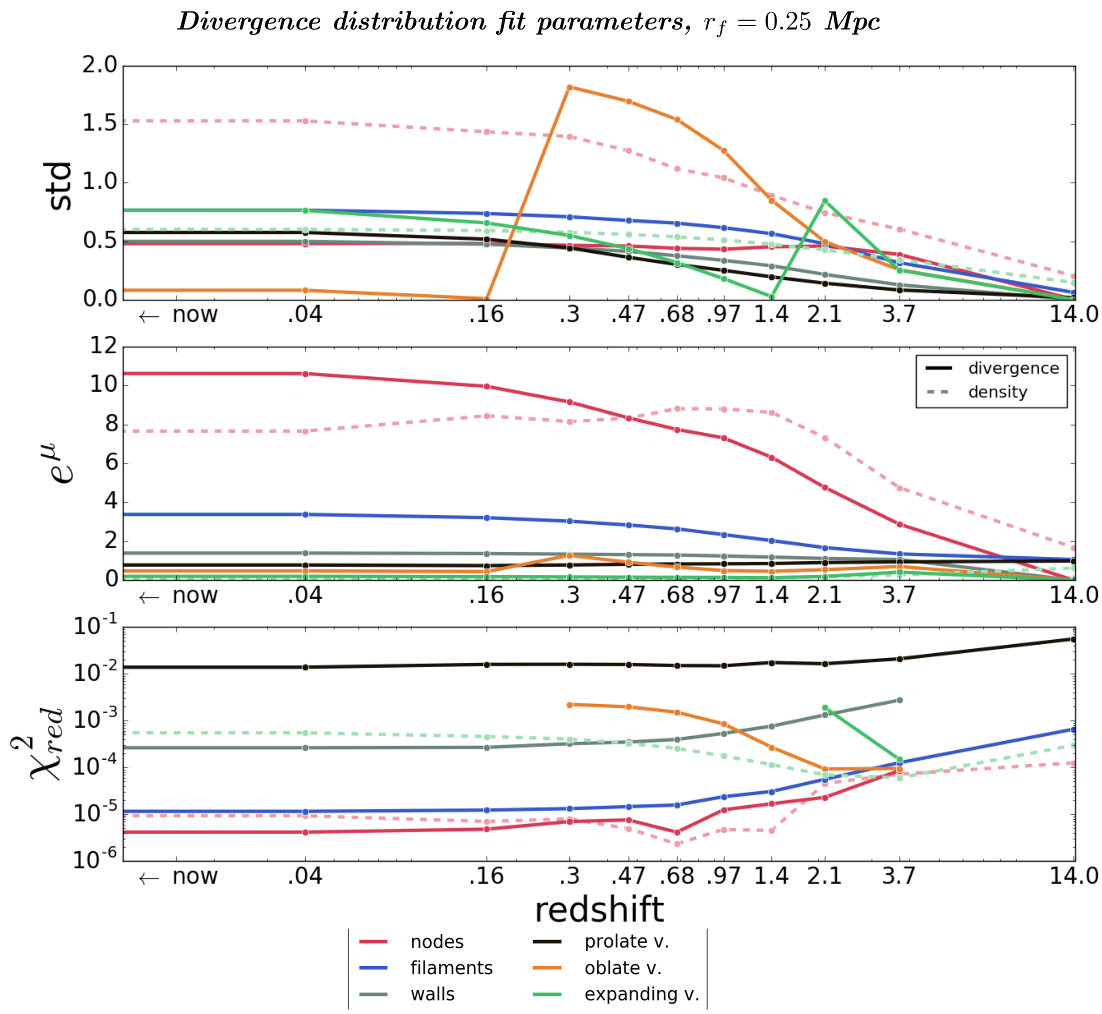


Figure 49: Redshift-evolution of parameters for lognormal fits to divergence distributions (solid lines). Dashed lines indicate the fit parameters to the density distributions from nodes and expanding void regions. Top panel: standard deviations, middle panel: medians, lower panel: reduced chi squares. Colours indicate the web components.

large and early *increase* in medians, contrary to all other regions[56].

All data from figure 47 are filtered at a scale of $r_f = 0.25$ Mpc. Small scale structures are expected to deviate from linearity earlier, and large ones later. Our analysis with different filter scales reproduces this successfully. Figure 48 shows examples of parameter evolutions from different spatial scales. It is clear that the parameters from different scales undergo the same changes, but that small scale distributions evolve earlier, and are followed by those from large scales.

A very important remark must be made about the interpretation of these results. The numbers shown in figure 47 come solely from *lognormal fits* to the distributions — not from the real distributions. Most velocity based quantities are reasonably well described by lognormal curves, but the density distribution less so. For as far as physical nonlinear growth deviates from lognormal distributions, those changes are not visible in this analysis.

The lower panel in figure 47 follows the reduced chi squares of the fits. This lets us know how well a given distribution is described by a lognormal curve. The increase of chi squares towards low redshifts indicates a deviation from lognormal distributions. It appears that filaments adhere to a lognormal distribution relatively well.

Figure 49 shows the results of the same analysis applied to the divergence distributions. The divergence distributions increase in standard deviations, parallel to the densities — and this is in line with expectations from gravitational instability. However, the medians for the divergence fits also increase significantly, contrary to the density fits.

The qualities of fits to the Divergence distributions vary to a great extent. Particularly the voids appear to deviate from expectations. In the nodes, the divergence measurements are generally closer to lognormal than the density measurements are. This reflects the differences in the high-density tails, also seen in figure 43(a).

As with the density data, the changes in the divergence distributions — as probed by lognormal fit parameters — are nearly the same between different spatial scales considered. All changes happen at an earlier time for lower scales, and at a later time for larger scales. This agrees well with expectations.

6.6 Deformation eigenvalues

The deformation tensor was defined in terms of shear and divergence:

$$D_{ij} = \sigma_{ij} + \frac{1}{3}(\nabla \cdot \mathbf{v})\delta_{ij}.$$

It is a symmetric, non-traceless tensor, so that it has three eigenvalues. These are ordered as:

$$\lambda_1 > \lambda_2 > \lambda_3.$$

Section 5.4 provides more detailed information on these quantities. In the present section, we explore the spatial and statistical distributions of the eigen-

[56]There is a very important caveat in interpreting these results, namely that the lognormal fits do not capture the true distributions. Deviations from lognormal distributions occur particularly in the high-density tail — see section 6.2.

values. We also follow their behaviour in different web components, and their relation to density.

Maps

Spatial distributions for the deformation tensor eigenvalues and their sums are shown in figures 50 for $z = 3.7$, and 51 for $z = 0.0$. The same colours are used as in the divergence maps, white areas indicate the zero level. One of the most readily visible aspects of the eigenvalue maps is that the largest eigenvalue, λ_1, exhibits a very deep hierarchy of structure, while the other two are far more homogeneous.

The largest eigenvalue indicates the largest expansion among the three principal axes of any region, and in that capacity it contains information on the anisotropy of structures. Hence, the filaments and walls are well visible in the spatial distribution. Small scale structures, like the colliding outflows from neighbouring voids, are not well visible in these plots. The least anisotropy is expected in nodes and expanding void regions, and it is indeed these regions that appear in lower λ_1 values on all redshifts.

The other two eigenvalues, λ_2 and λ_3, appear to follow the places where the density is highest. Notice that their spatial distributions do not appear to be statistically isotropic: the λ_2 map appears to be more correlated horizontally than vertically, and vica versa for the λ_3 map. This is an artefact that originates in the simulation — no physical quantity on Mpc scales is expected to be more spatially correlated along any specific direction. In this study, the horizontal and vertical alignments are not arbitrary, since the data are produced and stored in a regular grid. The imbalance in spatial correlations is due to the *fundamental mode* problem, addressed in section A.3.

The sum of eigenvalues is a close indicator of density, but is still a quantity derived from velocity derivatives. As such, the summed eigenvalue maps do not show sharp spikes where the density peaks. It is notable that the spurious double spined appearance of walls at $z = 0.0$ is more present both on the λ_1 and $\sum_i \lambda_i$ maps. The λ_1 distribution is expected to be closely related to the shear magnitude, and $\sum_i \lambda_i$ to the divergence. Nonetheless, the divergence maps in figure 32, as well as the shear magnitude maps in figure 57, show a clearer double spine structure than the eigenvalue plots.

Statistical distribution

Figure 52 shows the statistical distribution of all eigenvalues and their sum at $z = 3.7$. Conform our tradition, the plotted quantity is $1 - \frac{(\sum_i)\lambda_i}{Haf}$, this means that the blue line is not exactly the sum of the others — there is a factor 3 missing.

The distributions continue into the negative domain, meaning that the quantities $(\sum_i)\,\lambda_i$ exceed Haf in some places. In fact, λ_1 peaks at $2.982 \cdot Haf$; λ_2 at $2.89 \cdot Haf$; λ_3 at $2.542 \cdot Haf$; and $\sum_i \lambda_i$ at $1.895 \cdot Haf$.

A striking observation in this figure is that, while the eigenvalues sum to a smooth and ordinary-looking distribution, the individual distributions exhibit bumps and dents. Between $1 - \frac{\lambda_i}{Haf} \in \{0.2; 1.0\}$, a large dent occurs in the λ_1 distribution, which appears to be compensated by bumps in the other eigenvalue distributions. This effect persists throughout the redshifts, and at $z = 0.0$ it occurs at slightly higher λ_i values. Its nature is as of yet unclear.

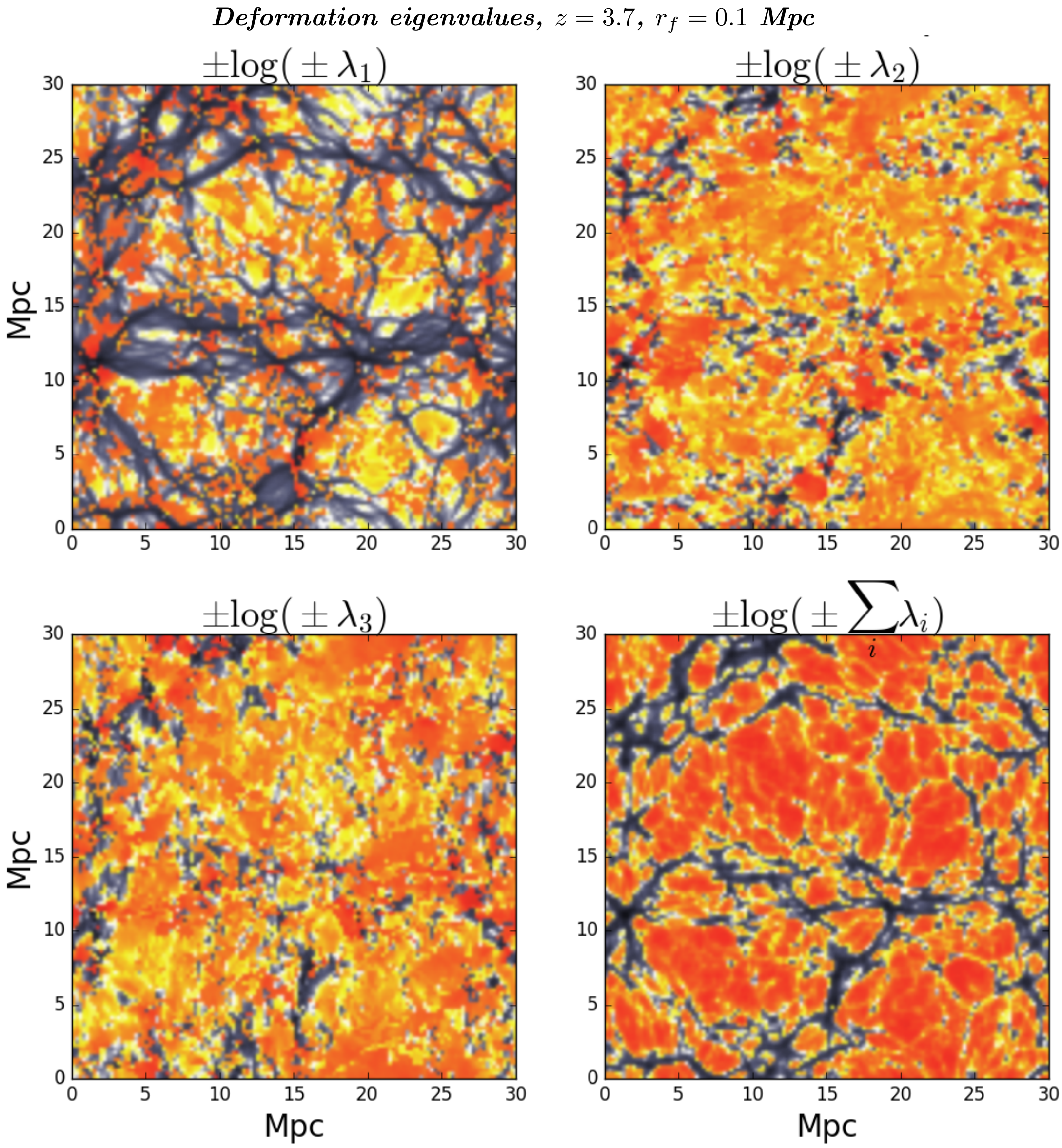


Figure 50: Spatial distribution of the deformation tensor eigenvalues at $z = 3.7$. The eigenvalues are ordered such that $\lambda_1 > \lambda_2 > \lambda_3$ at each grid cell. Positive eigenvalues are marked in orange and indicate expansion along their respective eigenvector direction. Negative eigenvalues — shown in dark shades — indicate contraction. The zero level is marked in white. The horizontal and vertical tendencies in the λ_2 and λ_3 maps are a numerical artefact.

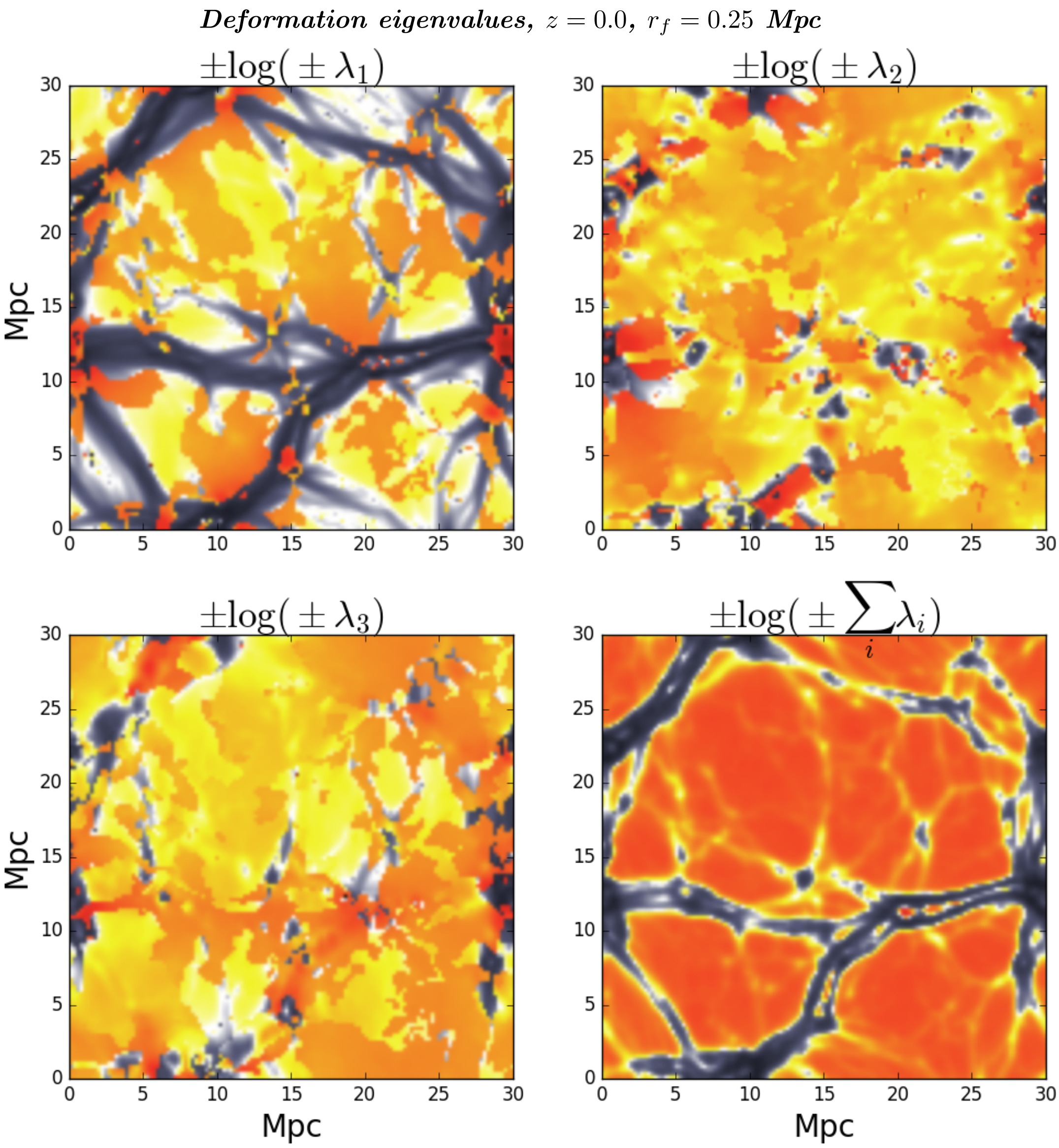


Figure 51: Spatial distribution of the deformation tensor eigenvalues at $z = 0.0$. Layout and colours are the same as in figure 50. The horizontal and vertical tendencies in the λ_2 and λ_3 maps are a numerical artefact, as is the double-spined appearance of walls in the $\sum_i \lambda_i$ map.

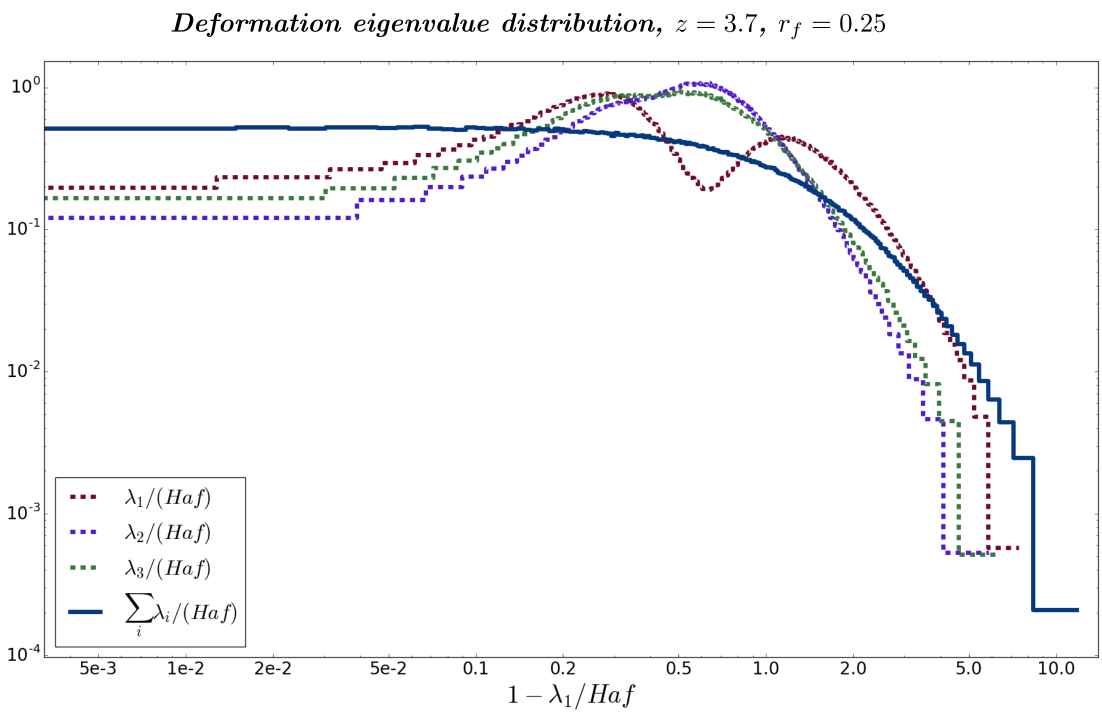


Figure 52: Statistical distributions of λ_1 (dashed, red), λ_2 (dashed, purple) λ_3 (dashed, green) and $\sum_i \lambda_i$ (solid, blue). The dent in λ_1 — and compensating bumps in λ_2, λ_3 — at ~ 0.5 are as of yet unexplained.

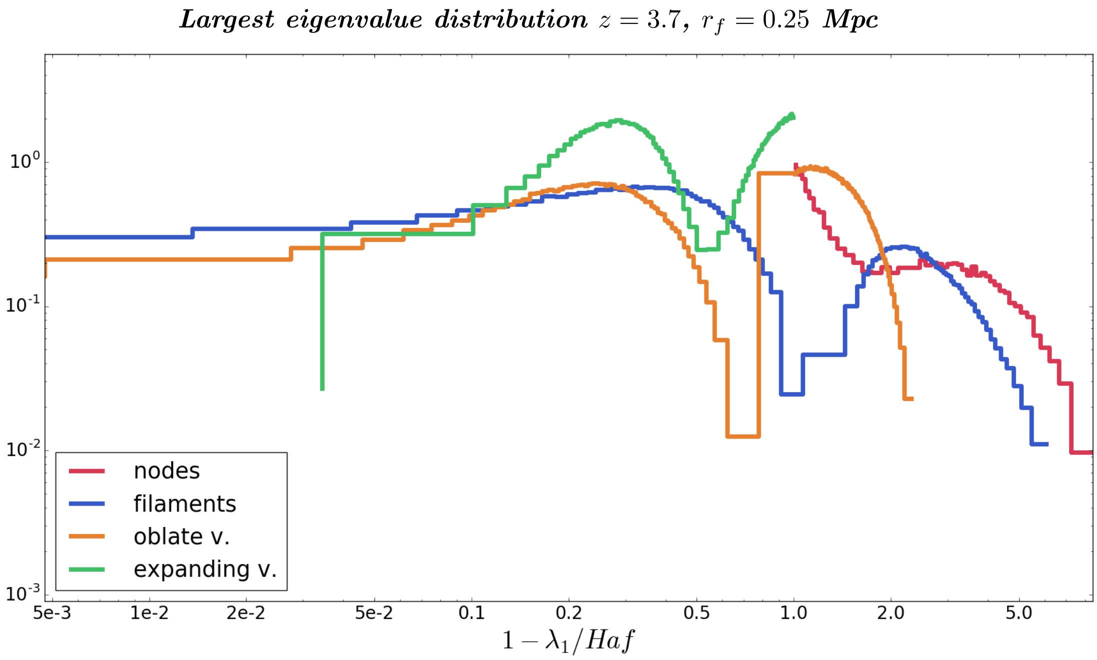


Figure 53: Statistical distribution of λ_1 at $z = 3.7$ for nodes, filaments, oblate and expanding void regions. Distributions from prolate void regions and walls contain areas of very low sampling density, and are thus not shown.

Deformation eigenvalues per web component

Separating the λ_1 values from different cosmic web components — figure 53 — reveals that the dent occurs at lower λ_1 for lower density regions. The node distribution appears to exhibit a decrease up to $\lambda_1 \simeq 3.0$, relative to the $\sum_i \lambda_i$ distribution[57]. Particulary in the oblate void regions and filaments, the λ_1 gap is deep — the sampling density gets very low.

Density-eigenvalue correlations

The correlation between density and summed eigenvalues — plotted in the upper panel of figure 54 — reflects some previously made observations. The densities continue to increase, while the sum of eigenvalue levels off. In the lower left corner we can see the bend towards negative values of $1 - \sum_i \lambda_i/(Haf)$. Interestingly, many readings from the oblate void regions appear to occur at even lower densities than the expanding voids.

The middle panel in figure 54 shows the correlation between density and the maximum eigenvalue λ_1 at $z = 3.7$. Several observations are made:

- Firstly, the slanted strip with a considerably lower sampling density is a manifestation of the dent in λ_1 values, found in figure 52. The slope of this divide on the $\lambda_1 - \delta$ plane indicates a certain density-dependency in this effect. This is in agreement with figure 53. The same density-dependence

[57]Note that $\sum_i \lambda_i = \nabla \cdot \mathbf{v}$, so the distribution of summed eigenvalues is given by figure 43.

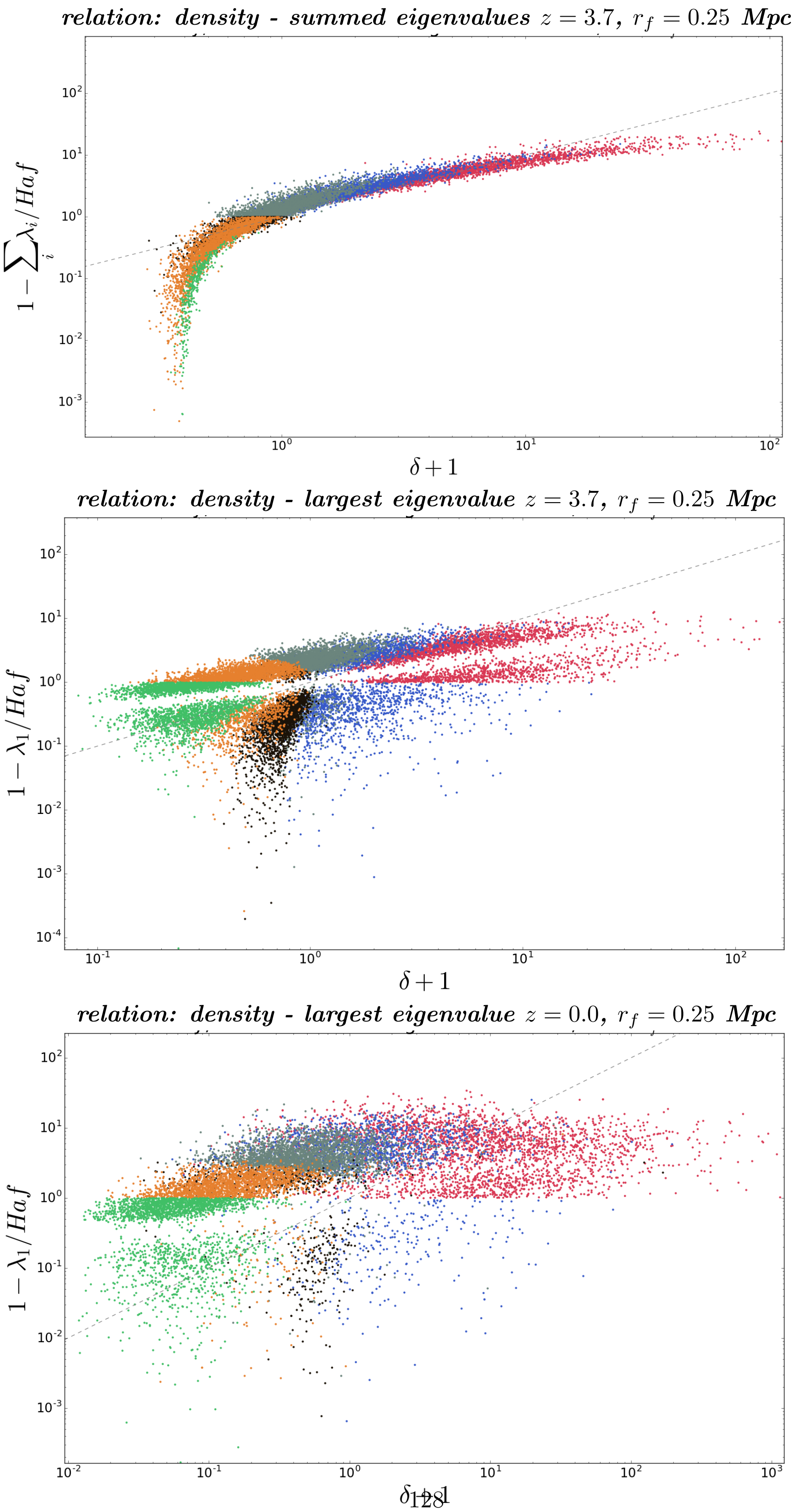


Figure 54: Correlations between density and deformation tensor eigenvalues — λ_1 and $\sum_i \lambda_i$. Data are a uniform random selection of grid cells from the Cosmogrid volume. colours indicate cosmic web region types; dashed grey lines indicate unity correlation. The gaps in the lower two panels are of unknown origin. See text for interpretations.

holds for all web components. It is shallower than $x = y$, and appears to coincide with the correlation of the whole data set.

- Contrary to the divergence-density correlation in figure 45, the boundary between overdense and underdense regions is not clear. For the nodes and expanding void regions, there is still a very sharp horizontal boundary at unity. However, walls, filaments, and collapsing void regions occur on both sides of this boundary. Do note that the oblate features — i.e. walls and oblate void regions — appear to be far more concentrated at lower eigenvalues, and that prolate void regions are more prevalent at high eigenvalues.

- Halfway the upper portion of the plane, we see a clear, linear boundary between void regions and walls. This is a consequence of our classification method, which identifies underdense walls as oblate void regions. Another clear linear boundary between prolate void regions and filaments is found just below it, but at a completely different slope. Similarly, the regions identified as prolate void regions also contain underdense filaments.

The density-maximum eigenvalue correlation for $z = 0.0$ is displayed in the lower panel of figure 54. The most notable difference is that the low sampling band has changed in appearance since redshift 3.7. While still clearly recognisable at low densities, the high density region is populated with readings from nodes. There is only a somewhat ambiguous indication that the 'rift' continues in the same direction, or that it bends towards a horizontal direction.

In general, the fact that some web components obey very clear boundaries in the $\lambda_1 - \delta$ plane, while others do not, may be related to the method of classification. The limitations of a classification based on deformation tensor eigenvalues are explored in sections 2.8 and A. The whole existence of the low sampling rift may be due to a numerical instability.

Eigenvalue ratios

In order to explore the cause of the relative decrease in frequency of λ_1 measurements, the ratios of eigenvalues were compared. Figure 55 displays the statistical distribution of the ratio λ_3/λ_1. This figure is also illustrative of the ratio λ_2/λ_1, which follows a very similar distribution. Both ratios are extremely closely concentrated around zero. In most cases, λ_1 far exceeds the other two eigenvalues. This is expressed in the large volume fraction where the λ_2/λ_1 and λ_3/λ_1 ratios are below unity. It indicates the strongly anisotropic nature of the velocity flow.

Figure 56 displays the ratios λ_2/λ_1 and λ_3/λ_1 at two redshifts. The first quartile of the distribution is shown in black, the second in indigo, the third in orange, and the fourth in bright yellow[58]. It is predominantly at the peripheries of massive regions, where extreme positive and negative spikes in both ratios are found. Such spikes occur if λ_1 is close to zero, which makes it likely that the other eigenvalues are negative. Therefore, there will often be some anisotropic collapse. A comparison with figures 35 and 36 suggests that these spikes typically occur in oblate collapsing void regions.

[58]These quartiles are based on the distribution from the entire Cosmogrid volume. The slice shown in figure 56 is not representative of this volume, and thus has an unequal spatial balance of quartiles.

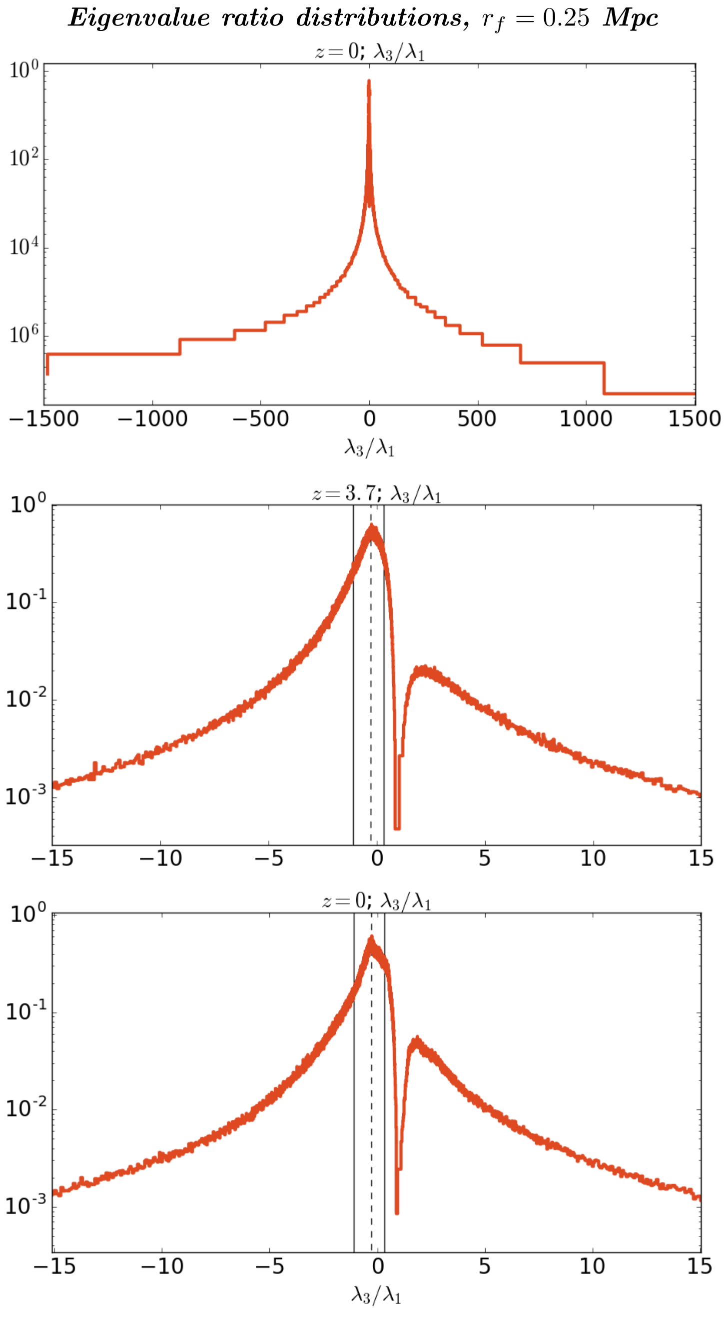


Figure 55: Statistical distributions of the ratio λ_3/λ_1. The top panel displays the distribution over the entire measured range — The ratios range from $\sim -10^7$ to $\sim 10^6$, this figure shows a fixed-weight histogram. The other two panels zoom in on a narrow range, at redshifts 3.7 and 0.0. In the lower two panels, vertical lines indicate the boundaries between the four quartiles of the distribution. The statistical distribution of the ratio λ_2/λ_1 assumes a closely similar appearance to the one shown here.

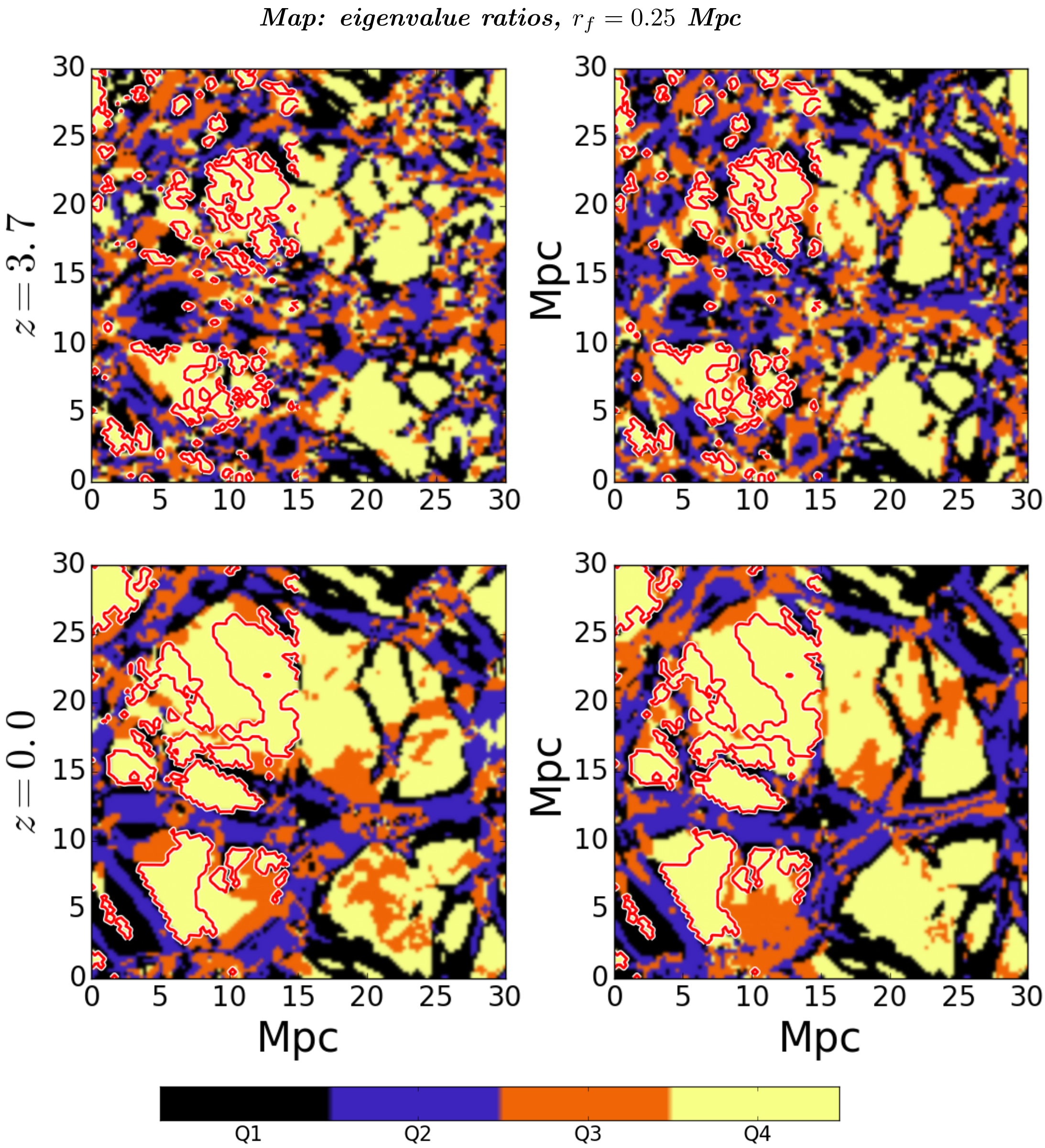


Figure 56: Spatial distributions of λ_2/λ_1 (left) and λ_3/λ_1 (right) at $z = 3.7$ (top) and $z = 0.0$ (bottom). As visualised in the colour bar, the first quartile of the distribution is shown in black, the second in indigo, the third in orange and the fourth in bright yellow. As can be seen in figure 55, the second and third quartiles occupy an extremely small range of ratios, centred around zero, while the first and fourth spread to extreme negative and positive values. In the left half of each panel, contours mark unity ratios. Note that $\lambda_3/\lambda_1 = 1$ is fully isotropic; and $\lambda_2/\lambda_1 = 1$ indicates either planar or full isotropy.

In the left half of each panel, contours mark the unity ratio. Physically, $\lambda_3 = \lambda_1$ implies that $\lambda_2 = \lambda_1$ also. This constitutes a completely isotropic signature — be it isotropically expanding or collapsing. More generally, $\lambda_2 = \lambda_1$ only implies an isotropic planar expansion or collapse — i.e. isotropy in the plane defined by the eigenvectors corresponding to the first and second eigenvalues. Since $\lambda_3 \leq \lambda_2$, two cases can be distinguished. If $\lambda_1 = \lambda_2 = \lambda_3$, this constitutes a fully isotropic signature. However, $\lambda_1 = \lambda_2 > \lambda_3$, is a specially *symmetric oblate* signature: a larger and equal expansion — or a slower collapse — in two directions, and a smaller expansion or faster collapse in the remaining direction. Among all places where $\lambda_2/\lambda_1 = 1$, the majority also has $\lambda_3/\lambda_1 = 1$. This means that fully isotropic signatures are detected far more often than symmetric oblate ones.

These forms of *complete* and *planar isotropy* occur mostly within underdense regions. This satisfies the expectations of relatively isotropic potential and velocity fields in isotropic spatial structures. Locally isotropic velocity flows are expected only in regions with a uniform density, and this occurs in the interiors of voids.

It is important to note that the velocity components incorporated in the deformation tensor eigenvalues do not include any contributions from pure bulk flow. Bulk flow is the zeroth order contribution to a flow field: a constant velocity field over an extended region. It is thus distinctly anisotropic, but not represented in the deformation tensor — i.e. the velocity gradient. This means that isotropic velocity gradient signatures can still be detected in areas dominated by bulk flows.

6.7 Shear magnitude

Velocity shear is defined as follows:

$$\sigma_{ij} = \frac{1}{2}(\partial_j v_i + \partial_i v_j) - \frac{1}{3}(\nabla \cdot \mathbf{v})\delta_{ij}. \tag{145}$$

It is also called the *symmetric part* of the velocity gradient $\partial_i v_j$. Shear measures anisotropy in potential velocity flows, and as such is an important aspect of the formation and evolution of anisotropic structures. This subsection explores the spatial and statistical distributions of the shear magnitude $|\sigma|$, which is defined as:

$$|\sigma| = \sqrt{\sum_i \sum_j \sigma_{ij}^2}. \tag{146}$$

This is still a linear combination of velocity gradients, and as such it is expected to scale with Haf during its linear growth.

This section also explores the behaviour of shear magnitude distributions from different web components, and its statistical relation to density. In an attempt at investigating the nonlinear evolution of shear magnitude, the lognormal fit parameters of shear magnitude distributions are compared to those of density distributions.

Maps

Since we have found the shear magnitude to be a very adequate indicator of structure formation, figure 57 follows the spatial distribution of shear magnitude

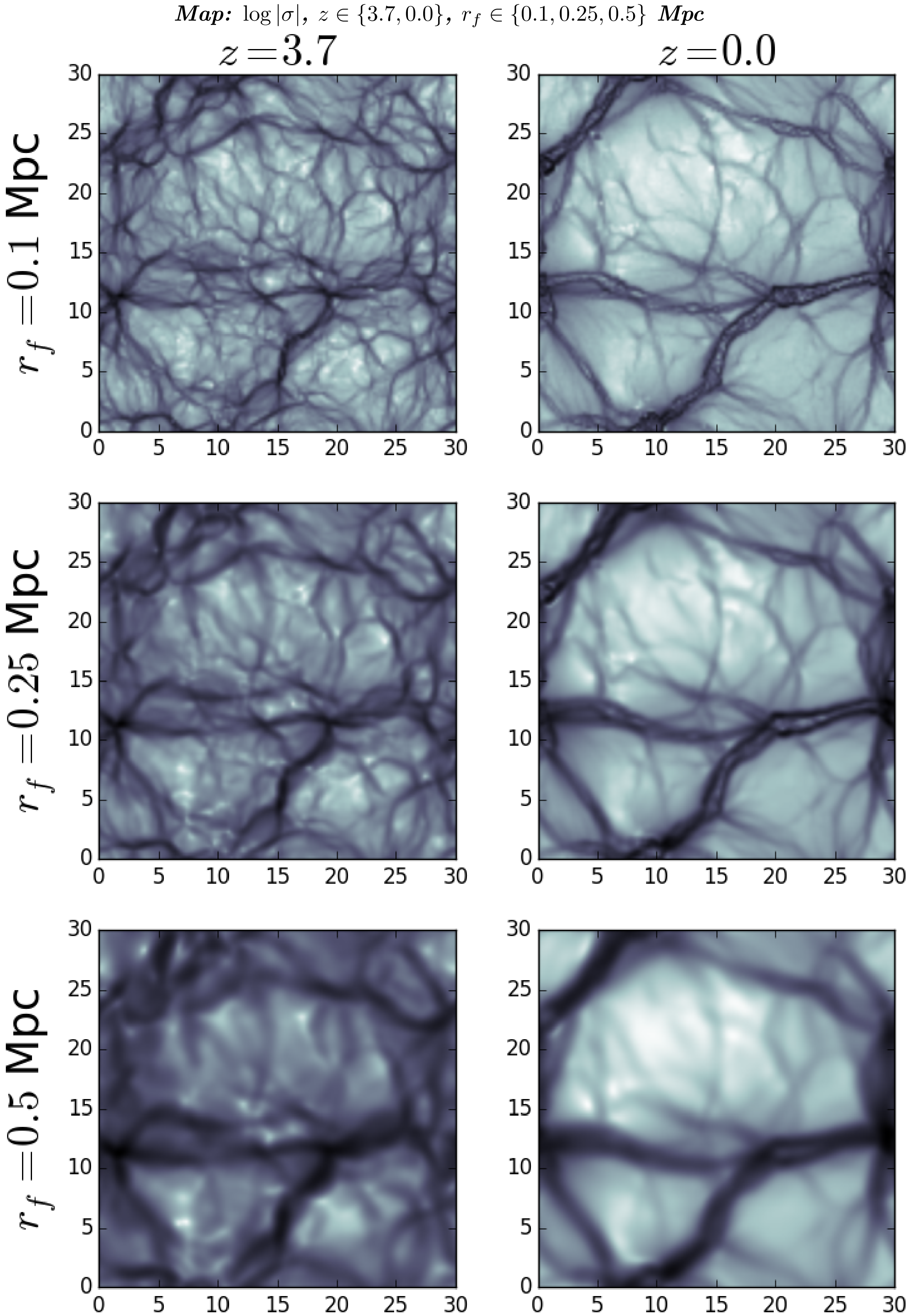


Figure 57: Spatial distribution of logarithmic velocity shear magnitude at $z = 3.7$ (left column) and $z = 0.0$ (right column). Maps in the top row are filtered at 0.1 Mpc, those in the middle at 0.25 Mpc, and those at the bottom at 0.5 Mpc. Bright regions indicate low $\log|\sigma|$, dark regions indicate high shear magnitude. The main conclusion drawn from this figure is that small scale structures — opposed to large ones — evolve by a greater number of individual mergers and relatively large displacements of structures in terms of their characteristic size. See text.

at $z = 3.7$ and $z = 0.0$ at three different filter scales. This allows us to assess the evolution of structure at three different scales in parallel.

Firstly, the spatial distribution shows a wealth of structure on various scales. The formation and evolution of anisotropic structures generally involves shearing of velocity flows, and so these maps highlight the regions of activity in anisotropic structure formation. This includes the shearing inflow of matter into filaments, walls and nodes. The merging and squeezing of voids also generates velocity shear, as matter is drained from the sheet-like region between voids — see e.g. Dubinski et al. (1993). This explains why shear flows show up in low-density areas as well.

The shear magnitude maps at $z = 0.0$ clearly show the unrealistic double spined appearance of massive, large scale walls[59]. At the edges of these regions, we see parallel strips where high shear values are measured. In part, these measurements come from the low density environments around walls. Here, velocity flows are bent towards the alignment with the walls, due to the anisotropic distribution of mass and gravity.

By comparing maps from different redshifts and filter scales, we gather insights in the evolution of structures from different scales. The small scale structure undergoes the greatest metamorphosis, while large scale structures remain relatively stable. At the largest scales, we witness the merging of structures, but hardly any displacement of structures over large scales. At smaller scales, we see more extensive changes. The structures are smaller, less massive, and more numerous; therefore, their evolution involves many mergers. In these mergers, structures get displaced over distances that are large in terms of the typical structure size.

Statistical distribution

In figure 58, the statistical distributions of shear magnitude at two different redshifts are compared. The shear magnitude distributions are displayed along with the corresponding density distributions. The distributions from $z = 3.7$ deviate mostly because of the high-density tail — explored in section 6.2. The distributions at $z = 0.0$ show very high shear magnitude measurements, relative to the density distribution. Recall that the comparison between these distributions rests on a normalisation — $\frac{}{Haf}$ — derived from the *linearised* continuity equation. The visual discrepancy between the distributions from $z = 0.0$ can in part be attributed to a nonlinear shear evolution.

In a comparison solely based on the shapes of the distributions — regardless of their position and horizontal extent — the distributions appear to be a rather well matched. At redshift 3.7, both distributions appear relatively close to a lognormal curve; at redshift 0.0 both have been stretched and acquired a more shallowly decreasing high-value tail. For as far as these distribution shapes represent realistic physical distributions of density and shear magnitude, the agreement between them suggests that a proper normalisation for the nonlinear shear growth does exist.

Shear magnitude per web component

It must be noted that the classification of web components depends largely on

[59] Recall that these have been observed in divergence and deformation eigenvalues as well, and are an artefact due to a poor performance of the DTFE estimator in multi-stream regions.

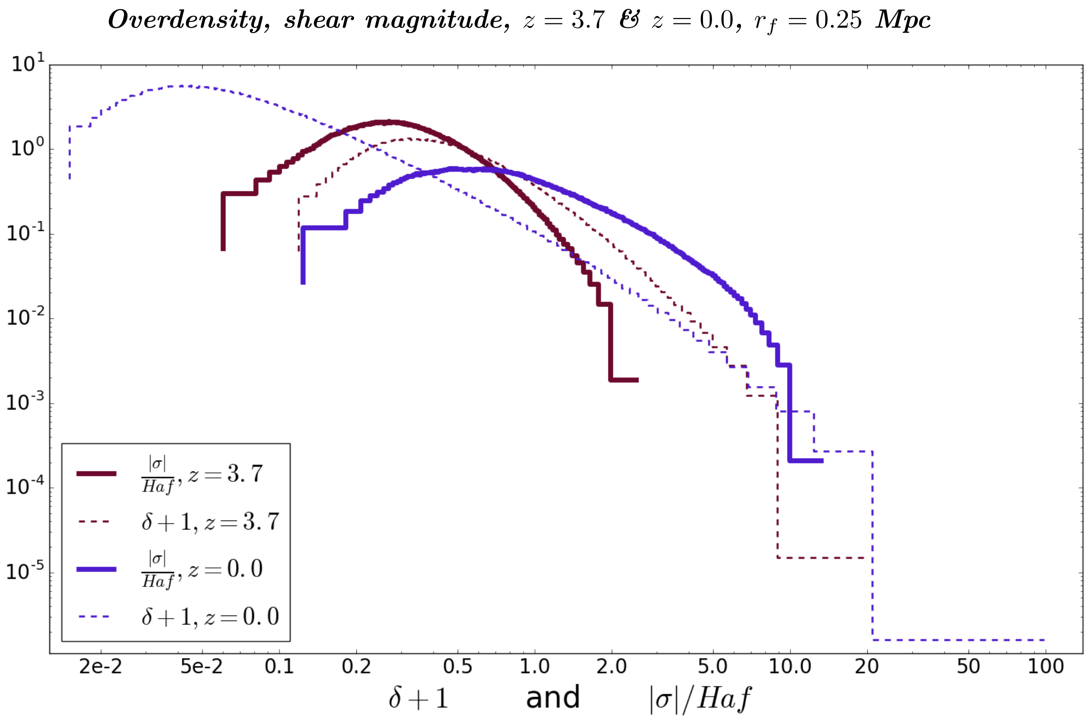


Figure 58: Statistical distributions of overdensity (dashed lines) and shear magnitude (solid lines), at redshifts 3.7 (in red) and 0.0 (in purple). The horizontal axis indicates both density and shear magnitude. The latter is normalised under the assumption of linear velocity growth, which does not hold at $z = 0.0$.

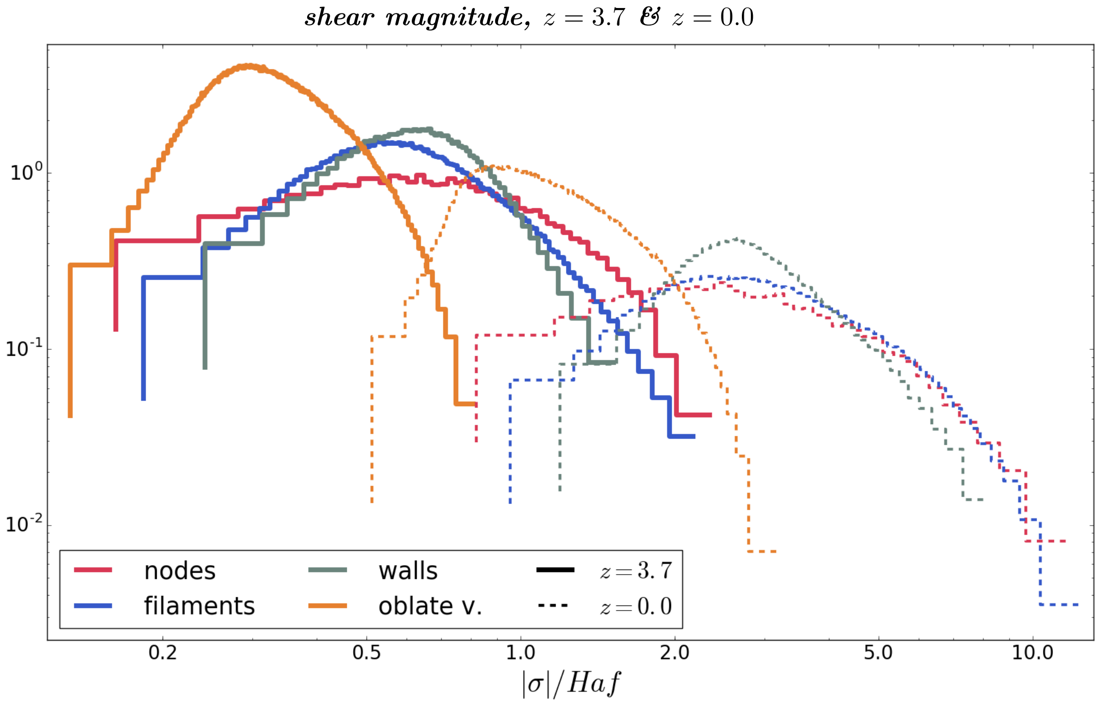


Figure 59: Statistical distributions of velocity shear from different cosmic web components at $z = 3.7$ (solid lines) and $z = 0.0$ (dashed lines). Colours indicate cosmic web region types — expanding and prolate void regions are not shown.

the velocity shear. This means that studying the shear in these regions is a 'circular experiment' that should be of limited value. However, the web classification is also dependent on divergence; and the study of shear magnitude has yielded some interesting observations.

The shear magnitude distributions from separate cosmic web components at redshifts 3.7 and 0.0 are displayed in figure 59. It shows distributions from the two redshifts that bookend the structure formation up to the present. The oblate voids, filaments and walls are the most important regions of anisotropic structure formation.

While figure 58 shows the overall shear magnitude distribution at redshift 0.0 to be enhanced at the high value tail, this seems to hold for none of the cosmic web components individually. Notice, however, that the highest shear magnitude values measured in walls, filaments and nodes continue towards $|\sigma| \simeq 10.0 \cdot Haf$. It can be seen that shear values increase significantly over the course of structure formation in all of the shown web components.

In velocity shear — other than e.g. in density or velocity divergence — it is not only the nodes where the highest peaks are measured: The enhancement of the high shear magnitude tail is due to shear occurring in walls and filaments. The importance of walls and filaments in the shear magnitude distributions is in agreement with the picture of velocity flows approaching *anisotropic* structures. Particularly in the case of walls — which are very prominent on the shear maps — it can be seen that the shear magnitude is centred around $|\sigma| \sim 2.1 \cdot Haf$ at $z = 0.0$. This constitutes a clear deviation from lognormal distribution.

The shear magnitude distributions from oblate void regions deviate from lognormal, by an apparent enhancement[60]. It is intuitive to attribute the increased frequency of these measurements to the scenarios of shear generation in void collisions. The distributions assume the same shapes at filter scales of 0.1 and 0.5 Mpc.

Notice that, in contrast to velocity divergence — but similar to density — there is a large overlap between shear magnitude values measured in different cosmic web components.

Density-shear magnitude correlations

The fact that sharp divisions between different web components occur neither in density nor in shear magnitude is clear in figure 60, where the correlation between the two is shown. The measurements from different web components form a clearly linear chain in the density-shear magnitude plane. However, within each individual component, no correlation can be seen. We see no *intra*-component correlation, but there is an *inter*-component correlation that is somewhat forced, since the components are in part defined by their shear flows.

Notice that the massive features occupy roughly the same range of shear magnitudes, and are differentiated between mainly by density. This indicates the relative importance of shear in walls and filaments.

It is important to note that the normalisation of $|\sigma|$ is not motivated by any full theory of nonlinear velocity growth. At redshift 0.0 and scales of 0.25 Mpc, the velocity shear has become nonlinear. This means that the vertical axis is arbitrary to some extent — a different normalisation would result in the vertical elevation or lowering of the data points. Figure 60 shows most data points to

[60] around $|\sigma| \sim 0.3 \cdot Haf$ at $z = 3.7$ and around $|\sigma| \sim 0.8 \cdot Haf$ at $z = 0.0$.

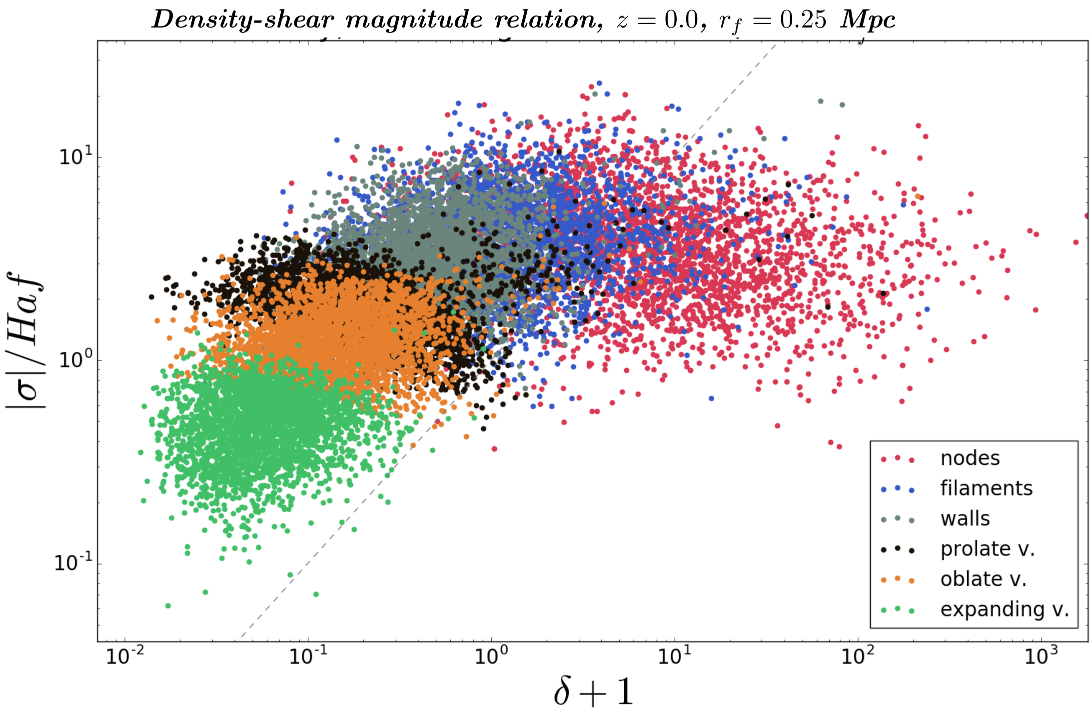


Figure 60: Correlation between density and velocity shear at $z = 0.0$. Colours indicate cosmic web components. Dashed grey line indicates unity correlation.

lie above the $y = x$ line. This suggests that the nonlinear velocity shear does not scale with $Haf(\Omega)$, but accelerates somewhat.

Lognormal fit parameters

As with the density and divergence distributions — section 6.5 — lognormal curves were fit to the shear magnitude distributions. Figure 61 shows the redshift-evolution of the fit parameters and the chi squares. The lognormal curve is determined by a standard deviation *std* and a median e^{μ}. Section 2.6 gives a deeper explanation, and figure 46 gives a few examples.

The shear magnitude distributions undergo only little change in standard deviations, but their medians increase considerably. This means a shift in balance towards high shear measurements; which reflects the importance of shear as an aspect of structure formation. The changes in fit parameters from different web components mostly run in parallel. This suggests that they are consequences of the same processes — i.e. gravitational instability.

The fit parameters from different scales undergo largely the same changes, but these changes consistently occur earlier for smaller scales and later for larger scales. This is in line with expectations for a hierarchical evolution of structures.

As mentioned before and discussed in section A, the redshift evolution of fit parameters yields only limited information on the evolution of these quantities. Deviations from lognormal distributions cannot be probed with this method. The lower panel in figure 61 shows the reduced chi squares of the lognormal fits. In general, the shear magnitude distributions are rather well described by log-

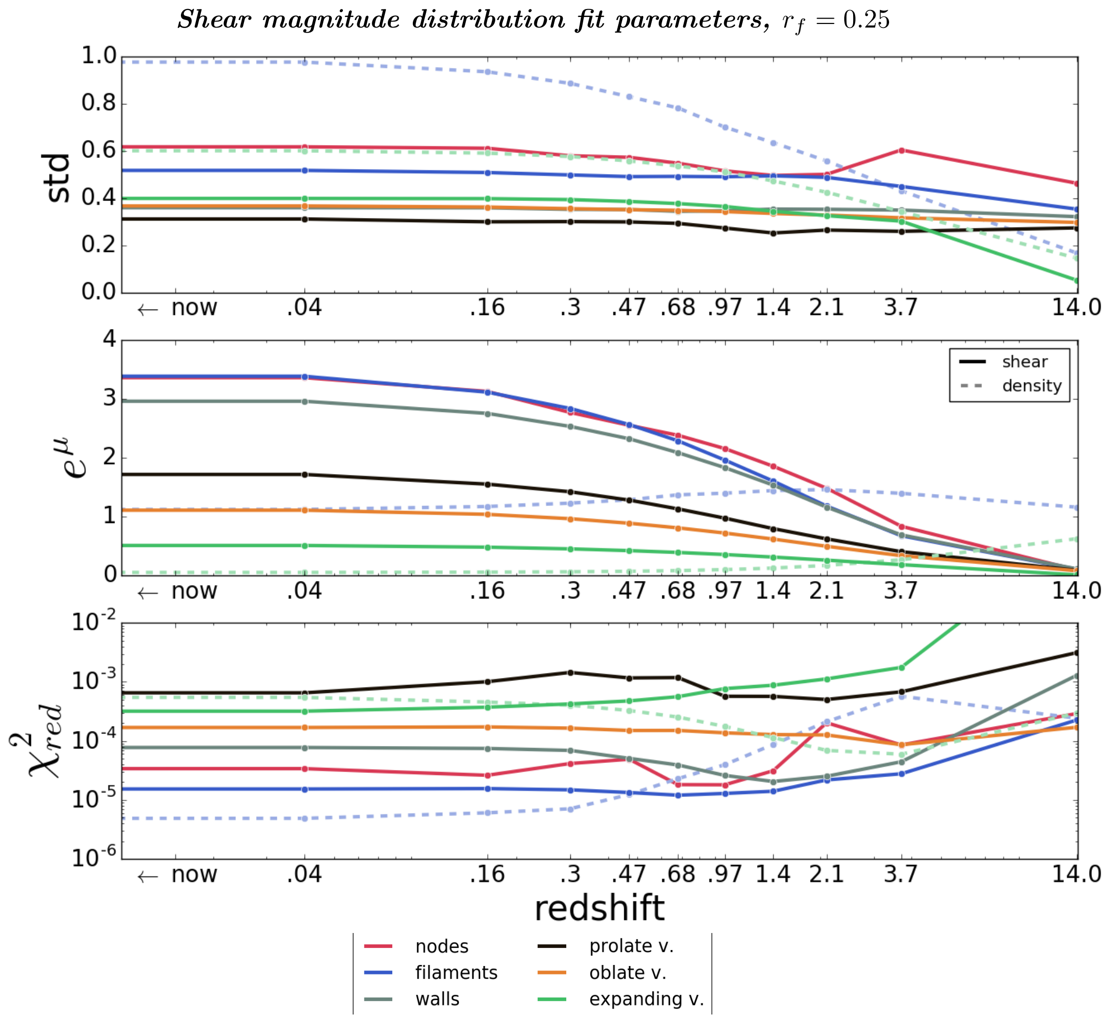


Figure 61: Redshift evolution of lognormal fit parameters to shear distributions (solid lines). Dashed lines relate to fits to density distributions from filaments and expanding void regions. Colours indicate web components. Horizontal axis: logarithmic redshift. Top panel: standard deviations, middle panel: medians e^{μ}, lower panel: reduced chi squares. All data are filtered at $r_f = 0.25$ Mpc.

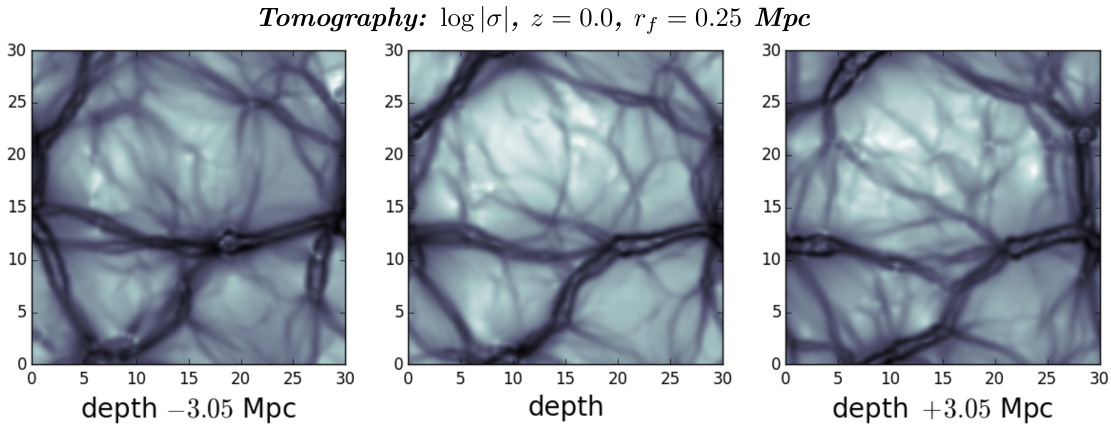


Figure 62: Partial tomography of the Cosmogrid volume in logarithmic shear magnitude at $z = 0.0$. Only three parallel slices are shown, spaced 3.05 Mpc apart, and centred on the same level as all other plots in this section so far. Notice that the large double spined features persist throughout these depth levels, while the thin filamentary features undergo greater variations.

normal curves. In line with previous findings from section 6.5, it is the massive features that match a lognormal distribution best.

6.8 Walls

Tomography

In previous subsections, we have seen — at low redshifts — a number of large, elongated features, appearing to consist of two parallel spines. Recall that these double spines are an artefact, because shell crossing occurs there, and the DTFE estimator unduly mixes several velocity flows. One way to demonstrate that these features are walls, is by *tomography*. The three-dimensional shape of structures is probed by considering successive slices of the cosmological volume, and studying the differences and similarities between features at nearby depth levels.

Density, divergence, deformation eigenvalue and shear magnitude maps were made of all slices in the Cosmogrid volume. When strung together into a moving `.gif` image, the speed at which structures appear to vary in the tomography is an indicator of the scale of their variations in the depth direction. Anisotropic features that appear, move and disappear relatively quickly span a limited range of depths, and are thus concluded to be filamentary in nature. Anisotropic features that appear, move and disappear more slowly are more stretched out in depth, and are thus sheet-like in nature.

In the tomographies, it was clearly the double spined features that moved slowly, and are thus identifiable as walls. Among the smaller, single spined features, most of them emerged and vanished more rapidly throughout the depth levels. The `.gif` files will be available at `www.astro.rug.nl/~mast/probes_and_agents`, and can be requested via `mast@astro.rug.nl`.

Figure 62 shows the shear magnitude maps from three levels. The map in the middle shows the same slice as all the previous maps in this chapter; the other two maps are situated at ~ 3.05 Mpc from the middle map on either side. In comparing the maps from three depth levels, it becomes apparent that some structures undergo more variation than others. Most notable are the double spined features, which remain present throughout all three maps and undergo only small changes in position and orientation. Spanning at least 6.1 Mpc in depth, they are classified as walls. This is in agreement with the cosmic web decomposition — section 6.4 and figure 36 — where these features are identified as walls.

Just another brick...

The slice of the Cosmogrid volume that has been mapped out many times so far is intersected by a few walls, but contains no face-on view on any of them. In order to explore the internal structure of a wall, maps from a different slice are on display in figure 63, in density, divergence and shear magnitude. These maps are filtered at 0.1 Mpc, but the purple contours — delineating the massive features — are filtered at 0.25 Mpc.

In all three quantities, we can see an intriguing substructure of filaments embedded in the face of the wall. In the density map, we see a few small spikes. Some filaments appear to form the edges of the sheet, and a rather fine network of filaments is spun between them. The divergence map indicates a rather homogeneous velocity convergence throughout the wall, interrupted by a few moderately sized centroids of positive divergence. The shear magnitude map, on top of a noisy small scale background, shows an irregular structure of thick, curvy features. These appear to be closely coupled to the regions of high velocity convergence.

Note that the tiny filaments embedded in walls cannot be detected in lower resolution simulations. Walls are sensitive to breaking up, but it is not specifically expected that they break up in filaments rather than blobs. (Bernard Jones, personal communication)

6.9 Vorticity magnitude

As explained in section 2.9 — and particularly equation 96 — the vorticity measurements in a DTFE-estimated velocity field may be to any extent clouded by a *projection effect*. In this study, the measured 'vorticity' appears to consist almost entirely out of that effect, from the projection of velocity streams onto Eulerian space. While there are clear indications of shell crossing in the low redshift snapshots of the Cosmogrid simulation, the vorticity measurements are of a very low quality and no reliable results about physical vorticity could be derived.

Maps

The spatial distributions of $|\omega|$ at $z = 3.7$ and $z = 0.0$ are mapped out in figure 64. While certain structures are clearly visible, the maps are far noisier than the other quantities in this investigation. Despite the small filter scale, small-scale structure is not clearly visible, and the walls appear like thick, solid regions rather than double spined features.

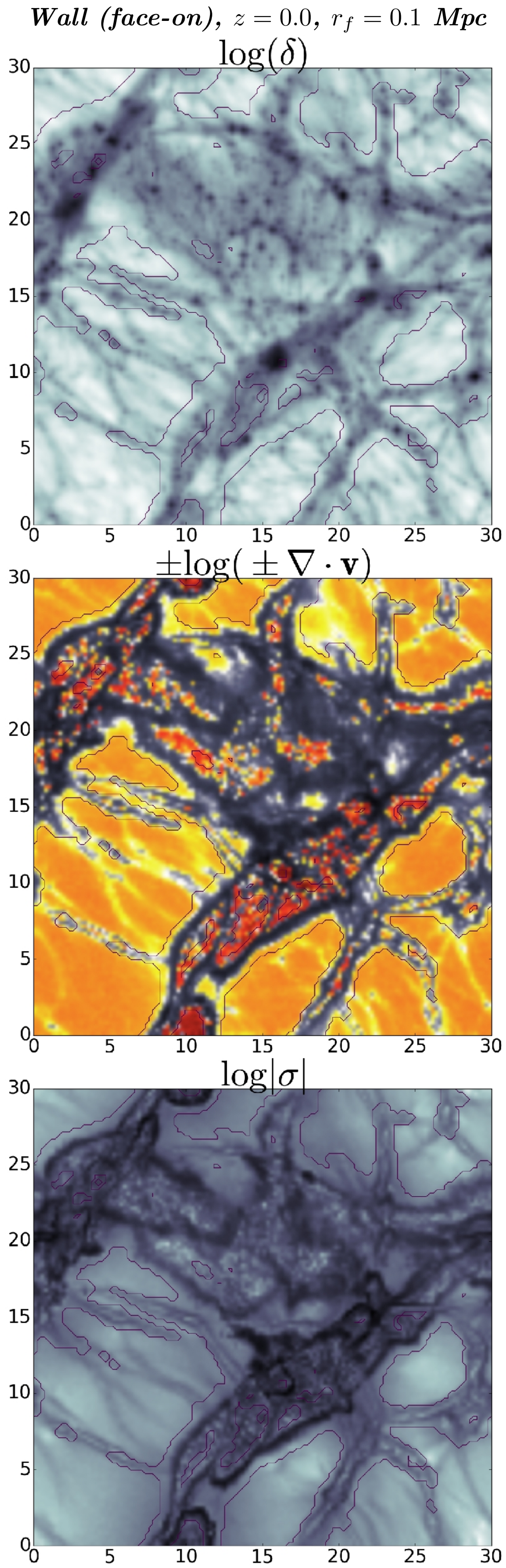


Figure 63: Logarithmic density, divergence and shear maps from a slice of the Cosmogrid volume that cuts along a wall in the cosmic web. These data are from $z = 0.0$ and filtered at 0.1 Mpc. Purple contours — filtered at 0.25 Mpc — delineate voids from massive features. Colour codes are the same as in figures 30, 32 and 57. Note that this face-on view reveals a rich substructure of small filaments embedded in the wall.

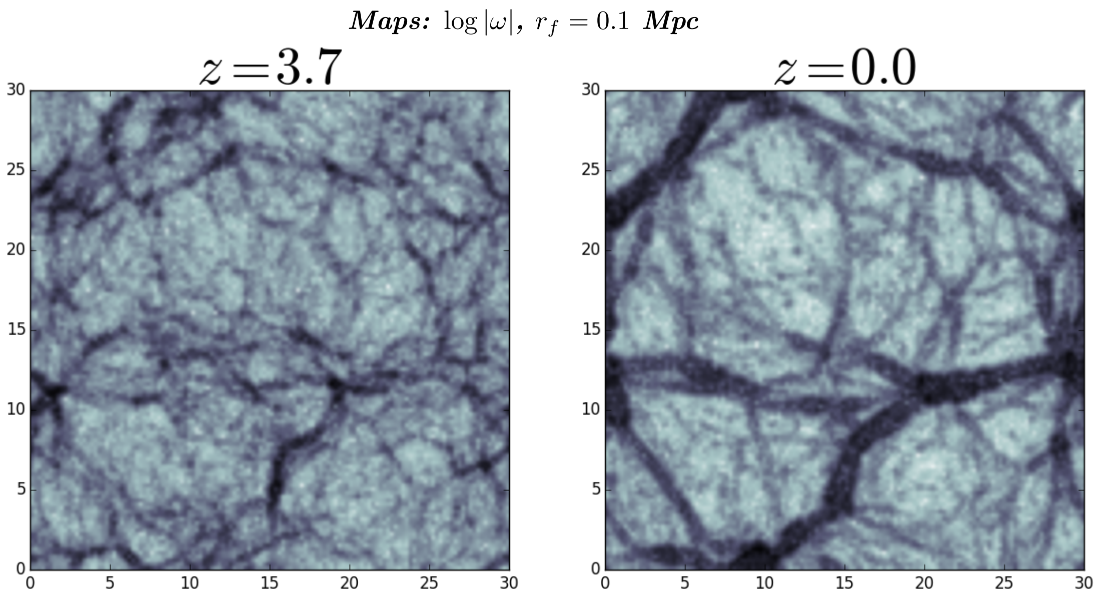


Figure 64: Spatial distribution of the measured vorticity magnitude at $z = 3.7$ and $z = 0.0$. Note that these data are very noisy, and consist mainly of a projection effect (Hahn et al., 2015). These measurements follow closely the structures in density, see figure 30.

From equation 96, we see that the projection effect

$$\langle(\nabla \log \rho) \times (\mathbf{v} - \langle\mathbf{v}\rangle)\rangle \tag{147}$$

is tightly coupled to the density. This introduces a cause to be particularly suspicious of the vorticity measurements from high-density regions, and it is these regions that dominate the maps in figure 64.

Statistical distribution

In figure 65, the statistical distributions of vorticity magnitude are compared with their corresponding density distributions. While the distributions at $z = 0.0$ are off by roughly one order of magnitude, those at $z = 3.7$ are separated by more than two orders of magnitude.

The vorticity magnitude distribution broadens by a very large amount over time, much more than the density distribution does. This indicates that the tentative normalisation $\frac{|\omega|}{Haf}$ fails in the nonlinear phase of velocity growth. A proper normalisation, reflecting nonlinear evolution, would be a nonlinear transformation of $|\omega|$.

Correlation with density

The correlations between the measured vorticity magnitude and density at redshifts 3.7 and 0.0 are plotted in figure 66. Most notable is that there is a certain correlation, but this may come entirely from the projection effect — expression 147 — which is itself a direct function of density. As with the shear magnitude, the fact that the data set from $z = 0.0$ is not centred around the

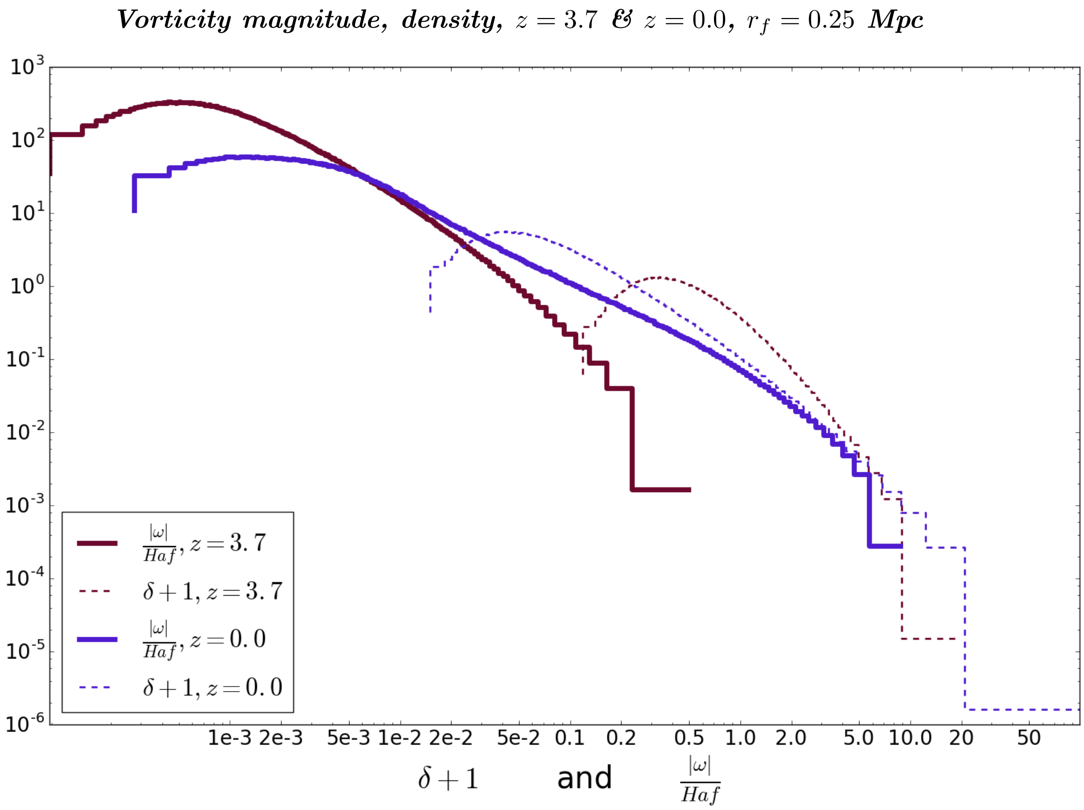


Figure 65: Statistical distributions of overdensity (dashed lines) and the measured vorticity magnitude (solid lines) at $z = 3.7$ (in red) and $z = 0.0$ (in purple). Note well that the vorticity magnitude measurements are of a very poor quality, and consist mainly of a projection effect (see Hahn et al., 2015).

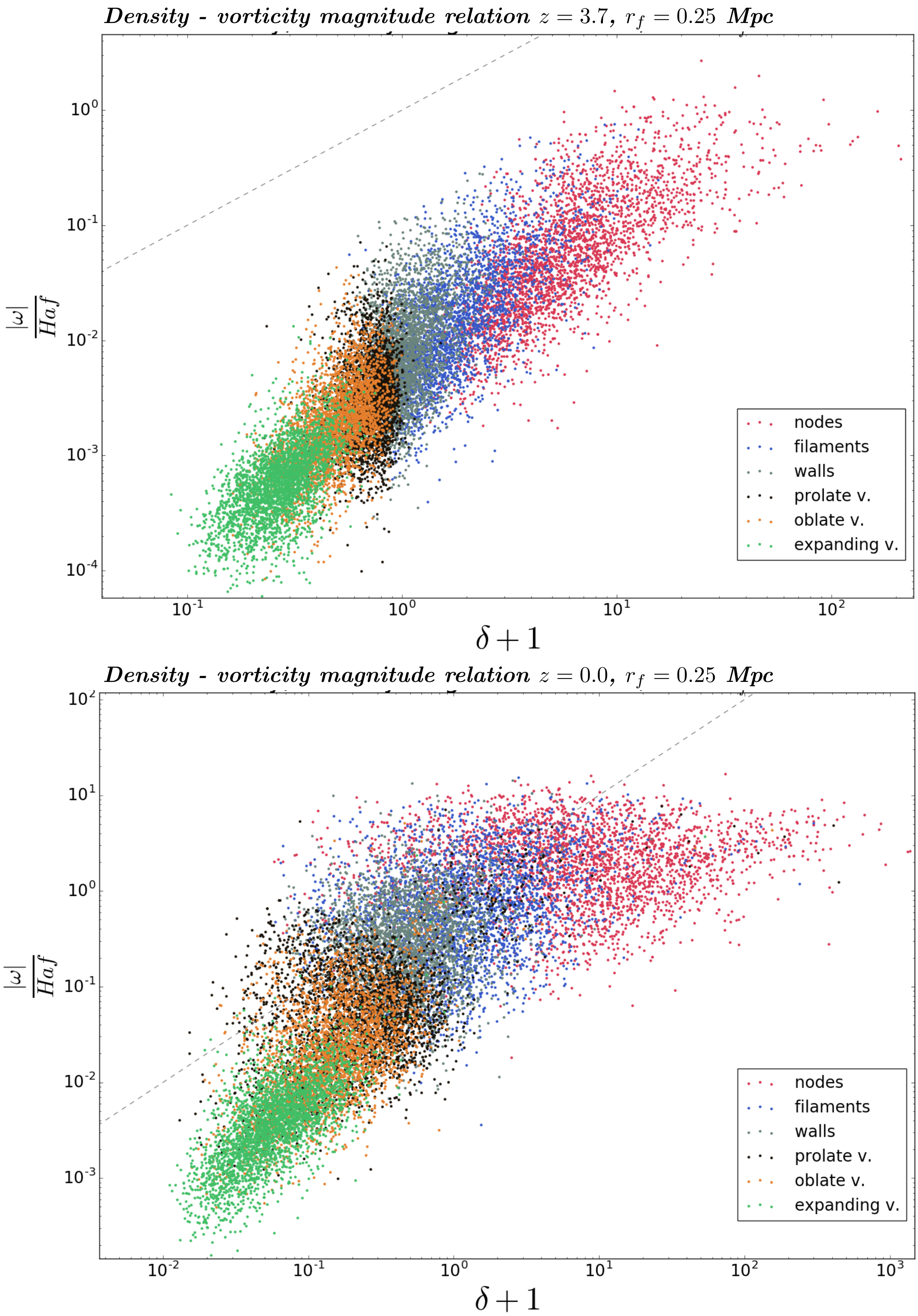


Figure 66: Correlations between density and the measured vorticity magnitude at $z = 3.7$ (upper panel) and $z = 0.0$ (lower panel). Colours indicate cosmic web components. Note that the vorticity magnitude measurements are of a very poor quality, see text.

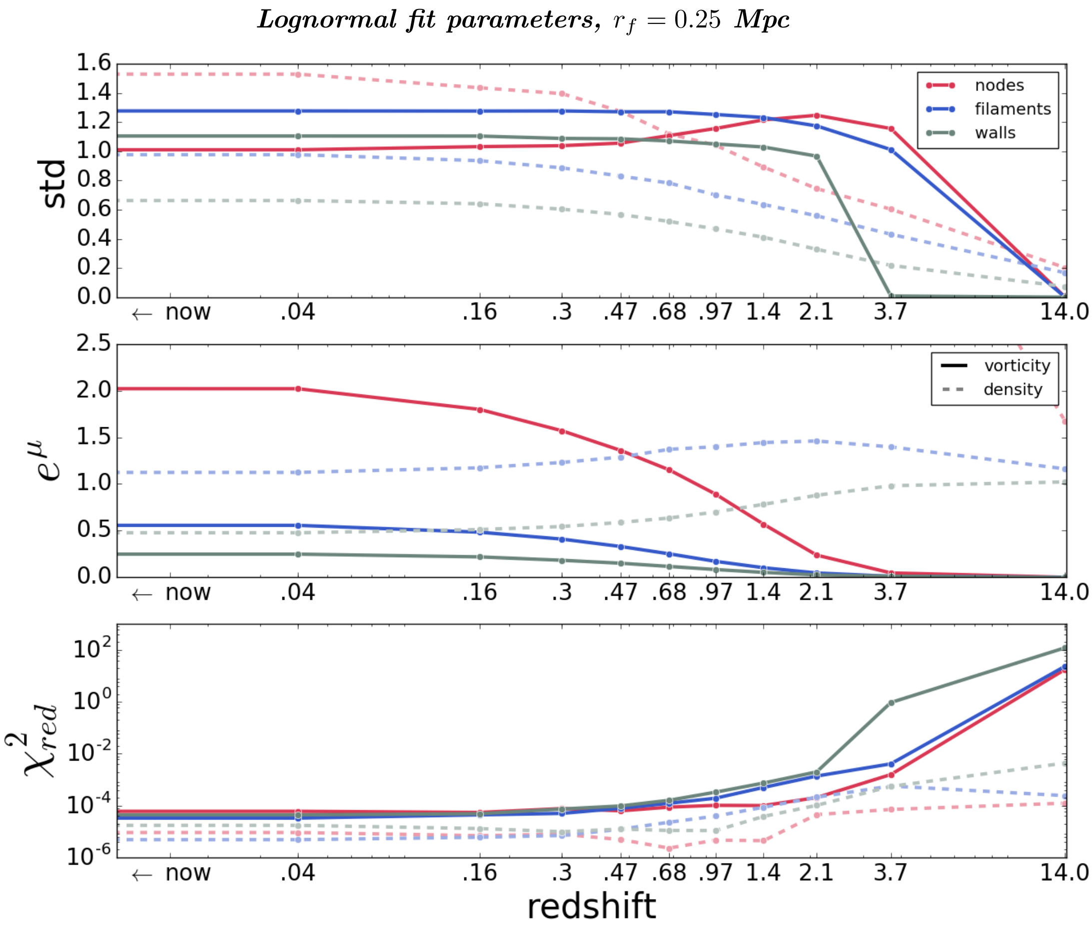


Figure 67: Redshift evolution of lognormal fit parameters to density (dashed) and vorticity magnitude (solid) distributions. Only data from walls, filaments and nodes are shown.

$y = x$ diagonal is arbitrary, since the proper nonlinear normalisation of $|\omega|$ is not determined. On top of that, the large deviation from the $y = x$ line at $z = 3.7$ — which is in the linear stage — is a consequence of the poor vorticity measurements.

At $z = 3.7$ there is a rather sharp — though not absolute — boundary between overdense and underdense regions. This time it is a vertical line, meaning that the vorticity magnitude measurement itself is has little bearing on the classification. There is a large overlap between the regions on the density-vorticity plane. At lower redshift, this overlap only appears to increase; and the boundary changes in orientation.

Lognormal fit parameters

In an analysis equivalent to the study of fits to the density, divergence and shear magnitude distributions — sections 6.5 and 6.7 — lognormal fits were made to the vorticity magnitude distributions. Here, we compare the medians e^{μ} and standard deviations *std* of the vorticity magnitude and density distributions. Figure 46 provides a few examples of lognormal curves, where the two parameters are varied.

Figure 67 presents the redshift evolutions of the fit parameters for walls, filaments and nodes. The void regions are not expected to contain vorticity flows. Interpretations of the density data have been made in section 6.5. Note that the medians of the density distribution fits from nodes increase by a great amount, which is visible in figure 47, but omitted here.

In contrast to the shear distributions, the vorticity distributions undergo a large and early increase in standard deviations. The medians increase strongly as well, but level off at lower values than the shear medians. Here, too, we see a largely parallel change in fit parameters from different web components. With relatively high standard deviations and low medians, it appears that most of the Cosmogrid volume is characterised by low vorticity magnitudes. It is clearly the nodes where these measurements peak — an unsurprising result in the light of equation 96.

On top of being suspicious about the vorticity measurements, we repeat the warning from section 6.5: An analysis of lognormal fit parameters limits our view on the distributions. Any deviations from the lognormal shape — which may prove a key aspect of nonlinear structure formation — does not show up in these graphs.

The lower panel of figure 67 shows a general improvement of fit qualities over the course of cosmic history. The same qualities of density fits are not reached, however. It would be informative to study a distribution of vorticity magnitudes determined by an estimation in phase space.

Despite these shortcomings, the hierarchical nature of structure evolution is retrieved in these results. When the redshift evolutions of fit parameters to distributions from different spatial scales are compared, we find only small differences. The small scale distributions evolve earlier, large scale distributions undergo the same changes some time later.

6.10 DTFE-TSC comparison

Maps

Figure 68 shows the spatial distributions of various physical quantities, from the lower right quadrant of the Cosmogrid slice that has been shown before. This figure allows the comparison between the Delaunay Tessellation Field Estimator — see section 4.5 — and the Triangular Shaped Cloud estimator — section 4.2. From the density maps, it is clear that the DTFE technique allows finer structures to be resolved better. The density peaks in the DTFE-estimated fields are not only smaller, but also less isotropic. In TSC estimation — as it is based on a fixed shape kernel — some information on the shape of anisotropic features is lost.

A comparison between velocity maps yields more dramatic differences. It is important to remark that the rippling artefacts surrounding the high density regions in the TSC maps are not a consequence of the TSC algorithm. This is aliasing due to the differentiation of the velocities, which was done in Fourier space — see section A.3 for an explanation. The fine structures in the divergence maps are blurred out in the TSC estimation. Those in the shear

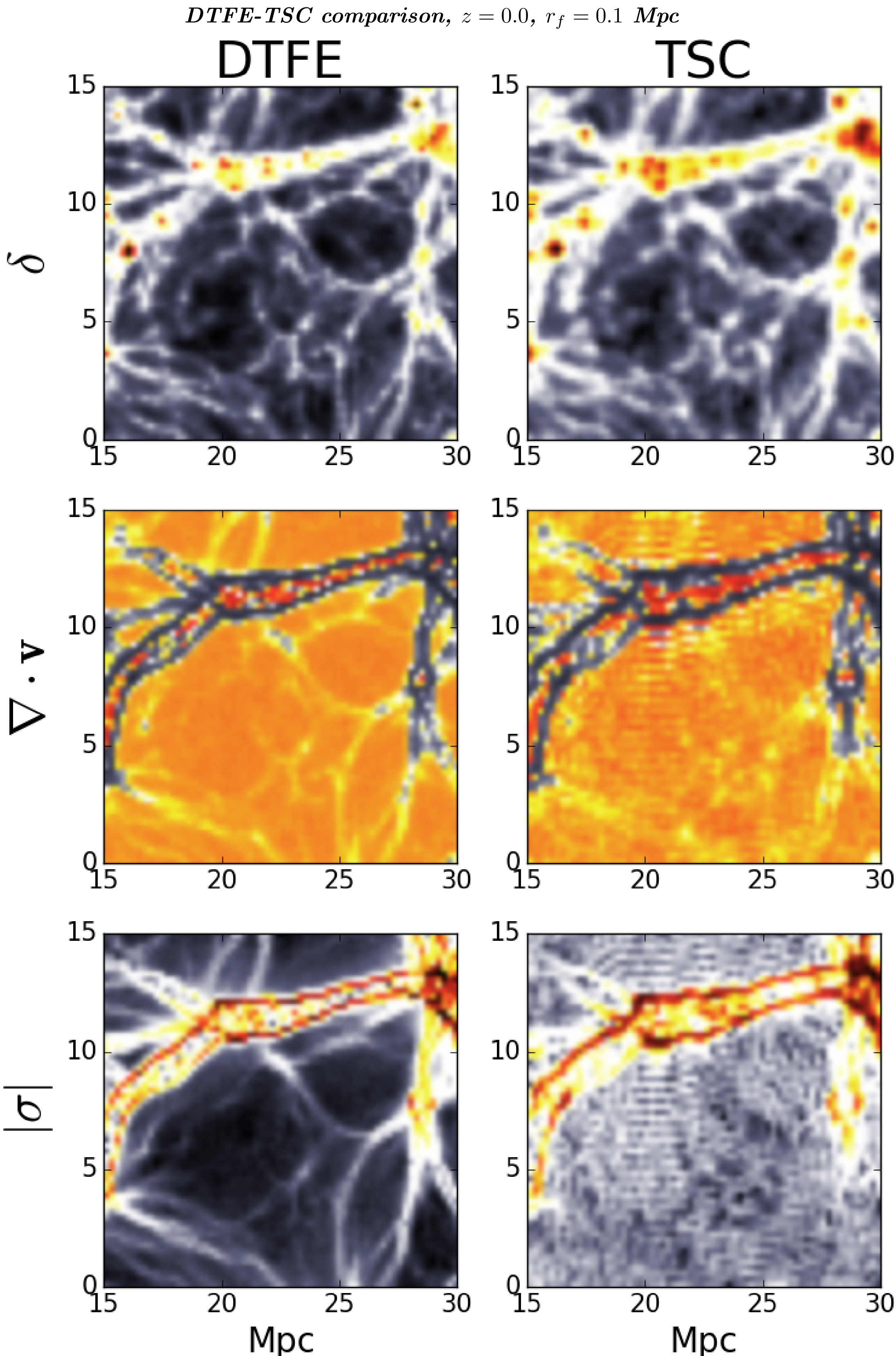


Figure 68: Spatial distributions of density (top row), divergence (middle row) and shear magnitude (bottom row) determined by the DTFE technique (left column) and the TSC estimator (right column). The fringes in the divergence and shear maps on the right are a consequence of the velocity derivation, not the TSC algorithm. The area shown is the South East quadrant of the familiar slice of the Cosmogrid volume.

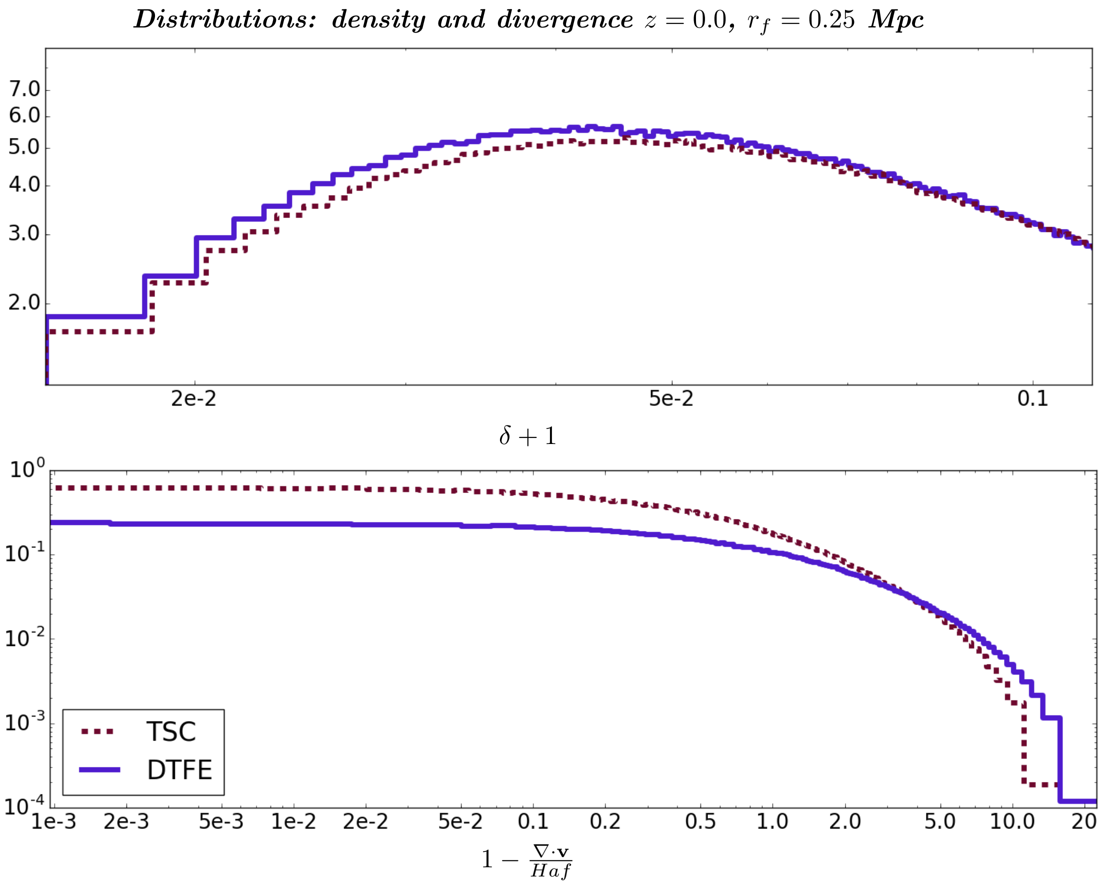


Figure 69: Statistical distributions of density (upper panel) and divergence (lower panel) determined by DTFE (solid lines) and TSC (dashed lines) algorithms. Only the low-density portion of the density distribution is shown — the two distributions coincide rather accurately up to $\delta + 1 \simeq 50$.

magnitude maps are completely lost in noise.

Statistics

Figure 69 presents the DTFE and TSC estimated densities and divergences. The discrepancy between the two divergence distributions is far larger, but it can be to any extent attributed to aliasing in the velocity differentiation.

The density distributions from different estimators agree very well. Figure 69 shows only the low-density portion, where a slight discrepancy occurs. The TSC method estimates a slightly smaller volume with these low densities, which seems to be a consequence of its limited spatial resolution. The density maps in figure 68 illustrates this too: densities from small scale massive features 'leak' into their underdense environments.

Web classification

The web classification based on the TSC-estimated deformation tensor[61] is

[61] id est a combination of the TSC-estimated shear and divergence.

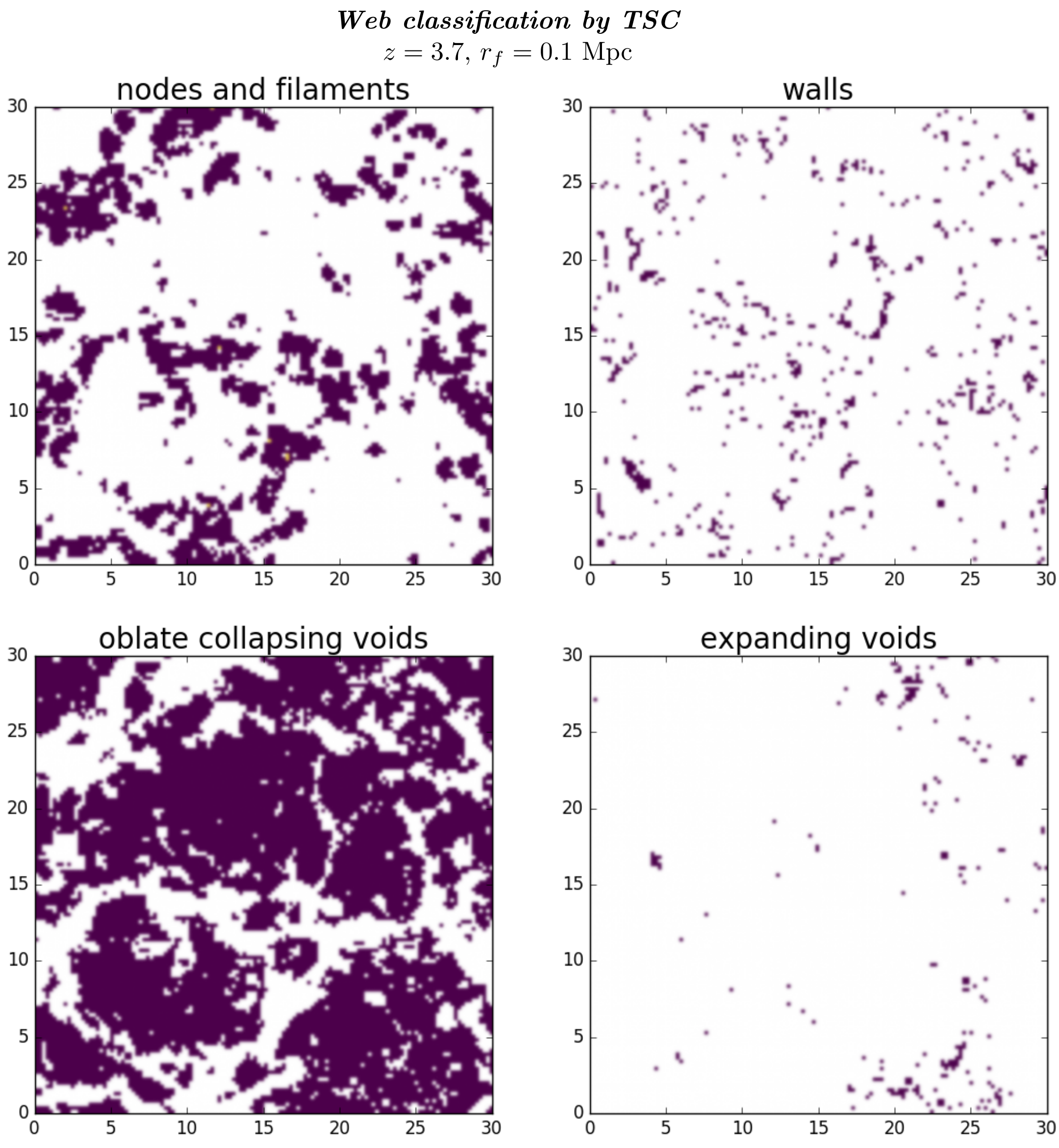


Figure 70: Web classification based on the TSC-estimated deformation tensor.

presented in figure 70. In comparison to the DTFE-based web classification — see figures 35 and 36 — this classification results in an extremely unrealistic prominence of oblate void regions, and almost no expanding void regions at all.

It is clear that the velocity measurements are very sensitive to artefacts showing up in both TSC estimation and velocity differentiation via Fourier space. This study has a great deal to thank the DTFE algorithm for.

7 Discussion and possible artefacts

This section provides a brief account of the caveats and artefacts of relevance to the data and analysis techniques used in this study. We have applied a powerful field estimator and web classifier to the results of a high resolution N-body simulation. Here, we explain the mechanisms that determine to what extent the produced results are reliable. We also identify a number of phenomena that arise purely from the numerical methods, and represent no physical entities.

A more complete assessment of the encountered artefacts is presented in appendix A.

The Cosmogrid simulation

The Cosmogrid simulation — upon which we based these investigations — has a very high spatial and mass resolution. A box of $(30\ \mathrm{Mpc})^3$ box contains 2048^3 particles, although we use only 512^3 particles, due to computational limitations. Much of our analysis is conducted upon a 128^3 cell grid, yielding a spatial resolution of 0.234 Mpc.

Besides the simulation's numerous strengths, the limited box size of 30 Mpc has two disadvantages. Firstly, it is less than the scale of homogeneity — see the introduction — and the volume is thus too small to be representative of the whole Universe. This brings the risk of misrepresenting the prominence of any feature contained in the volume. Secondly, the *fundamental mode* of the largest structures is subject to heightened uncertainties, due to the small number of long wavelength Fourier components that can be defined. Both of these effects are stronger at lower redshifts — in the nonlinear regime.

The Cosmogrid simulation contains only dark matter. The effects of baryons, radiation and dark energy on structure formation and evolution have not been represented. Nonetheless, the simulation is a powerful tool for the purpose of investigating gravitational structure formation.

The Delaunay Tessellation Field Estimator

The DTFE — see section 4.5 — is among the most advanced methods for field estimation. It conducts a piecewise linear interpolation of field quantities between sampling points. It ensures minimal interpolation errors because it is based on the most compact possible triangulation of the sampling points: the Delaunay triangulation.

Since no fixed kernel size is defined, the DTFE technique performs equally well in regions of extremely low and extremely high sampling densities. It is not dependent on any fixed kernel shape or orientation either, so that it reproduces well the shape of anisotropic features. Another important strength of this method is that it yields *volume-weighted* averages, rather than *mass-weighted* ones. This is preferable since most analytically derived quantities are volume-weighted; the DTFE is among the very few techniques that satisfies this.

We have been able to confirm the superiority of the DTFE technique with respect to the Triangular Shaped Cloud estimator — see section 4.2. The TSC method depends on a kernel that is fixed in size and aligned to the estimation grid. It also produces mass-weighted averages.

It is in multi-streaming regions where a limitation of the DTFE technique is exposed. By interpolation between nearby particles, no distinction is made between the different overlapping velocity flows (Hahn et al., 2015). Particularly the measurements of velocity vorticity are affected by this, as vorticity is generated precisely in these regions.

Visually, massive walls in the cosmic web have assumed a *double-spined* appearance in velocity-related fields at low redshifts — these artefacts are visible in divergence, eigenvalues and shear magnitude. High velocity divergence is measured in the innermost areas of these features. These artefacts are a consequence of shell crossing that occurs in high-density areas at low redshifts.

Deformation eigenvalues

The statistical distribution of the largest eigenvalues exhibits an unexpected and unexplained dent around $1 - \frac{\lambda_1}{haf} \in \{0.2; 1.0\}$. This dent is precisely compensated by bumps in the other eigenvalue distributions, resulting in a smooth $\sum_i \lambda_i$ distribution. This effect persists throughout all redshifts, but occurs at slightly higher eigenvalues at later times. A decomposition of web components reveals that the decrease in λ_1 samples occurs at lower eigenvalues for regions that are characterised by lower densities — i.e. voids. The density dependence of this effect matches the correlation between λ_1 and density: they have precisely the same slope. Towards low redshifts, this effect takes on a strange and unexplained shape in the $\lambda_1 - \delta$ plane. We have investigated a range of possible causes, but have not yet succeeded in identifying it.

Web classification

We have identified cosmic web regions on the basis of the eigenvalues of the deformation tensor, which is determined by velocity shear and divergence. Our classifier is unique in making a distinction on the basis of the *overall* expansion or contraction of a region, on top of the eigenvalues individually. We make no arbitrary assumption for the threshold λ_{th}, and our classifier operates on only one spatial scale at a time — we apply it to several.

Velocity flows are a large scale phenomenon, which means that a web classification on the basis of velocity flows is subject to certain resolution limitations. A greater limitation arises from ambiguities in the collapse of structures: without a well-chosen λ_{th}, slightly contracting regions are treated as if they are collapsing fully. This brings misclassifications, we find oblate void regions in particular to be very prominent.

Essentially, our classification assumes $\lambda_{th} = 0$; figure 71 shows the result of a decomposition where we empirically tweaked λ_{th} to yield a more visually realistic balance of region types.

Fourier artefacts

Aliasing is a virtually inevitable effect encountered in discrete Fourier transforms (DFT) of sampled signals. It occurs whenever a wave component of the signal does not match any of the frequencies encoded in the Fourier transform. That Fourier power is then transferred to other frequencies. This means that components with higher frequencies than half of the sampling frequency — the *Nyquist frequency* — can never be represented in the DFT.

One important artefact in our data set is a consequence of the limited box size of the Cosmogrid simulation. The longest wavelength components that can be defined in any box are only few in number. This means that long wavelength components are subject to higher statistical uncertainties. At low redshift, the largest structures in the simulation are dominated by wavelengths comparable

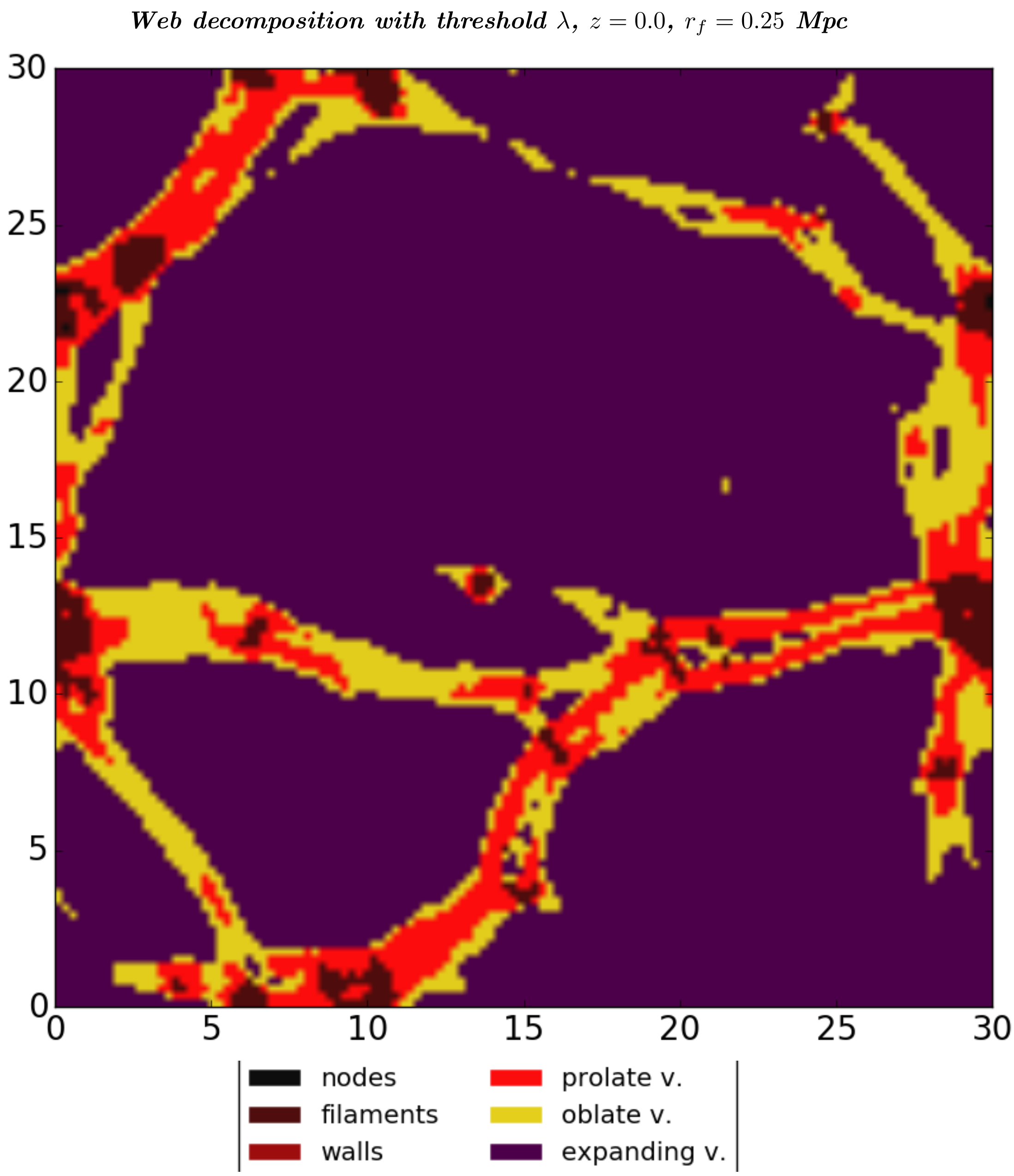


Figure 71: Spatial layout of web components at $z = 0.0$, classified by the same method introduced in section 5.4, only with a non-zero threshold eigenvalue.

to the box size of 30 Mpc. This may cause rippling imprints on the spatial distributions, as can be seen in figures 50 and 51.

8 Conclusions

Our milky way — the spiral galaxy in which we spend all of our time — is only one of billions of galaxies that occupy the observable Universe. At far larger scales, the way matter is distributed in space is referred to as the *cosmic web*. This project has been an exploration of velocity flows from a simulated Universe, and their roles as probes and agents in the formation of this structure.

We have studied the density and velocity fields from the Cosmogrid simulation (Ishiyama et al., 2013). Our methods include a Delaunay Tessellation Field Estimation (DTFE) of density and velocity fields from the simulation particles. We decomposed the velocity gradient into divergence, shear and vorticity, and classified six different components of the cosmic web on the basis of the eigenvectors of the deformation tensor.

Section 6 presents the spatial and statistical distributions of these quantities, and a decomposition of these fields into the contributions from different web components. We have studied the correlations between density and various velocity-related quantities, and followed the redshift evolution of parameters for the lognormal fits to the statistical distributions.

From these results, we distil a number of conclusions: We find various forms of evidence for hierarchical evolution of cosmic structures; we determine the extent to which density and velocity divergence are correlated; we explore the formation and interactions between different structures that make up the cosmic web; we specifically probe the evolution of anisotropic structures; and follow the time evolutions of density, divergence, shear and vorticity.

Appendix B provides a detailed list of the conclusions that are drawn from our results. Here, we advert the main points.

We have recovered various signs of hierarchical structure formation. The density field, exhibits mergers of small scale structures into larger ones, both in overdense and underdense regions. Observable moments in structure evolution — a balance between expanding voids and filaments, and an increase in oblate void regions — occur at later times on larger spatial scales. The same goes for the changes in statistical distributions of various quantities as probed by lognormal fits: small scale structures evolve earlier than large scale structures.

We visualised the evolution of shear fields at several spatial scales in parallel. Small scale structures undergo more intensive changes, involving many mergers; while large scale structures remain relatively stable.

Particularly at high redshifts, a close correlation between density and divergence is recovered, both in spatial and statistical distributions. Above a certain threshold, however, the density keeps increasing while the convergence levels off — the same happens in voids, where divergence no longer increases towards extremely underdense regions. The density distribution exhibits a greater deviation from the lognormal curve than the velocity-related quantities[62]. The median velocity convergence increases over cosmic time, while the median density remains stable.

In our unique classification of cosmic web structures, we find oblate void regions to enclose expanding void regions virtually everywhere. Between different

[62] A study by Uhlemann et al. (2015) has provided analytical support for this observation.

spatial scales, we find variations in the proportions of space occupied by various web components. Notably, we find oblate void regions to be very prominent, and this finding has been reproduced by a separate numerical stochastic experiment.

We have probed the formation and evolution of anisotropic structures by studying the velocity shear and the eigenvalues of the deformation tensor. Both the $|\sigma|$ and λ_1 fields show a deep hierarchy of anisotropic structures, spanning broad ranges of density. We find rather high shear measurements in filaments and walls. We also see an intricate filamentary and clumpy substructure within the face of a wall.

Densities were found to spread both to far lower and far higher levels as cosmic time progresses, reflecting the unbounded growth of the density contrast under gravitational instability. The density field acquires sharp, localised peaks — both in nodes and in filamentary regions. Velocity-related fields do not mimic this.

The divergence distribution generated by a random Gaussian velocity field match well with the measured divergences from high redshifts. This also holds for contributions from all cosmic web components separately. The shapes of shear magnitude distributions evolve mostly parallel in all region types. At low redshifts, the shear grows nonlinearly — somewhat faster than Haf. Like the density, shear acquires a more shallowly decreasing high-value tail. Vorticity measurements are dominated by a projection effect (Hahn et al., 2015), which is coupled to density.

We have studied the formation and evolution of a great hierarchy of structures. Some of them so vast that they take many millions of human lifespans to cross, even at the speed of light. In at least one extremely small, nonlinear component of this hierarchy, the inhabitants have acquired a taste for solving the mysteries that populate the Universe. Scientific investigation can be seen analogous to a journey — and at times, when one journey comes to a close, it kindles the opening of another.

9 Acknowledgements

Nanos gigantum humeris insidentes

I express my gratitude to Rien van de Weygaert, for his guidance in this project, for providing valuable insights in the topic, and constructive comments on the thesis. Cosmological research is a challenging but very enjoyable activity.

It is by virtue of the Cosmogrid simulation that this project could take place, so I am indebted to the authors. Steven Rieder, in particular, has provided the data set and some help in its exploitation. Similarly, this project has relied on the `DTFE` code, which conducts the field estimation. I thank Marius Cautun for developing the software, and for some support in its usage.

I also owe gratitude to Johan Hidding, Bernard Jones, and Patrick Bos, for useful and amusing conversations about this field of study. As coordinator, Marco Spaans has offered some practical advice regarding this project. On occasion, the computer group — Martin, Eite and Valentin — have cast some helpful magic my way.

Lastly, my parents belong in this section too, for myriad reasons. Not the least among them are a cultivation of curiosity; encouragement when it was needed; and the occasional indispensable arm-twisting.

References

????

Aarseth, S. J. 1963, MNRAS, 126, 223

Abel, T., Hahn, O., & Kaehler, R. 2012, MNRAS, 427, 61

Adler, R. J. 1981, The Geometry of Random Fields

Aghanim, N., & Baccigalupi, C. 2015, IAU General Assembly, 22, 2255659

Alpaslan, M., Robotham, A. S. G., Obreschkow, D., et al. 2014, MNRAS, 440, L106

Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., & van der Hulst, J. M. 2007a, A&A, 474, 315

Aragon-Calvo, M. A., & Szalay, A. S. 2013, MNRAS, 428, 3409

Aragón-Calvo, M. A., van de Weygaert, R., & Jones, B. J. T. 2010, MNRAS, 408, 2163

Aragón-Calvo, M. A., van de Weygaert, R., Jones, B. J. T., & van der Hulst, J. M. 2007b, ApJ Letters, 655, L5

Arnold, V. I., Shandarin, S. F., & Zeldovich, I. B. 1982, Geophysical and Astrophysical Fluid Dynamics, 20, 111

Avila-Reese, V., Colín, P., Gottlöber, S., Firmani, C., & Maulbetsch, C. 2005, ApJ, 634, 51

Bagla, J. S. 2002, Journal of Astrophysics and Astronomy, 23, 185

Bagla, J. S., Prasad, J., & Khandai, N. 2009, MNRAS, 395, 918

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15

Bardeen, J. M., Steinhardt, P. J., & Turner, M. S. 1983, Phys. Rev. D, 28, 679

Barnes, J., & Hut, P. 1986, Nature, 324, 446

Baugh, C. M. 2006, Reports on Progress in Physics, 69, 3101

Bennett, D. N., Halpern, M., & Hinshaw, G. e. 2003, ApJS, 148, 1

Bernardeau, F., & Reimberg, P. 2015, ArXiv e-prints, arXiv:1511.08641

Bernardeau, F., & van de Weygaert, R. 1996, MNRAS, 279, 693

Bertschinger, E. 1987, ApJ Letters, 323, L103

—. 2001, ApJS, 137, 1

Binney, J., & Quinn, T. 1991, MNRAS, 249, 678

Birdsall, C., & Langdon, A. 1985, Plasma Physics via Computer Simulations, 56

Birdsall, C. K., & Fuss, D. 1969, Journal of computational physics, 141

Blanton, M. R., Eisenstein, D., Hogg, D. W., Schlegel, D. J., & Brinkmann, J. 2005, ApJ, 629, 143

Bode, P., & Ostriker, J. P. 2003, ApJS, 145, 1

Bode, P., Ostriker, J. P., & Xu, G. 2000, ApJS, 128, 561

Bond, J. R., & Szalay, A. S. 1983, ApJ, 274, 443

Bond, N. A., Strauss, M. A., & Cen, R. 2010a, MNRAS, 406, 1609

—. 2010b, MNRAS, 409, 156

Bond, R., Kofman, L., & Pogosyan, D. 1996, Nature, 380

Cautun, M., van de Weygaert, R., & Jones, B. J. T. 2013, MNRAS, 429, 1286

Cautun, M., van de Weygaert, R., Jones, B. J. T., & Frenk, C. S. 2014, MNRAS, 441, 2923

Ceccarelli, L., Ruiz, A. N., Lares, M., et al. 2015, ArXiv e-prints, arXiv:1511.06741

Colberg, J. M., Krughoff, K. S., & Connolly, A. J. 2005, MNRAS, 359, 272

Coles, P., & Jones, B. 1991, MNRAS, 248, 1

Colless, M., Peterson, B. A., Jackson, C., et al. 2003, ArXiv Astrophysics e-prints, astro-ph/0306581

Crocce, M., Fosalba, P., Castander, F. J., & Gaztañaga, E. 2010, MNRAS, 403, 1353

Dekel, A. 1994, ARA&A, 32, 371

Delone, B. N. 1934, Bull. Acad. Sci. USSR: Classe Sci., 793

Dolag, K., Meneghetti, M., Moscardini, L., Rasia, E., & Bonaldi, A. 2006, MNRAS, 370, 656

Dolag, K., & Stasyszyn, F. 2009, MNRAS, 398, 1678

Doroshkevich, A. G., Kotok, E. V., Poliudov, A. N., et al. 1980, MNRAS, 192, 321

Doroshkevich, A. G., & Shandarin, S. F. 1978a, MNRAS, 182, 27

—. 1978b, SOVAST, 22, 653

Dressler, A. 1980, ApJ, 236, 351

Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, MNRAS, 401, 1429

Dubinski, J., Nicolaci da Costa, L., Goldwirth, D. S., Lecar, M., & Piran, T. 1993, ApJ, 410, 458

Efstathiou, G. 1986a, Nature, 322, 311

Efstathiou, G. 1986b, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 267, The Use of Supercomputers in Stellar Dynamics, ed. P. Hut & S. L. W. McMillan, 36

Efstathiou, G., Davis, M., White, S. D. M., & Frenk, C. S. 1985, ApJS, 57, 241

El-Ad, H., & Piran, T. 1997, ApJ, 491, 421

El-Ad, H., Piran, T., & da Costa, L. N. 1996, ApJ Letters, 462, L13

Evrard, A. E., MacFarland, T. J., Couchman, H. M. P., et al. 2002, ApJ, 573, 7

Fergusson, J. R., & Shellard, E. P. S. 2009, Phys. Rev. D, 80, 043510

Fisher, J. D., Faltenbacher, A., & Johnson, M. S. T. 2016, MNRAS, arXiv:1511.05971

Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., & Yepes, G. 2009, MNRAS, 396, 1815

Forero-Romero, J. E., Hoffman, Y., Yepes, G., et al. 2011, MNRAS, 417, 1434

Fossen, H. 2016, Structural Geology: Cambridge University Press, ISBN 9781107057647

Fry, J., & Ma, C. 2001, Annals of the New York Academy of Sciences, 927, 143

Ganon, G., & Hoffman, Y. 1993, ApJ Letters, 415, L5

Gao, L., Springel, V., & White, S. D. M. 2005, MNRAS, 363, L66

Gramann, M. 1993, ApJ Letters, 405, L47

Gregory, S. A., & Thompson, L. A. 1978, Nature, 274, 450

Groen, D., Portegies Zwart, S., Ishiyama, T., & Makino, J. 2011, Computational Science and Discovery, 4, 015001

Groen, D., Rieder, S., & Portegies Zwart, S. F. 2012, Acta Cosmologica, 161

Guth, A. H., & Pi, S.-Y. 1982, Physical Review Letters, 49, 1110

Hahn, O., Angulo, R. E., & Abel, T. 2015, MNRAS, 454, 3920

Hahn, O., Porciani, C., Carollo, C. M., & Dekel, A. 2007, MNRAS, 375, 489

Hamilton, A. J. S. 1998, in Astrophysics and Space Science Library, Vol. 231, The Evolving Universe, ed. D. Hamilton, 185

Hawking, S. W. 1982, Physics Letters B, 115, 295

Hellaby, C., & Lake, K. 1985, ApJ, 290, 381

Hidding, J., van de Weygaert, R., Vegter, G., & Jones, B. J. T. 2012, ArXiv e-prints, arXiv:1211.5385

Hockney, R. W., & Eastwood, J. W. 1981, Computer Simulation Using Particles

Hoffman, Y., Metuki, O., Yepes, G., et al. 2012, MNRAS, 425, 2049

Hoffman, Y., & Ribak, E. 1991, ApJ Letters, 380, L5

Hogg, D. W., Eisenstein, D. J., Blanton, M. R., et al. 2005, ApJ, 624, 54

Horváth, I., Hakkila, J., & Bagoly, Z. 2014, A&A, 561, L12

Hubble, E. 1934, ApJ, 79, 8

Huchra, J., Jarrett, T., Skrutskie, M., et al. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 329, Nearby Large-Scale Structures and the Zone of Avoidance, ed. A. P. Fairall & P. A. Woudt, Fairall

Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, ApJS, 199, 26

Hui, L., & Bertschinger, E. 1996, ApJ, 471, 1

Icke, V. 1973, A&A, 27, 1

Icke, V., & van de Weygaert, R. 1987, A&A, 184, 16

Iliev, I. T., Mellema, G., Pen, U.-L., et al. 2006, MNRAS, 369, 1625

Ishiyama, T., Fukushige, T., & Makino, J. 2009, PASJ, 61, 1319

Ishiyama, T., Rieder, S., Makino, J., et al. 2013, ApJ, 767, 146

Jõeveer, M., Einasto, J., & Tago, E. 1978, MNRAS, 185, 357

Jenkins, A., Frenk, C. S., Pearce, F. R., et al. 1998, ApJ, 499, 20

Jones, B. J., Martínez, V. J., Saar, E., & Trimble, V. 2004, Reviews of Modern Physics, 76, 1211

Kauffmann, G., & Fairall, A. P. 1991, MNRAS, 248, 313

Khandai, N., & Bagla, J. S. 2009, Research in Astronomy and Astrophysics, 9, 861

Kim, J., Park, C., Gott, III, J. R., & Dubinski, J. 2009, ApJ, 701, 1547

Kirshner, R. P., Oemler, Jr., A., Schechter, P. L., & Shectman, S. A. 1981, ApJ Letters, 248, L57

Kitaura, F.-S., Angulo, R. E., Hoffman, Y., & Gottlöber, S. 2012, MNRAS, 425, 2422

Klypin, A. A., & Shandarin, S. F. 1983, MNRAS, 204, 891

Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102

Knebe, A., Gill, S. P. D., Gibson, B. K., et al. 2004, ApJ, 603, 7

Laigle, C., Pichon, C., Codis, S., et al. 2015, MNRAS, 446, 2744

Lambas, D. G., Lares, M., Ceccarelli, L., et al. 2016, MNRAS, 455, L99

Lavaux, G., & Wandelt, B. D. 2010, MNRAS, 403, 1392

Lemson, G., & Kauffmann, G. 1999, MNRAS, 302, 111

Li, B., & Barrow, J. D. 2011, Phys. Rev. D, 83, 024007

Libeskind, N. I., Frenk, C. S., Cole, S., et al. 2005, MNRAS, 363, 146

Libeskind, N. I., Hoffman, Y., & Gottlöber, S. 2014, MNRAS, 441, 1974

Libeskind, N. I., Knebe, A., Hoffman, Y., et al. 2011, MNRAS, 411, 1525

Linde, A. D. 1982, Physics Letters B, 116, 335

Linder, E. V. 2006, Journal of Physics Conference Series, 39, 56

Little, B., Weinberg, D. H., & Park, C. 1991, MNRAS, 253, 295

Lloyd, S. P. 1982, IEEE Transactions on Information Theory, 28, 129

Lonsdale, C. J., & Barthel, P. D. 1986, ApJ, 303, 617

Macciò, A. V., Moore, B., Stadel, J., & Potter, D. 2006, ArXiv Astrophysics e-prints, astro-ph/0609146

Makino, J., Fukushige, T., Koga, M., & Namura, K. 2003, PASJ, 55, 1163

Makino, J., & Taiji, M. 1998, Scientific Simulations with Special-Purpose Computers–the GRAPE Systems, 248

Marinacci, F., Vogelsberger, M., Mocz, P., & Pakmor, R. 2015, MNRAS, 453, 3999

Maulbetsch, C., Avila-Reese, V., Colín, P., et al. 2007, ApJ, 654, 53

Melott, A. L. 1983, MNRAS, 204, 7P

Monaghan, J. J. 1992, ARA&A, 30, 543

Mücket, J. P. 1985, Astronomische Nachrichten, 306, 305

Mukhanov, V. F., & Chibisov, G. V. 1981, Jetp letters, 33, 532

Novikov, D., Colombi, S., & Doré, O. 2006, MNRAS, 366, 1201

Nusser, A., Dekel, A., Bertschinger, E., & Blumenthal, G. R. 1991, ApJ, 379, 6

Padilla, N., Paz, D., Lares, M., et al. 2014, ArXiv e-prints, arXiv:1410.8186

Pauls, J. L., & Melott, A. L. 1995, MNRAS, 274, 99

Peacock, J. A. 1999, The Observatory, 119, 296

Peacock, J. A., & Heavens, A. F. 1985, MNRAS, 217, 805

Peebles, P. J. E. 1980, The large-scale structure of the universe

—. 1982, ApJ, 258, 415

—. 1993, Principles of Physical Cosmology

Petkova, M., & Springel, V. 2009, MNRAS, 396, 1383

Pichon, C., & Bernardeau, F. 1999, A&A, 343, 663

Planck Collaboration, ., Ade, P. A. R., Aghanim, N., et al. 2015a, ArXiv e-prints, arXiv:1502.01589

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015b, ArXiv e-prints, arXiv:1502.01592

Platen, E., van de Weygaert, R., & Jones, B. J. T. 2008, MNRAS, 387, 128

Pogosyan, D., Bond, J. R., Kofman, L., & Wadsley, J. 1998, in Wide Field Surveys in Cosmology, ed. S. Colombi, Y. Mellier, & B. Raban, 61

Portegies Zwart, S., Ishiyama, T., Groen, D., et al. 2010, IEEE Computer, v.43, No.8, p.63-70, 43, 63

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Prunet, S., Pichon, C., Aubert, D., et al. 2008, ApJS, 178, 179

Pueblas, S., & Scoccimarro, R. 2009, Phys. Rev. D, 80, 043504

Råde, L., & Westergren, B. 2004, Mathematics Handbook for Science and Engineering, doi:10.1007/978-3-662-08549-3

Raychaudhuri, A. 1955, Phys. Rev., 98, 1123

Rieder, S., van de Weygaert, R., Cautun, M., Beygu, B., & Portegies Zwart, S. 2013, MNRAS, 435, 222

Rojas, R. R., Vogeley, M. S., Hoyle, F., & Brinkmann, J. 2005, ApJ, 624, 571

Romano-Díaz, E., & van de Weygaert, R. 2007, MNRAS, 382, 2

Romano-Díaz, E., & van de Weygaert, R. 2007, ArXiv e-prints, arXiv:0707.2607

Ruth, R. 1983, IEEE Transactions on Nuclear Science, 30, 2669

Salmon, J. 1996, ApJ, 460, 59

Sarkar, P., Yadav, J., Pandey, B., & Bharadwaj, S. 2009, MNRAS, 399, L128

Sathyaprakash, B. S., Sahni, V., & Shandarin, S. F. 1996, ApJ Letters, 462, L5

Saunders, W., Sutherland, W. J., Maddox, S. J., et al. 2000, MNRAS, 317, 55

Schaap, W. E., & van de Weygaert, R. 2000, A&A, 363, L29

—. 2001, ArXiv Astrophysics e-prints, astro-ph/0109261

—. 2002, A&A, L29

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521

Scrimgeour, M. I., Davis, T., Blake, C., et al. 2012, MNRAS, 425, 116

Shandarin, S., Habib, S., & Heitmann, K. 2010, Phys. Rev. D, 81, 103006

—. 2012, Phys. Rev. D, 85, 083005

Shandarin, S. F., Sheth, J. V., & Sahni, V. 2004, MNRAS, 353, 162

Shandarin, S. F., & Zeldovich, Y. B. 1989, Reviews of Modern Physics, 61, 185

Shapiro, P. R., Struck-Marcell, C., & Melott, A. L. 1983, ApJ, 275, 413

Sheth, R. K., & van de Weygaert, R. 2004, MNRAS, 350, 517

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJ Letters, 396, L1

Sousbie, T. 2011, MNRAS, 414, 350

Sousbie, T., Pichon, C., Colombi, S., Novikov, D., & Pogosyan, D. 2008, MNRAS, 383, 1655

Spergel, D. N., Bean, R., Doré, O., & et al. 2007, ApJ, http://arxiv.org/abs/astro-ph/0603449

Springel, V. 2005, MNRAS, 364, 1105

—. 2010, ARA&A, 48, 391

Springel, V., & Hernquist, L. 2003, MNRAS, 339, 312

Springel, V., Yoshida, N., & White, S. D. M. 2001, NA, 6, 79

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Stanonik, K., Platen, E., Aragón-Calvo, M. A., et al. 2009, ApJ Letters, 696, L6

Starobinsky, A. A. 1982, Physics Letters B, 117, 175

Sugimoto, D., Chikada, Y., Makino, J., et al. 1990, Nature, 345, 33

Summers, F. J. 1993, PhD thesis, California University

Sutter, P. M., Elahi, P., Falck, B., et al. 2014, MNRAS, 445, 1235

Sylos Labini, F., Vasilyev, N. L., Pietronero, L., & Baryshev, Y. V. 2009, EPL (Europhysics Letters), 86, 49001

Tegmark, M., Blanton, M. R., Strauss, M. A., et al. 2004, ApJ, 606, 702

Tempel, E., Libeskind, N. I., Hoffman, Y., Liivamägi, L. J., & Tamm, A. 2014, MNRAS, 437, L11

Teyssier, R., Pires, S., Prunet, S., et al. 2009, A&A, 497, 335

Uhlemann, C., Codis, S., Pichon, C., Bernardeau, F., & Reimberg, P. 2015, ArXiv e-prints, arXiv:1512.05793

van de Weygaert, M. A. M. 1991, PhD thesis, Ph. D. thesis, University of Leiden (1991)

van de Weygaert, R. 1994, A&A, 283, 361

—. 2002, ArXiv Astrophysics e-prints, astro-ph/0206427

—. 2006, ArXiv Astrophysics e-prints, astro-ph/0607539 van de Weygaert, R., & Babul, A. 1997, (in preparation)

van de Weygaert, R., & Bertschinger, E. 1996, MNRAS, 281, 84

van de Weygaert, R., Platen, E., Tigrak, E., et al. 2010, in Astronomical Society of the Pacific Conference Series, Vol. 421, Galaxies in Isolation: Exploring Nature Versus Nurture, ed. L. Verdes-Montenegro, A. Del Olmo, & J. Sulentic, 99

van de Weygaert, R., & Schaap, W. 2009, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 665, Data Analysis in Cosmology, ed. V. J. Martínez, E. Saar, E. Martínez-González, & M.-J. Pons-Bordería, 291–413

van de Weygaert, R., & van Kampen, E. 1993, MNRAS, 263, 481

van Haarlem, M., & van de Weygaert, R. 1993, ApJ, 418, 544

Vanmarcke, E. H. 1983, Random Fiels: Analysis and Synthesis. MIT Press, Cambridge, Massachusetts.

Vazza, F., Brüggen, M., Gheller, C., & Wang, P. 2014, MNRAS, 445, 3706

Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518

Wambsganss, J., Bode, P., & Ostriker, J. P. 2004, ApJ Letters, 606, L93

White, M., Cohn, J. D., & Smit, R. 2010, MNRAS, 408, 1818

White, S. D. M. 1996, in Formation and evolution of galaxies: Les houches lectures. (Dordrecht: Elsevier, astro-ph/9410043), ed. R. Schaefer, J. Silk, M. Spiro, & J. Zinn-Justin, Cosmology and Large-Scale Structure

Xu, G. 1995, ApJS, 98, 355

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Yoshikawa, K., & Fukushige, T. 2005, PASJ, 57, 849

Zel'dovich, Y. B. 1970, A&A, 5, 84

Zel'dovich, Y. B., Einasto, J., & Shandarin, S. F. 1982, Nature, 300, 407

Appendix A Inventory of caveats and artefacts

The main issues that are to be kept in mind in the interpretation of our results are discussed in section 7. This appendix elaborates on those points, to provide a closer view on the strengths and limitations in our data and methods.

A.1 Data

Cosmological volume and resolution

What the Cosmogrid simulation excels in is resolution. The main run, with 2048^3 particles in a $(30\ \mathrm{Mpc})^3$ box, has a sufficient spatial- and mass resolution to resolve ultra faint haloes the size of dwarf galaxies. In the 512^3 particle run, the density and velocity spaces have not been smoothed in post processing. These data have offered a high spatial resolution that is important for the analysis of small scale nonlinear velocity flows.

In view of the available computational power, this study has only used the 512^3 particle simulation, and the analysis was conducted on a 128^3 cell grid. In future, a more industrial analysis, making use of all the available data, would certainly benefit the quality and statistical significance of the results.

One limitation of the Cosmogrid volume is its relatively small volume. As a result, it cannot show us a representative picture of structure in the Universe. The principal homogeneity of the Universe only holds for distance scales $\gtrsim 100$ Mpc (Hogg et al., 2005; Sarkar et al., 2009; Scrimgeour et al., 2012; Sylos Labini et al., 2009). A 30 Mpc box size is comparable to the characteristic scales of the largest actual cosmological structures themselves — structures of these scales are especially dominant at low redshift.

Other than the consequence of this fact in Fourier space — see section A.3 — this introduces a substantial risk that the actual cosmic structures are statistically misrepresented in the simulation. In fact, the structures in the Cosmogrid volume are dominated by a large central void, and thus the overall expansion of the volume is greater than the Hubble expansion.

The authors of the Cosmogrid Universe have found the limited box size to cause an artificial increase in the number of intermediate mass haloes. This is due to the absence of perturbations at long-wavelengths (Ishiyama et al., 2013). They have compared the halo mass function of the 30 Mpc box to 45 Mpc and 60 Mpc simulations to confirm this. Bagla et al. (2009) have asserted the effects of limited box size for the mass functions, and the skewness of velocity distributions.

Cosmological components

Simulating only dark matter, Cosmogrid ignores baryonic matter, radiation, and dark energy. It has left out all effects from pressure, relativistic gravity, magnetic forces, radiative processes, hydrodynamic processes, star formation, and the feedback interactions that are known to be of influence in reality. These factors are particularly important below the scales of galaxy clusters — see section 3.1 — but may also constitute visible corrections at larger scales, to any purely gravitational dark matter simulation. The proper inclusion of all of these factors into a full, realistic simulation of structure formation will be a great challenge.

On top of all this comes a somewhat philosophical critique: simulations, however potent, remain an entity fundamentally different from the Universe around us — it would be more honest and modest to refer to them as 'attempts at simulation'. In his lectures, Rien van de Weygaert warns of a misguided "attitude [...] in which the reproduction of observed patterns or results by a computer model is considered to be an explanation in itself. [...] A true physical understanding may at best only start or be guided by computer experiments."

A.2 Field estimation

DTFE

As a method to determine a field value at any grid location, based on an inhomogeneous sampling of particles, the DTFE technique offers many advantages over alternative methods. Field estimation is, in essence, a difficult task, and no perfect solution has been developed. DTFE is among the most advanced methods available today, and is merely based upon a piecewise linear interpolation. However, the Delaunay Tessellation offers a highly optimised configuration for this interpolation, as it erects the shortest possible vertices between sample points.

One of the greatest strengths of the DTFE method is its unhindered applicability to a wide range of spatial scales and sampling densities. It is not dependent on a fixed kernel size — like, e.g. the TSC method is — and as such it can interpolate between particles at any conceivable sampling density. Neither is it dependent on any kernel orientation — since it is not a grid based estimator. As a result, it is equally well suited for the processing of particles in all anisotropic structures, regardless of their orientation. In contrast, grid based estimators may produce biased results, depending on the degree of alignment between anisotropic structures and the estimator's grid axes.

Another important advantage of the tessellation field estimator is that it yields volume-weighted averages. This constitutes a distinct advantage over grid based field estimators, which yield mass-weighted averages. The analytically derived quantities of interest are nearly always volume-weighted, so the DTFE results are particularly suited for a comparison with analytical work — see section 4.3 for a more detailed discussion.

We have been able to confirm that the DTFE algorithm resolves compact density peaks better than then Triangular Shaped Cloud (TSC) estimator. DTFE also reproduces anisotropic shapes better. Only a slight discrepancy occurs between the density distributions: The TSC method estimates a slightly smaller volume with low densities. This is probably a result of its limited spatial resolution. Significant artefacts in the velocity fields result from both the TSC estimator and from differentiation of velocities via Fourier transforms.

Anisotropic features and grid alignment

While the field estimation in this study is not based on a grid — and thus does not introduce any bias in orientation — the same cannot be said of the subsequent steps in the analysis. The estimated field values were represented in a grid, and their spatial and statistical properties were evaluated from those discrete and regular grid positions.

An anisotropic feature that is well aligned with one of the three grid axes — i.e. it is oriented close to the x, y or z direction — will be represented differently than a feature that is not. As long as the spatial grid resolution is sufficiently high, this difference is limited, but a grid will always introduce some arbitrary effect on the data.

Similarly, the statistical distribution of field values from anisotropic structures is not expected to be influenced significantly, given a sufficiently high spatial resolution. The diversity in scales and orientations of structures occurring in the Cosmogrid volume is of importance. No preferred inherent scale or orientation exists for these structures; therefore, alignment with an arbitrarily oriented grid cannot systematically enhance or suppress any properties of the simulation.

Web classification

Our web classifier, described in detail in section 5.4, shares some of the strengths and weaknesses of those proposed by e.g. Hahn et al. (2007); Forero-Romero et al. (2009); Hoffman et al. (2012). It is based on velocity flows, it assumes no specific threshold eigenvalue λ_{th}, and it acts on only one distance scale — though it has been separately applied to several.

The fact that our classification method is based on the velocity-related deformation tensor is of relevance in interpreting the results. Since velocity flows are large-scale phenomena, and they are influenced by the distribution of matter over large distances, some misclassifications can occur. This happens in cases where a region in the neighbourhood of a given structure exhibits velocity flows characterising that type of structure. The same happens in web classifiers based on the tidal field, which is also of influence at great distances.

Misclassification can also occur due to ambiguity in the collapse or expansion of structures. If a region contracts or expands only slightly along one of its principal directions, our classifier picks up a small but definite positive or negative eigenvalue, and treats this movement as an absolute indicator of a structure's dimensionality. Forero-Romero et al. (2009) work around this problem by introducing a threshold eigenvector, λ_{th} — see section 2.8 for details. The drawback of this method is that it introduces a free parameter, an arbitrary choice in the classification method. This is not adopted in our strategy[63], sparing this study the burden of an arbitrary free parameter, but at the cost of some adequacy in the classification. For example, the prominence of oblate void regions in our results is in part due to this.

Figure 71 presents an alternative — visually more realistic — classification of web components, where a non-zero λ_{th} has been used. Note that the expanding void regions are nearly ubiquitously surrounded by oblate regions. It must be mentioned that the shift in threshold eigenvalues also applies to the separation between generally expanding and contracting regions. As a result, the regions marked as oblate collapsing void regions may include features one could call walls; and prolate collapsing void regions contain filamentary features.

Like many web classifiers that have previously been developed, ours depends on an assumed spatial scale at which structures are identified[64]. Our web classification is applied to scales of 0.1 Mpc, 0.25 Mpc and 0.5 Mpc, and the difference

[63] In essence, we take $\lambda_{th} = 0$.

[64] The Nexus and Nexus+ algorithms (Cautun et al., 2014) form a notable exception to this.

in structures at these scales can be seen e.g. in figures 39 and 57.

It is a fundamental aspect of the cosmic web that no absolute differences in dynamical development between types of regions can be defined from first principles. Not every region can be unambiguously labelled. This makes it impossible to identify structures with full accuracy, or even to quantify the accuracy of any given identification. This is also why web classifiers are fundamentally dependent on free parameters like threshold eigenvalues and distance scales.

Hahn's algorithm

One limitation of the field estimation in this study is that it does not distinguish between the particles that come from different velocity flows. In the case of shell crossing — see section 2.9 — particles from one velocity flow are positioned among those from any number of spatially overlapping flows. The Delaunay tessellation field estimator sec will then interpolate between particles from different streams, and this way information about separate velocity flows is lost. This may account for a part of the uncertainty in the higher density regions at low redshifts — which is where shell crossing occurs.

In particular, the detection of velocity vorticity suffers from this limitation. Vorticity is generated in multi-stream regions, and its detection is dependent on a proper separate treatment of the streams involved. The quality of the vorticity results in this study — section 6.9 — show how little can be achieved without such a treatment.

As explained in more detail in section 2.9, Hahn et al. (2015) have developed a method to disentangle the particles from different velocity flows, based on the Lagrangian position of all particles. Interpolating their properties is done between particles that are close in Lagrangian, rather than Eulerian space. This method is illustrated in figure 12 (see also Abel et al., 2012; Hahn et al., 2015). A future study in which this algorithm is applied to a high-resolution data set like Cosmogrid is expected to yield far better vorticity measurements.

The fact that the DTFE estimator does not respect the phase space configuration of velocity flows has another prominent consequence, and that is the *double spined* appearance of massive walls in many velocity-related fields. These artefacts can be seen in figures 32, 36, 51, and 57. They are the consequence of shell crossing that occurs in those regions at low redshifts. An estimator that properly disentangles different velocity flows may result in velocity fields where these double spined artefacts do not appear.

Deformation Eigenvalues

The statistical distribution of the largest eigenvalues exhibits an unexpected and unexplained dent around $1 - \frac{\lambda_1}{haf} \in \{0.2; 1.0\}$. This dent is precisely compensated by bumps in the other eigenvalue distributions, resulting in a smooth $\sum_i \lambda_i$ distribution. This effect persists throughout all redshifts, but occurs at slightly higher eigenvalues at later times. A decomposition of web components reveals that the decrease in λ_1 samples occurs at lower eigenvalues for regions that are characterised by lower densities — i.e. voids. The density dependence of this effect matches the correlation between λ_1 and density: they have precisely the same slope. Towards low redshifts, this effect takes on a strange and unexplained shape in the $\lambda_1 - \delta$ plane. We have investigated a range of possible causes, but have not yet been successful in identifying it.

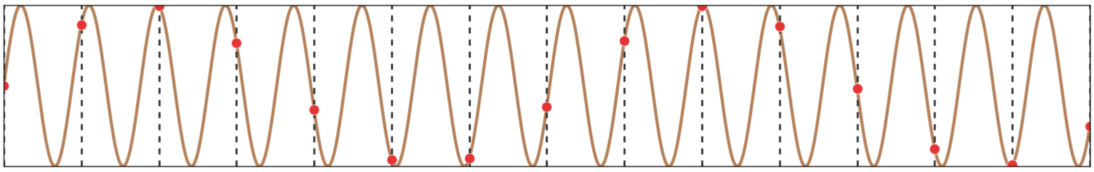


Figure 72: Aliasing illustrated. The orange sinusoid represents the original signal, the vertical dashed lines are the sampling points along one dimension. A sampling of the signal at the present resolution consists solely of the red markers. Since the signal's wavelength is shorter than the sampling width, the red markers appear to come from a waveform with the same amplitude but a longer wavelength Fourier component — about seven times the sampling width. As a result, original signal's wavelength will not appear in the discrete Fourier transform, and the samples wrongfully contribute to a higher wavelength component.

A.3 Fourier artefacts

Spectral leakage and aliasing

For a one-dimensional signal $f(x)$ that is discretely sampled at N points, regularly spaced a distance h apart, spanning a total extent $L = Nh$, the discrete Fourier representation consists of N — generally complex — Fourier coefficients at a range of wavenumber indices k. The number of Fourier coefficients is equal to the number of sample points, and they are ordered by wavenumber.

The smallest wavelength that can be represented is twice sampling resolution, and this is called the *Nyquist* limit. If a wave component of the signal has a shorter wavelength than $2h$, the sampling of that waveform will contribute to the Fourier component of a higher wavelength. This is illustrated in figure 72, each sampling 'overshoots' the wavelength by a little, resulting in a new sinusoid. Such a distortion of the original signal is called *aliasing*[65]. To some degree, this effect will always emerge in the DFT of a signal consisting of components with wavelengths close to the sampling resolution.

Aliasing is a specific case of a broader phenomenon called *spectral leakage*. Spectral leakage occurs when a signal is convolved — see section 4.1 — with a certain convolution kernel. In the general case, this process creates spurious wave components, which contribute to the Fourier transform of the resulting signal. The operation of sampling can be seen as a convolution of the signal with a *Dirac comb*, which causes leakage in the form of aliasing — an effect spanning the whole sample. Convolution with other kernels — e.g. smoothing windows, see equations 108 and 109 — generally produces leakage that affects a smaller range. In this study, we have used a Gaussian kernel for the filtering of data at various length scales. While, in theory, this kernel produces leakage that affects all of space, this effect falls off quickly over distance — in real space, most of the leakage from any point emerges very close to that point, so that damage to the data is limited.

[65] The original signal and the slower sinusoid described by the samples are said to be 'aliases of each other', as they are both sampled precisely the same way.

Cosmogrid fundamental mode

For the representation of small scale structure in $f(x)$ — assuming the Nyquist limit is respected — there are relatively many small-wavelength indices where this structure can be encoded. This large quantity of information results in a high statistical certainty. For this reason, it is generally preferable to sample structures over a scale that appreciably exceeds the length scale at which those structures occur — provided a good sampling resolution. In the case of Cosmogrid, we see that the extent of the simulation volume — 30 Mpc, with a sampling resolution of 0.234 Mpc — is sufficient for a reliable representation of structure at, say, 0.5 Mpc.

For the representation of larger scale structures, though, the number of Fourier coefficients that can encode $f(x)$ is limited — at the same sampling resolution. Long wavelength components, encoded by fewer Fourier coefficients, are subject to greater uncertainty. In the case of Cosmogrid, the largest structures appear at a scale comparable to the sample length of 30 Mpc — the largest wavelength is called the *fundamental mode* of the spatial distribution. There are too few Fourier components available to accurately represent such large scale structures. This decreases the statistical certainty of measurements from large scale structures.

In figures 50 and 51, it can be seen that the spatial distributions of λ_2 and λ_3 are subject to an artificial enhancement of horizontal and vertical correlation, respectively. These effects are a consequence of the fundamental mode problem: The longest wavelength Fourier coefficients in each direction $\mathbf{k}_x, \mathbf{k}_y, \mathbf{k}_z$ are only few in number, and therefore of limited accuracy. This introduces the risk that the Fourier power in one of them visibly — and unduly — exceeds that in the other directions. When the largest wavelength component in one direction is spuriously amplified, the resulting spatial distribution of $f(\mathbf{x})$ is subject to an artificial wavelike enhancement in that direction.

On top of the artificial enhancement of long-wavelength components, there is an arbitrariness in the directionality of these effects. Since the data are sampled on a regular cubic grid, they can only occur in the spatial orientations of the grid axes.

Appendix B Inventory of findings

As summarised in section 8, this study has yielded insights and confirmations relating to hierarchical structure evolution; relations between density and divergence; the decomposition of the cosmic web; the formation and evolution of anisotropic structures; and the evolution of statistical distributions of various quantities. This appendix aims to provide a more complete picture of these conclusions.

Evidence of hierarchical structure evolution

- A visual assessment of the density field reflects the expected hierarchical evolution of cosmic structures by the merging of small scale structure into larger features. Mergers occur both in massive regions — i.e. nodes, filaments and walls — as in voids.
- It appears that the early stage of structure formation is characterised by (i) a transition where expanding void regions become more spatially dominant than filaments; and (ii) a temporary increase in the prominence of oblate collapsing void regions. Both of these indicators appear to be coupled, and occur at later times when larger spatial scales are considered. This suggests that small scale structures mature before large scale structures do.
- A study of lognormal fit parameters supports the expectations of hierarchical evolution as well. We find that the general redshift-evolution of density distributions match very closely between different spatial scales. However, changes systematically occur earlier at smaller scales and later at larger scales. The same is found in fits to divergence and shear magnitude distributions — and even in vorticity magnitude distributions.
- By comparing shear magnitude maps from different redshifts and filter scales, we gather insights in the evolution of anisotropic structures from different scales. The small scale structure undergoes the greatest metamorphosis, while large scale structures remain relatively stable. The merging of large scale structures involves only relatively small displacements. At smaller scales, we see more extensive changes. The structures undergo many mergers, where structures get displaced over distances that are large in terms of the typical structure size.

Density-divergence relations

- At high redshifts, the density and divergence distributions match well, but discrepancies increase over time.
- In agreement with gravitational instability, the velocity divergence field follows the peaks and troughs in density. The general correlation between density and divergence is very clear, particularly at high redshifts.

- However, past some density threshold, velocity convergence no longer increases along with the density contrast. This is also manifested in the density and divergence distributions — the density distribution has an elevated long tail, but no similar high-convergence tail is found. The divergence distributions taken from the whole cosmic volume appear to be better modelled by a lognormal curve than the density distributions.

- Similarly, the divergence in voids levels off, where the density continues to lower levels. In other web components, the two quantity distributions appear to match more closely.

- A study of lognormal fit parameters to the divergence distributions indicates that they evolve somewhat differently from density distributions: the median convergence increases, while the median density does not.

Classification of cosmic web structures

- Oblate void regions can be seen ubiquitously flanking expanding void regions. Even if a threshold eigenvalue is used for detecting cosmic web structures, oblate features surround expanding void regions nearly everywhere. This is in agreement with expected scenarios where matter flows into walls and filaments towards nodes, in a shearing motion; as well as the draining of matter from oblate regions trapped between merging voids.

- The volume filling percentages of various cosmic web components vary according to the spatial scales considered. Notably, filaments are more prominent at large scales than at small scales.

- Concerning the relative volume occupation percentages of different cosmic web components, most of the shifts in balance occur at relatively high redshifts, on all spatial scales studied. This supports existing expectations that most of structure formation happens at early times.

- We find unexpectedly high volume occupation percentages of oblate collapsing void regions[66]. This result is to a great extent reproduced by a numerical experiment where the volume filling percentages of cosmic web components are simulated, based on Gaussian random velocity fields.

- The prominence of oblate void regions is attributed to an overestimation of collapse in our web classification algorithm. We have found that this can be remedied by setting a non-zero threshold eigenvalue (see Forero-Romero et al., 2009).

Formation and evolution of anisotropic structures

- Anisotropy in cosmic structures is well probed by the largest deformation eigenvalue λ_1. We find that the spatial distribution of λ_1 exhibits a very deep hierarchy of structure, more than the other eigenvalues. Notably,

[66]As well as a strikingly low volume occupation of prolate void regions.

filaments and walls appear clearly on the λ_1 maps, while the nodes and expanding void regions do not — they are typically more isotropic features. The spatial structures in λ_2 and λ_3 appear to follow more closely the density peaks.

- Velocity shear is a major aspect of anisotropic structure evolution. A visual spatial mapping of the magnitude of the shear tensor clearly shows anisotropic structures in high and low density regions, at all redshifts and spatial scales. This supports expectations of anisotropic structure evolution in dense and sparse environments.
- In the whole cosmic volume, there is a clear linear correlation between shear magnitude and density. Within each component separately, however, this correlation becomes unclear[67].
- The massive features occupy roughly the same ranges in shear magnitude. This shows the importance of shear in filaments and walls.
- In general, the shear magnitude distributions are rather well described by lognormal curves. Just as with density and divergence, it is the massive features that match a lognormal distribution best.
- In density, divergence and shear, we can see an intriguing substructure of walls. There are spikes in density, and an intricate substructure of filaments embedded in the face of the wall. Detecting this would not have been possible in a simulation of lower spatial or mass resolution.

Evolution of statistical distributions

In density:

- A decomposition of the density distribution into its different contributions from individual web components reproduces the findings by Hahn et al. (2007); Aragón-Calvo et al. (2010); Cautun et al. (2013)[68].
- With the passing of cosmic time, the density distribution broadens to both far lower and far higher values. All separate cosmic web components spread to higher densities, resulting in an increased overlap between the density distributions. This happens most notably in node and filament distributions.
- Sharp, localised peaks[69] are found in the density fields — more so at low redshifts than in the early stages. These cause an elevated high-density tail in the density distribution, which constitutes a clear deviation from the lognormal curve[70].
- A study of lognormal fit parameters shows that density distributions — in all web components separately — deviate from a Gaussian appearance over time: they acquire ever higher density peaks and ever larger regions of

[67] In other words, there is an *intra*-component correlation, but no *inter*-component one.
[68] Note that their web classification is based on the tidal field, while ours is based on velocity shear and divergence.
[69] Not only at the nodes of the network, but along filaments also.
[70] A study by Uhlemann et al. (2015) has provided analytical support for this observation.

lower density. This reflects the unbounded growth of the density contrast. This effect is more pronounced in voids than in massive region types. A decrease in fit qualities towards low redshifts indicates that densities deviate from a lognormal distribution. The filaments appear to follow a lognormal model the best.

In divergence:

- The divergence distributions from different cosmic web components — particularly at high redshifts — are visually well reproduced by our separate numerical experiment. We have generated divergence distributions from a random Gaussian velocity field, and they assume the same appearance. In our divergence measurements, there are continuities between the distributions from regions that share the same eigenvalue, but differ in the sign of $\sum_i \lambda_i$. Our numerical experiment also reproduces these continuities in divergence distributions.

- The divergence distributions from massive regions appear to be better described by lognormal curves than those from void regions.

In velocity shear:

- There are signs of a nonlinear shear evolution towards redshift 0.0, where shear magnitude growth accelerates somewhat. In the nonlinear regime, velocity shear measurements do not scale with $Haf(\Omega)$, but grow slightly faster.

- The shear magnitude distributions at $z = 0.0$ show very high measurements, relative to the densities at that time frame. This visual discrepancy between the distributions from $z = 0.0$ can in part be attributed to the nonlinear shear evolution.

- When regarding solely the shapes of the distributions, and not their position and horizontal extent, the density and shear magnitude distributions appear to be a rather well matched. At redshift 3.7, both distributions appear relatively close to a lognormal curve; at redshift 0.0 both have been stretched and acquired a more shallowly decreasing high-value tail. This observation yields potential insights into a proper normalisation for the nonlinear shear measurements.

- The enhancement of the high shear magnitude tail is in part due to large shear measurements occurring in walls and filaments.

- Over the course of structure formation, we find a significant increase of shear values in oblate void regions, walls, filaments and nodes.

- At all considered scales, the shear magnitude distributions from oblate void regions have an enhancement of at $|\sigma| \in \{0.3; 0.8\}Haf$ — a range that is associated with low density areas. This can be attributed to the draining of matter from underdense walls between merging voids.

- We studied the redshift evolution of parameters of lognormal fits to the shear magnitude distributions. The results indicate a shift in balance towards high shear measurements as cosmic time progresses. Different web components mostly have the same changes in fit parameters — there is a mostly parallel evolution of quantities in all types of regions.

And in vorticity:

- Vorticity measurements are dominated by an artificial *projection effect*, which is tightly coupled to the local density. This results in high measurements in dense areas, but no significant detection of vorticity.

- We find great discrepancies between the statistical distributions of density and vorticity measurements. There is a certain correlation between the two, but this may come entirely from the projection effect.

Velocity fields: probes and agents of Cosmic Web evolution

We have studied the density and velocity fields from the Cosmogrid simulation (Ishiyama et al., 2013). Our methods include a Delaunay Tessellation Field Estimation (DTFE) of density and velocity fields from the simulation particles. We decomposed the velocity gradient into divergence, shear and vorticity, and classified six different components of the cosmic web on the basis of the eigenvectors of the deformation tensor.

This thesis presents the spatial and statistical distributions of these quantities, and a decomposition of these fields into the contributions from different web components. We have studied the correlations between density and various velocity-related quantities, and followed the redshift evolution of parameters for the lognormal fits to the statistical distributions.

From these results, we find various forms of evidence for hierarchical evolution of cosmic structures; we determine the extent to which density and velocity divergence are correlated; we explore the formation and interactions between different structures that make up the cosmic web; we specifically probe the evolution of anisotropic structures; and follow the time evolutions of density, divergence, shear and vorticity. Lastly, we have identified a few artefacts resulting from the data and methods; and we assert the merit of the DTFE algorithm.