

# Dark matter haloes: a multistream view

Nesar S. Ramachandra, <sup>\*</sup> Sergei F. Shandarin,

*Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045*

17 May 2017

## ABSTRACT

Mysterious dark matter constitutes about 85% of all mass in the Universe. Clustering of dark matter plays the dominant role in the formation of all observed structures on scales from a fraction to a few hundreds of Mega-parsecs. Galaxies play a role of lights illuminating these structures so they can be observed. The observations in the last several decades have unveiled opulent geometry of these structures currently known as the cosmic web. Haloes are the highest concentrations of dark matter and host luminous galaxies. Currently the most accurate modeling of dark matter haloes is achieved in cosmological N-body simulations. Identifying the haloes from the distribution of particles in N-body simulations is one of the problems attracting both considerable interest and efforts.

We propose a novel framework for detecting potential dark matter haloes using the field unique for dark matter – multistream field. The multistream field emerges at the nonlinear stage of the growth of perturbations because the dark matter is collisionless. Counting the number of velocity streams in gravitational collapses supplements our knowledge of spatial clustering. We assume that the virialized haloes have convex boundaries. Closed convex regions of the multistream field are detected by making use of the Hessian of the multistream field. The halo candidates obtained from the multistream field are compared with the FOF and AHF halos.

**Key words:** methods: numerical – cosmology: theory – dark matter – large-scale structure of Universe

## 1 INTRODUCTION

The web-like distribution of matter initially revealed by redshift surveys (with less than 300 galaxies by ? and around 1000 galaxies by ?) and numerical modeling (using N-body simulations of around 30000 particles by ? and ?) pioneered morphological investigations of the cosmic web structures (see ?, also reviews by ? and ?). Detailed mapping of the Universe has crossed three million objects today, by catalogues such as the Sloan Digital Sky Survey (SDSS; ?). The upcoming Large Synoptic Survey Telescope (LSST; ?) is expected to probe the nature of dark matter using several billion galaxies. On the other hand, cosmological simulations have improved immensely in several aspects – numerical techniques, parallelization schemes, inclusion of various physical processes, volume and resolution (some of these developments are summarized in ? and ?). Modern state-of-the-art simulations like the Illustris Project ? and Q-Continuum ? use more than a billion dark matter particles. Finally, the ever improving data analysis techniques have resulted in new and sophisticated

density estimators, geometrical and topological indicators. A plethora of algorithms for identifying and characterizing dark matter structures have emerged in last two decades (a summary on cosmological data analysis is highlighted in ?). Considering all these improvements, it is worth noting that the proto-structures detected in the modern simulations are qualitatively similar to the quasi-linear description of clustering by Zeldovich Approximation (ZA; ?). Location and properties of these structures, i.e., the voids, walls, filaments and haloes maybe inconsistent across different structure finding algorithms, but that is primarily due to varied definitions.

Most of structure finders are halo finders only and majority of them are stemmed from three underlying algorithms. One of them is the SO (Spherical Over-density) halo finder that defines halos as spherical regions whose mass density exceeds the mean density by a specified factor (?). Another is the FOF (Friends-of-Friends) halo finder describing haloes as the groups of particles separated less than a specified linking length often chosen as 0.2 times the mean particle separation (?). The FOF can be also used for identifying filaments and walls/pancakes by increasing the linking length (?, ?, ?) Finally the DENMAX (DEN-

\* E-mail: nesar@ku.edu

sity MAXimum) halo finder assumes that the halos are the peaks of the density fields and thus selects the particles concentrated in the vicinity of the density maxima (?). One of the common features of these techniques is that all three are based on density, in one form or another. And all of them depend on free parameters that are chosen chiefly on the ‘merits principle’ (?) rather than on physics. Over the years all three kinds of the halo finders have been experiencing various modifications and improvements. A few examples from a long list of these modifications may include:

- (i) Improvised techniques of generation of the density field from the particle positions, and finding spherically bound over-densities (?, ?, ?, ?, ?, ? etc.)
- (ii) Adaptive methods controlling the linking length in methods using FOF (?, ?, ?, ?, ?, ? etc.)
- (iii) Adaptive methods for searching the positions of density maxima (For example, ?.)
- (iv) Generalization of FOF and DENMAX techniques to six-dimensional phase space, and many others (such as 6DFOF by ? and ROCKSTAR by ? use velocity-position space with parameters analogous to linking-length.)
- (v) Computing hierarchical tree of clusters in the phase-space such as the Hierarchical Structure Finder ?, and the 6-D Hierarchical Over-density Tree ?.
- (vi) Hybrid algorithms: frameworks such as the Hierarchical Bound-Tracing algorithm ? and SURV ? identify haloes at multiple time steps from the simulation to find prospective sub-haloes. In addition, there are HOP methods by ?, ? and ?.

A detailed comparisons of several halo/sub-halo finders is provided in ?, ?, ?. In a nice summary discussing these developments as well as describing a few new suggestions they concluded that there was no general consensus for a precise definition of a halo or a sub-halo. Consequently, there were different estimates of number of haloes, halo mass functions, halo centre locations, boundaries and other parameters.

There are significant concerns with SO, DENMAX and FOF algorithms - both in terms of underlying mechanisms of halo formation and the parameters used in halo identification. SO is motivated by the analytical toy model of the collapse of a top-hat spherical density perturbation. Parameters of the virial radii  $r_{vir}$  and virial mass  $M_{vir}$  are determined by the regions with density  $\rho_{vir} \geq \Delta_{vir} \times \rho_b$ , where  $\rho_b$  is the background density of the simulation box.  $\Delta_{vir}$  is generally taken around 180 or 200, derived for an isolated spherical collapse model, and it varies for different cosmologies and redshift. The peaks in CDM models not only aspherical, but their collapse is subject to tidal forces, mergers and presence of sub-structures - none of these complexities are weighed in the spherical collapse model.

For FOF, the free parameter of linking length is generally taken as  $b = 0.2$  times the mean separation of particles at  $z = 0$ . This inter-particle separation corresponds to  $\Delta_{vir} \approx 180$  if the halo has an isothermal density profile,  $\rho \propto r^{-2}$ . Using percolation theory, ? argued that this corresponds to a rather wide range of over-densities depending on halo mass and density profiles. They found out that  $b = 0.2$  corresponds to local over-density  $\delta$  within the enclosed halo to be in the range of 250 to 600. Moreover, the resulting FOF-haloes need not have a compact geometry: often the haloes are irregularly shaped, which is unlikely if the haloes are virialized. Hence modern algorithms

re-define the halo boundaries by excluding particles using post-processing techniques. In recent simulations with clear delineation of walls and filaments ?,  $b = 0.05$  was used for finding FOF-haloes since the traditional value of  $b = 0.2$  corresponded to structures that percolate into the web structure.

Absence of dynamical traits in the FOF and SO algorithms are arguably more crucial. In phase-space, the halo collapse models show collisionless DM particles in oscillatory motions about a core, at successive foldings of the phase-space sheet. The velocity field within each oscillatory spiral is multi-valued in physical space. Incrementing multistream shells, separated by caustic surfaces sequentially trace the structures of the cosmic web - walls, filaments and the haloes. Majority of the mass accretion into the haloes along the filaments: from lower multistreams into higher. Thus the DM haloes are not independent of filaments around them, and the hierarchical layers of multistreams represents this portrait precisely. This picture of structure formation was initially theorized using ZA ? and in context of caustics ? as well as in the Adhesion Approximation (?, ?). ? reviewed gravitational evolution of density perturbations in the context.

It has been demonstrated that the multistream field in Eulerian space can be computed directly from the Lagrangian sub-manifold (? and ?). About 90% of the field is single-streaming voids, and the rest of the volume comprises of multistream walls, filaments and haloes. ? found the multistream value of  $n_{str} \approx 90$  corresponds to virial density  $\Delta_{vir} = 200$ . On the other hand, DM particles are identified by (?) as belonging to haloes if they undergo flip-flop along 3 orthogonal axes. These analyses have opened up a new avenue in studies of halo formation, both qualitative and quantitative. Re-investigations of halo spins, physical radii of the halo, sub-structure in the light of streaming phenomena have shown that the halo structures and formations are more complicated than previously believed. ? investigated the distribution of streams in small haloes at various redshifts. They concluded that tracking caustics and streams is better than density, since density fields are noisy in the dense inner regions of haloes. In another study, ? argued that the ‘splashback radius’ - distance from the halo core to the first caustic enumerated from outside - is a better physical indicator of DM halo boundary than the virial radius (also see the discussion on turn-around radius of bound objects by ?). ? also agree with the view that the locally overdense regions correspond better with the volumes within the first caustic than the virialized DM clumps. Recent toy model of anisotropic halo collapse by ? considers intersecting multistream filaments forming spinning nodes. ? showed that the virial surfaces of FOF haloes have varying number of streams, including single-streams. Study by ? delineated the rich sub-structure of haloes using another derivative parameter from the Lagrangian sub-manifold called the ‘flip-flop’ defined on the Lagrangian space.

In this paper, we identify potential haloes by utilizing multistreaming as the governing dynamical phenomena. Review of the DM particle clustering in a one-dimensional dimensional universe is made in Section 2, with a description of multistream fields. The cosmological simulation is described in Section 3. In Section 4, we prescribe a framework to iden-

tify dark matter haloes using Hessian eigenvalues in the multistream field. We also discuss a few properties of the haloes. Comparison of the multistream haloes with AHF and FOF algorithms is done in 6. We also discuss the spatial distribution of the dark matter haloes along the percolating structure.

## 2 PHASE-SPACE REPRESENTATION OF GRAVITATIONAL CLUSTERING

### New section

We begin with a simple illustration showing the formation of a few haloes in a one-dimensional simulation. Dark matter clustering in a (1+1)-dimensional phase-space ( $\mathbf{p}, \mathbf{x}$ ) (where  $p$  is the momentum and  $x$  is the co-moving Eulerian coordinate) at four successive time steps is shown in the top panels of Figure 1. The lower panels show the corresponding multistream field ( $n_{str}$ ) (red) and density field  $\rho(\mathbf{x}, z)$  (gray). At  $z_1$  (left-most panel), velocity is single-valued in Eulerian co-ordinates shown, except at a small three-stream region near  $\mathbf{x} = 5\pi/4$ . This is the first instance of multistreaming in the region, which was previously had  $n_{str} = 1$  throughout. The interface of  $n_{str} = 1$  and  $n_{str} = 3$  regions is also the location of the first caustic. On the other hand, the density calculated at a high resolution shows variations, even in the mono-streaming regions. The variations are especially more pronounced around the caustic (near  $\mathbf{x} = 5\pi/4$ ).

The gravitational clustering is more evolved in the two center panels ( $z_2$  and  $z_3$ ) with three prominent phase-space spirals. The regions between the spirals have sparsely distributed dark matter particles, and have  $n_{str} = 1$ . Each spiral corresponds to a location of gravitational collapse with  $n_{str} > 1$  region, and higher density. A few of these regions within three-streaming regions are elevated to  $n_{str} = 5$ . The corresponding density field is not only noisier, but also reaches very high values at the caustics. This is also a primary distinguishing feature between mass density fields and multistream fields: At the locations of caustic, the density (regardless of how it is calculated) is not smooth. Computational limitations on simulation resolutions and refinement of density calculations soften the fields, exceptionally at the zero volume measure regions of caustic surfaces. On the other-hand, multistream values are increased by finite values at caustic surface locations - This property is preserved at higher simulation resolutions and any refinements of multistream field calculations - although  $n_{str}$  may be resolved enough to have intermediate even-values. Multistream fields are also intrinsically discrete valued, which is not true with density fields. Discreteness of multistream fields is discussed in more detail in ?.

The right-most panel in Figure 1 shows the final structure at  $z = 0$ . Two large spirals have spatially merged. These collapse environments are naturally very complex, with an increased number of successive caustic formation and merging. The corresponding velocity streams also show a more complicated structure. Clearly, the multistream field has a saddle point that is not as low as  $n_{str} = 1$ . This poses a bigger problem in the context of most of halo detection algorithms, and we discuss this in Appendix A.

### 2.1 Collapse in higher dimensions

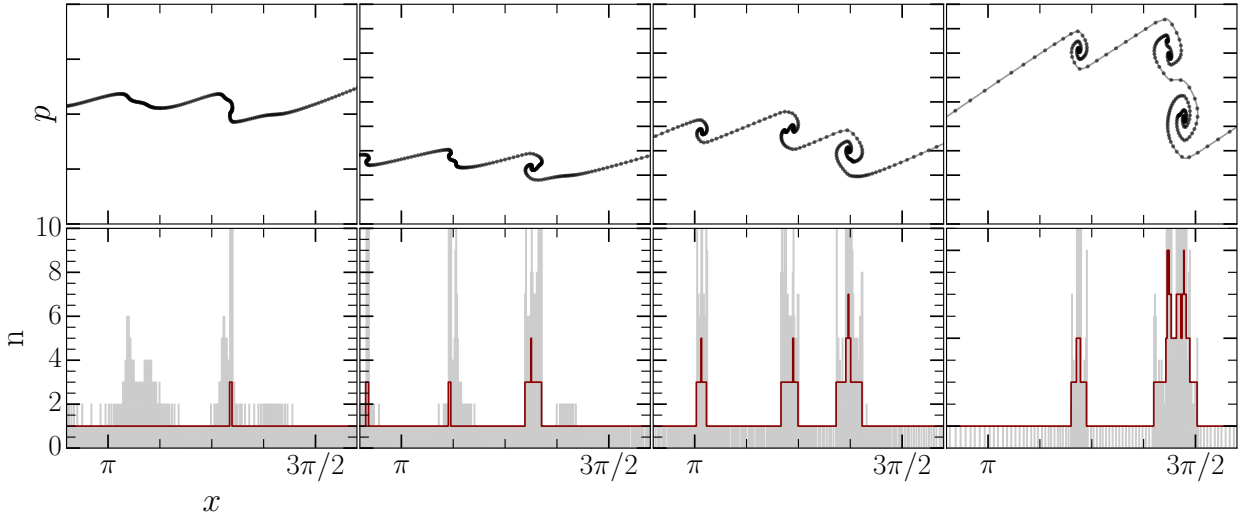
Extending the above results of one-dimensional collapse into higher dimension is vital, primarily in the context of halo formation. The individual spiral collapses in the one-dimension happen at a small region (left-most panel in Figure 1), and the region grows by volume, whilst increasing the spiral twists within. This is in contrast with the theoretical top-hat spherical model of halo formation when the shell crossing would not happen until the final moment of the collapse of the entire halo into a point-like singularity. Thus all shell crossings happen at a single point and at a single instant of time. The collapse of an isolated, spherically symmetric density peak is a very exceptional case, because every spherical shell feels only the force due to interior mass until it collapses into the caustic region. The collapse of the real peak proceeds in the field generated by the mass distribution - in both the mass within the forming halo, and the mass outside the halo.

The collapse of a uniform ellipsoid also results in a simultaneous collapse of the entire ellipsoid however this time not into a point but into a two-dimensional ellipse (?). Another customarily used spherical model of halo formation by ? and ? does not consider the initial collapse at all. Instead it assumes self-similar initial conditions and the halo at advanced stage with formally infinite number of spherical caustic shells.

The ‘core’ in a collision-less dark matter collapse (in Figure 1) is a region where a multistream region is first formed due to caustic formation. This is conceptually similar to a shell crossing. However, there are successive caustic formations that follow the first shell crossing, and they are not limited to the halo cores. Each caustic increases the multistream value within by a finite number. The cores of the multistream haloes obviously have the local maxima of velocity streams in Eulerian coordinates. On the contrary, mass densities have infinite values at the caustics surfaces, including the core. Discontinuities in densities at these regions of sharp multistream transitions are clearly seen if the mass and spatial resolutions were sufficiently high (see two-dimensional simulations by ? as well as in three-dimensional simulations by ?, ?, ? etc.).

In three-dimensional simulations, the Lagrangian sub-manifold twists in complicated ways in a six-dimensional phase space. This is due to complexities involving caustic formations in higher dimensions, which is true even in the ZA, see ? and ? for detailed analyses of caustic formation. The resulting intricate geometrical structures can be characterized by the  $n_{str}$  field. Nearly 90% of the volume in N-body simulations are single-streamed voids at  $z = 0$  (? , also see ? for a percolation analysis of single-streaming voids). From the visualizations in ? and percolation analysis of ?, we also know that the  $n_{str} = 3$  regions mostly form connected wall-like structures, unlike the isolated patches as seen in one-dimensional simulations of Figure 1. The structures become predominantly filamentary at higher thresholds of  $n_{str} \gtrsim 17$ . Subsequently, the regions around the multistream maxima have isolated closed surfaces (for example, in Figure 2), which may be identified as halo locations.

Caustic formations and mass accretion are also seen to occur more along the higher streams, which makes the haloes non-spherical, with the alignment generally determined by a



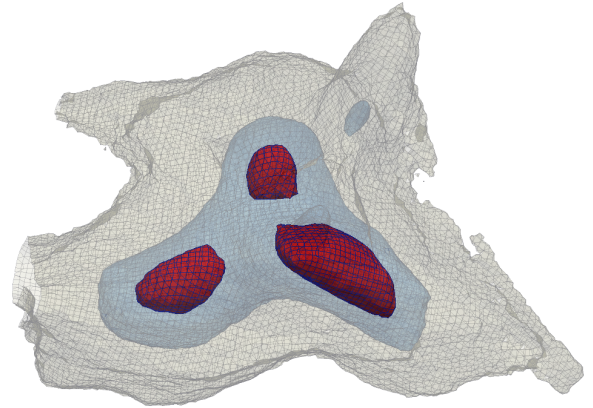
**Figure 1.** Dynamical collapse of dark matter in one-dimensional universe: Top panels show the  $(\mathbf{p}, \mathbf{x})$  phase-space manifold of the dark matter sheet at redshifts  $z_1$ ,  $z_2$ ,  $z_3$  and  $z = 0$ . Dots represent the dark matter particles. The momentum values are chosen at arbitrary scales. Bottom panels show the corresponding multistream field  $n_{str}(\mathbf{x}, z)$  (red) and density field  $\rho(\mathbf{x}, z)$  (gray).

complicated interplay of the intensities of the streams from neighboring filamentary structures. Number of streams corresponding to the dark matter halo also has a local environment dependence. The three small haloes in Figure 2, where the number of streams are higher than the neighbouring filaments, are aligned along three intersecting filaments. Halo environment studied in ? show similar hierarchical variation in  $n_{str}$  values. The halo environments are thus very complicated, and empirical thresholds (on multistream or density fields) may not account for all the haloes uniformly. Hence we use a local geometrical method to identify potential haloes in multistream fields.

The first non-linear structures in the web have  $n_{str} = 3$ . By visual inspection, these regions generally form a fabric-like open structures that resemble walls. The surface contours of higher  $n_{str}$  are embedded within the walls. Figure 2 shows a filamentary structure of the web at  $n_{str} \gtrsim 17$ . The figure also shows regions around local maxima of the multistream field, which are generally located at the intersections of filaments.

### 3 SIMULATIONS AND TOOLS

The emphasis of this paper is to demonstrate the use of multistream field in identifying potential dark matter haloes, and not a full statistical analysis of halo properties. For this purpose, we have run simulations at two different mass resolutions (number of particles  $N_p = 128^3$  and  $256^3$ , and respective mass of particles,  $m_p = 3.65 \times 10^{10} h^{-1} M_\odot$  and  $4.57 \times 10^9 h^{-1} M_\odot$ ), with the same periodic side length  $L = 100 h^{-1} \text{Mpc}$ . The gravitational softening length  $\epsilon = 20 h^{-1} \text{kpc}$  and  $10 h^{-1} \text{kpc}$  for low and high resolution simulation respectively. The initial conditions are generated by MUSIC (?) with the transfer function from ? at a redshift of  $z_{ini} = 80$ . The  $\Lambda\text{CDM}$  cosmological simulation is run using GADGET-2 (?) and ? is similar to the ones used in ?. The cosmological parameters used in the simulation are  $\Omega_m = 0.276$ ,  $\Omega_\Lambda = 0.724$ , the Hubble parameter,  $h = 0.703$ ,



**Figure 2.** Multistream field contours: The multistream field is calculated at 8 times the native resolution. Each inner convex blobs (red) surround local multistream maxima inside. Surrounding outer shell (blue) is not convex throughout the surface, and the outermost gray multistream surface displays a filamental geometry.

the power spectrum normalization,  $\sigma_8 = 0.811$  and the spectral index  $n_s = 0.961$ .

Multistream field  $n_{str}(\mathbf{x})$  is calculated on the GADGET-2 snapshots at  $z = 0$  using the tessellation scheme by ?. The multistream field can be computed at the native resolution of the Lagrangian grid of the simulation, i.e., at refinement factor of  $l_i/l_d = 1$  (where  $l_i$  is the inter-particle separation in Lagrangian grid and  $l_d$  is the side length resolution of diagnostic grid). Arbitrarily high refinement factors can be utilized in computing multistream fields as well, for example  $l_i/l_d = 8$  for the halo multistream environment shown in Figure 2. For analysis of full simulation boxes, we restrict  $l_i/l_d$  to 1 and 2.

Two halo finders are also used to identify potential haloes with 20 or more particles at  $z = 0$ : a classic



Friends-of-Friends method (FOF-?) using a popular linking length,  $b = 0.2$  (e.g. ? and ?) and the Adaptive Mesh Investigations of Galaxy Assembly (AMIGA halo finder or AHF-?, ?). Halo catalogue from these halo finders are used to compare with our implementation of halo detection in the multistream field. The haloes candidates from AHF and FOF algorithms are hereafter referred to as AHF-haloes and FOF-haloes respectively.

#### 4 HALOES IN THE MULTISTREAM FIELD

We intend to identify haloes in the  $n_{str}(\mathbf{x})$  field instead of using just the position coordinate data. While the eigenvalue analysis itself is done at a chosen time, the multistream field inherently has data from six-dimensional Lagrangian space  $(\mathbf{q}, \mathbf{x})$  that contains the full dynamical information, similar to the phase-space sheet albeit in a different form. Dynamical history that is embedded in the multistream field is crucial in understanding the strength of gravitational binding of the particles. A physically motivated distinction between void and gravitationally collapsed regions – voids are the regions with a single stream – is a unique feature of multistream analysis ( ? and ?). Thus the haloes detected from local maxima of the  $n_{str}$  field can be ensured to be away from the mono-streaming voids. Methods based on linking-length or density fields may not be able to ensure that all the particles in haloes are away from voids (as shown for FOF haloes in ?).

Numerical analyses of scalar fields generally depend on resolution as opposed to particle coordinates analysis tools like FOF. The multistream field conveniently has an advantage of being less noisy than mass density ( ?, also see the Appendix in ? ).

##### 4.1 Hessian of multistream fields

###### New sub-section

Hessian matrix  $\mathbf{H}(f)$  of a scalar field  $f$  involves local second-order variations in three orthogonal directions. Each element of the Hessian matrix  $\mathbf{H}_{ij}(f)$  (where  $i$  and  $j$  can be any of  $x$ ,  $y$  or  $z$  directions) is given by Equation 1.

$$\mathbf{H}_{ij}(f) = \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (1)$$

By choosing a  $f = -n_{str}(\mathbf{x})$  (smoothened if necessary), local multistream variations can be diagnosed. The Hessian matrices at each point on the configuration space are always symmetric matrices, resulting in real eigenvalues. The Hessian eigenvalues in multistream fields differ from that in density, gravitational potential or velocity shear tensor. We refer the readers to ? for an extensive analysis on multistream Hessians and their geometrical significance. Some of the salient features of Hessian eigenvalues of multistream field are as follows:

(i) Every element of Hessian matrices  $\mathbf{H}(-n_{str})$ , and consequently the eigenvalues  $\lambda_i$ 's are zero in single-stream voids. Even if the multistream field is a smoothed, the eigenvalues peak at zero. This property is unique to multistream

fields. Eigenvalues of Hessians of density, velocity shear tensor ? do not peak at zero, and the eigenvalues of deformation tensor are negative in voids (by Zeldovich formalism).

(ii) The eigenvalues of these Hessian matrices are always real, and depending on if their values are positive or negative, one may infer local geometrical features in the field. For our choice of  $-n_{str}(\mathbf{x})$  as the domain of Hessian, at least in principle, the conditions for geometric criteria are:  $\lambda_1 > 0 > \lambda_2 \geq \lambda_3$  for locally flat regions,  $\lambda_1 \geq \lambda_2 > 0 > \lambda_3$  for locally tubular structures and  $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$  for clumped blobs.

(iii) Convex neighbourhoods around local maxima of the multistream field are isolated by the positive definite Hessian matrices.

(iv) The resulting Hessian eigenvalues characterize the geometry in a four-dimensional hyper-space of  $(-n_{str}, x, y, z)$ . The boundary of a region with  $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$  is a closed convex contour in this hyper-space, and thus its projection on to the three-dimensional Lagrangian space is also closed and convex.

Of the three geometries that are characterised by the eigenvalue conditions, we investigate the convexity of multistream contours in the context of halo finding in the section below.

##### 4.2 Halo finder algorithm

Our goal is to isolate the locations of convex geometries in the multistream flow field. Prospective regions of the halo locations in the web structure are selected by positive definite condition of the Hessian  $\mathbf{H}(-n_{str})$ :  $\lambda_1 > 0$ ,  $\lambda_2 > 0$  and  $\lambda_3 > 0$ , or simply the smallest of the eigenvalues,  $\lambda_3 > 0$ . We also filter out the regions if the multistream values inside do not suggest gravitational collapse into haloes. The sequence of our halo detection framework is listed below:

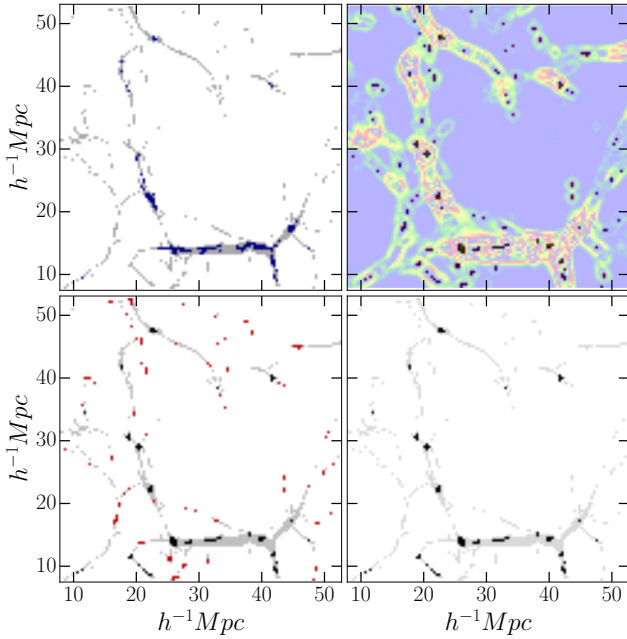
(i) The multistream flow field is calculated on a diagnostic grid. The number of tetrahedra that encompass each vertex in the grid gives the  $n_{str}$  field. Top left panel of Figure 3 shows the multistream web structure in a slice of the simulation with  $n_{str} > 1$  in gray and  $n_{str} \geq 7$  in blue.

(ii) The discrete multistream flow field is smoothed in order to reduce numerical noise. We have used Gaussian kernel for smoothing in our analysis. Effect of smoothing scales in the halo identification is discussed in Section 4.3.

(iii) Second order variations of the smoothed  $-n_{str}(\mathbf{x})$  is computed at each point in the field. This gives symmetric Hessian matrices for this field whose eigenvalues are real. Ordered eigenvalues of the Hessian,  $\lambda_1 \geq \lambda_2 \geq \lambda_3$  are calculated. The  $\lambda_3$  field is shown in the top right panel of Figure 3.

(iv) Using segmentation techniques, each region with  $\lambda_3$  strictly greater than 0 within  $n_{str} \geq 3$  regions of multistream field are isolated and labelled. This condition for each halo candidate guarantees that it is in the region where at least one gravitational collapse happened within the halo boundary. Mass particles belonging to these regions are shown shown as dark spots in the top right panel of Figure 3.

(v) The multistream field has a range of values within the isolated sites. We impose constraints on the isolated regions to rule out the labels with low multistream values. The local maxima of  $n_{str}$  inside each halo must be at least 7. By this restriction, it is ensured that the halo sites with three



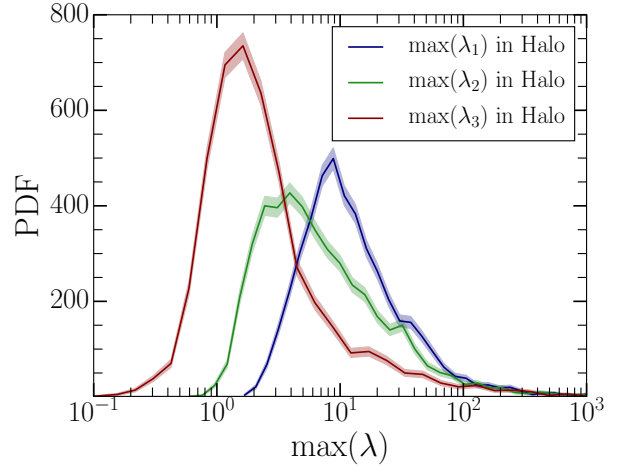
**Figure 3.** Detection of potential halo candidates in the multistream field: algorithm of segmentation and filtering are illustrated in a smaller slice of  $40h^{-1}Mpc \times 40h^{-1}Mpc$  slice of the simulation box. Top left figure shows the multistream field of the slice. Voids (white) are the regions with  $n_{str} = 1$ , rest are non-void structures. Blue patches within the structure (gray) are the regions with gravitational collapses in more than one direction, i.e.,  $n_{str} \geq 7$ . Top Right figure shows the smallest eigenvalue  $\lambda_3$  field. The value of  $\lambda_3$  is close to 0 in most of the regions (yellow), including the voids. Regions with  $\lambda_3 > 0$  and  $n_{str} > 1$ , are isolated (black spots) using image segmentation techniques. Bottom left panel shows the filtering scheme: the red patches do not have maxima of  $n_{str} \geq 7$  in the regions, hence are filtered out. The remaining potential halo regions with more than 20 particles are shown in the bottom right panel.

Lagrangian sub-manifold foldings are selected. Bottom left panel of Figure 3 shows patches that are discarded in red. The resulting  $\lambda_3$ -haloes are shown in the bottom right.

(vi) In our comparisons with other halo finders in Section 6, we also used an additional constraint on the minimum number of mass particles in the haloes to be 20 - which is generally used as a criteria in several halo finders.

For the illustration halo detection framework in this section, we have calculated the number-of-streams at refinement factor of 2 and smoothing scale of  $0.39h^{-1}Mpc$  (equal to the grid length of the multistream field) for the simulation box of  $128^3$  particles and size  $L = 100h^{-1}Mpc$ . Hessian matrices and eigenvalues are calculated on the same diagnosis grid. Results of the halo detection scheme for simulation box of higher mass resolution, and different smoothing factors are discussed in Sections 6 and 4.3. Hereafter we refer to the potential dark matter haloes detected from the Hessian analysis of the multistream field as  $\lambda_3$ -haloes for brevity.

Applying the above scheme on the simulation with side length of  $100h^{-1}Mpc$  and  $128^3$  particles (with cosmological parameters mentioned in Section 3), we detected approximately 50000 regions satisfying  $\lambda_3 > 0$  within the non-void in the multistream field of refinement factor  $l_l/l_d = 2$  and



**Figure 4.** PDF of highest  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  values in each of 4492 haloes detected by our algorithm. The peaks of the PDF are in the range 1-10. Shaded regions represent  $1\sigma$  error.

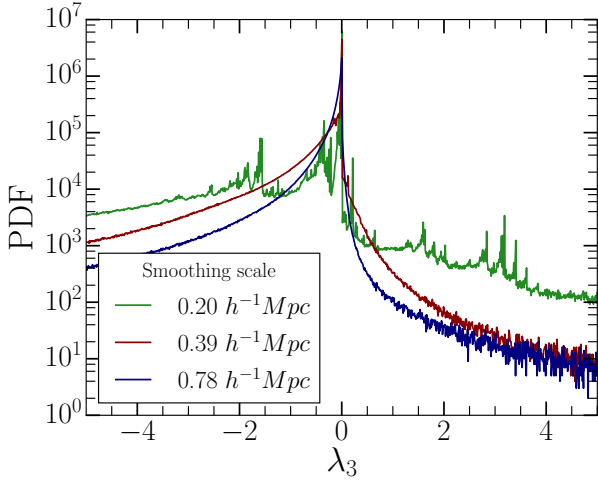
**Table 1.** Statistics of the Hessian eigenvalues in halo candidates: Local maxima of

Statistics	$\lambda_1$	$\lambda_2$	$\lambda_3$
Minimum	1.5	0.5	$1.3 \times 10^{-2}$
Maximum	$1.7 \times 10^3$	$1.5 \times 10^3$	$1.1 \times 10^3$
Median	10.5	5.5	1.9

smoothing scale of grid length, i.e.,  $0.39h^{-1}Mpc$ . We filtered out the segments with local maxima of  $n_{str} < 7$ , and around 14000 regions remained as prospective haloes. Majority of these regions have less than 20 particles, which are excluded in the halo catalogues. On the whole, our algorithm detected about 4500 haloes with more than 20 particles in the entire simulation box. We have not applied virialisation to define the halo boundaries. A more detailed study of halo edges, and comparison with that of FOF-haloes and AHF-haloes is done in Section 5. Here we concentrate on the three vital factors in our framework: local geometrical indicators  $\lambda_i$ 's, the softening scale of the field and multistream thresholds.

The maximum values of  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  in each of the haloes have peaks away from 0 as shown in Figure 4. The median values of  $\max(\lambda_1)$  and  $\max(\lambda_2)$  are in the range of 1-10 (Table 1), in spite of the threshold for  $\lambda_3$  being barely positive, by definition. Hence the interior of the potential halo segments is quite convex, with a local maxima inside. In some haloes, the local maxima of eigenvalue are in the order of thousands, as tabulated in Table 1.

With this algorithm, we obtain prospective dark matter haloes - regions with a local maximum of the multistream field in the interior of their closed convex surfaces. The haloes are detected without using density fields or linking lengths between particles. The parameters in the algorithm are entirely based on features of the multistream field and local geometry using Hessian matrices.



**Figure 5.** The distribution of  $\lambda_3$  in the simulation box of  $128^3$  particles and multistream field of refinement factor  $l_l/l_d = 2$ . Three smoothing scales are shown.

#### 4.3 Effect of smoothing

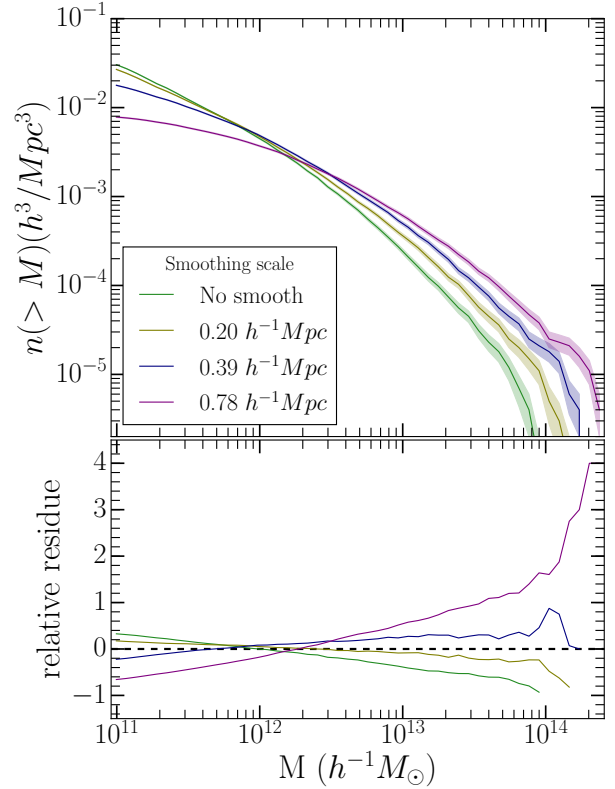
In order to reduce noise, the field is smoothed for our analysis using a Gaussian filter. The effect of smoothing scale on the distribution of the eigenvalue  $\lambda_3$  in the simulation of  $128^3$  particles is shown in Figure 5. Effect of softening on the multistream fields does not alter the distribution of multistream distribution significantly (Seen in Figure 10 of ?). However, the second order variation (and consequently the Hessian eigenvalues) is significantly changed due to the softening of the edges of structures. PDF of  $\lambda_3$  at multistream smoothing scale of the half the side length of diagnostic grid,  $0.5 \times l_d = 0.20h^{-1}$  Mpc is noisier than in the scales of  $l_d$  and  $2 \times l_d$ . However, at every scale, the PDF peaks at 0. The volume fraction of regions with  $\lambda_3 > 0$  (i.e. with positive curvature) is 2.4%, 2.3% and 2.5% for scales  $0.20h^{-1}$  Mpc,  $0.39h^{-1}$  Mpc,  $0.78h^{-1}$  Mpc respectively. For the detection of haloes in Section 4, we only look at these regions.

In addition to reducing the numerical noise, smoothing of the multistream field also results in softening of the substructures ?. Since our framework of detecting haloes isolates the multistream regions with local maxima, the closed curvatures are resolved separately. The halo or a sub-halo regions, that enclose the local maxima of the  $n_{str}$  field, vary with the smoothing scale of the multistream field. By increasing smoothing of the multistream field, the number of haloes are reduced as shown in Table 2. In the simulation with  $256^3$  particles, 27929  $\lambda_3$ -haloes are detected at smoothing scale equal to the diagnostic grid length,  $l_d = 0.20h^{-1}$  Mpc. The number of haloes decreases to 18221 and 7897 at softening scales of two- and four times the grid lengths respectively.

Moreover, since the spatial resolution is higher at the low softening, more small haloes are detected, as shown in lower mass regime of halo mass functions in Figure 6. The tail of halo mass functions reveal that large haloes are more massive for higher softening scales. For instance, the largest haloes for the same simulation with multistream softening

**Table 2.** Number of  $\lambda_3$ -haloes identified at smoothing of  $n_{str}(\mathbf{x})$  at different scales.

$N_p$	$0.20h^{-1}$ Mpc	$0.39h^{-1}$ Mpc	$0.78h^{-1}$ Mpc
$128^3$	5181	4492	2923
$256^3$	27929	18221	7897



**Figure 6.** Top panel shows Halo Mass functions of the potential  $\lambda_3$ -haloes in the multistream field refinement factor  $l_l/l_d = 2$  with various smoothing scales. Simulation box has  $256^3$  particles. Lower panel shows the deviation of the each halo mass function with respect to their average.

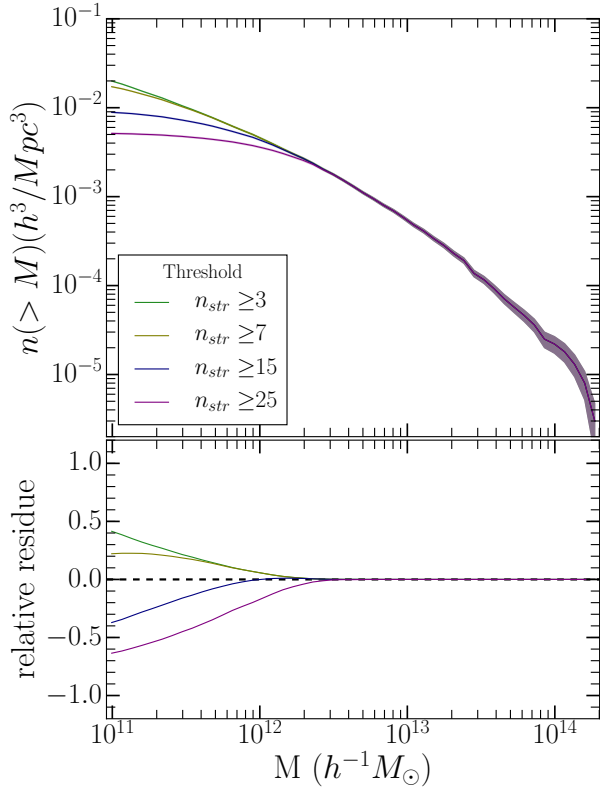
length of  $0.20h^{-1}$  Mpc,  $0.39h^{-1}$  Mpc and  $0.78h^{-1}$  Mpc have 30650, 38333 and 56257 particles respectively.

The sub-halo finder methods (see ? and references therein) identify substructures within a large host halo. The sub-haloes are resolved individually as  $\lambda_3$ -haloes at different scales from our algorithm if the local maxima of the smoothed multistream field is enclosed within the boundary.

#### 4.4 Effect of multistream thresholds

##### New subsection

Environmental dependence of the haloes results in various multistream values for the halo core. Theoretical toy models of halo formation, such as the tetrahedral collapse model ? describes a three-dimensional halo with four filaments accreting mass into it, has 15 stream crossings. ? have previously showed that a high threshold of  $n_{str} \geq 90$



**Figure 7.** Halo Mass functions of the potential  $\lambda_3$ -haloes in the multistream field refinement factor  $l_l/l_d = 2$  (in Simulation box with  $N_p = 256^3$ ) with various thresholds on local maxima of  $n_{str}$  within the halo. Lower panel shows the deviation of the each halo mass function with respect to their average.

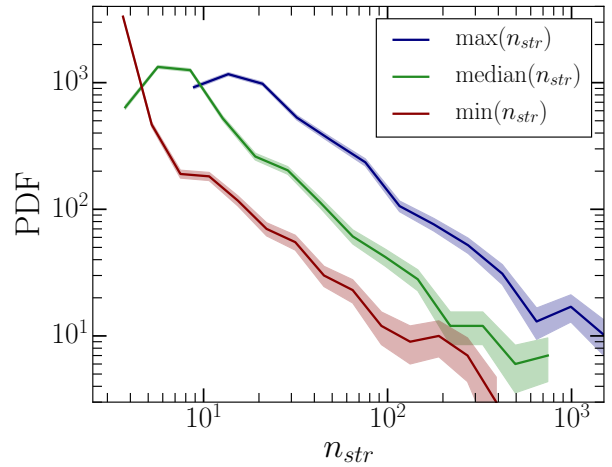
is equivalent of virial density of  $\rho_{vir} = 200$ , and filters most of the large haloes above  $10^{13} M_\odot$ .

The algorithm used for detecting multistream haloes initially detects all the closed regions in the multistream ( $n_{str} > 1$ ) regions of the cosmological simulation. In order to exclude some of the obvious non-halo sites, we impose a lower threshold of  $n_{str} \geq 7$  on the multistream maximum (these regions were also seen as parts of walls or filaments in ?), so that all the sites with three or more foldings in the Lagrangian sub-manifold are chosen. Combining this with the conditions on local eigenvalues, number of particles in haloes etc, we got a pretty good correspondence with other halo finders as demonstrated in section 6.

Although this condition is by no means strict, it is necessary to check the validity of the assumption. Figure 7 shows the halo mass functions for the haloes detected with changing thresholds on the multistream values of the halo cores. The figure demonstrates that increasing the cut-off from  $n_{str} \geq 3$  to  $n_{str} \geq 25$  systematically excludes small mass haloes while the number of haloes with  $M \gtrsim 2 \times 10^{12} M_\odot$  remains the same.

## 5 HALO PROPERTIES

Multistream environment of haloes can be very diverse. ? demonstrated that a majority of the FOF-haloes are in contact with the single-streaming voids. Illustration in ? also



**Figure 8.** Maximum, minimum and median of  $n_{str}$  in each of 4492 halo candidates. The closed contours of haloes encompass a wide range of multistream values. None of the haloes are in contact with the void region, since lowest value of  $\min(n_{str})$  is 3. Shaded regions are the  $1\sigma$  absolute errors in the number of  $\lambda_3$ -haloes.

**Table 3.** Local maxima and minima of  $n_{str}$  in each of 4492 haloes. The highest  $n_{str}$  values in the interior of haloes span over a large range of values. Low values of  $n_{str}$  in haloes, which are generally near halo boundaries, have a median of 3.

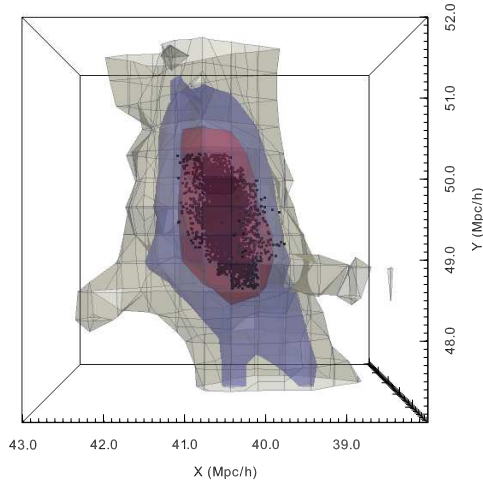
Statistics	$n_{str}^{high}(H_i)$	$n_{str}^{low}(H_i)$
Minimum	7	3
Maximum	2831	459
Median	17	3

shows that a large number of FOF-haloes have more than 10 per cent void on the spherical surfaces with virial radii. The  $\lambda_3$  haloes are significantly different: none of the  $\lambda_3$ -haloes are in contact with the regions where gravitational collapse has not occurred. This is guaranteed by the lower bound of  $n_{str} = 3$  on all potential halo candidates. Condition on the multistream field within the potential halo sites also ensures that there are collapses along more than one direction, which corresponds to  $n_{str} = 7$ . Hence by definition, for any multistream halo  $H_i$ , highest and the lowest multistream value are  $n_{str}^{high}(H_i) \geq 7$  and  $n_{str}^{low}(H_i) \geq 3$  respectively.

The potential haloes  $H_i$ s selected by eigenvalue condition  $\lambda_3 > 0$  have a local maxima of  $n_{str}^{high}(H_i)$  inside their boundaries. For a large number of these  $\lambda_3$ -halo candidates, the maximum  $n_{str}^{high}$  is higher than the bound of  $n_{str} \geq 7$ , as shown in Table 3 and Figure 8. For simulation with  $128^3$  particles, the median of this peak  $n_{str}^{high}(H_i)$  value is 17. Unsurprisingly, the global maximum of the multistream field ( $n_{str} = 2831$ ) is also a local maximum for one of the haloes. On the other hand, the median of lowest multistream value  $n_{str}^{low}(H_i)$  in the haloes is 3 (Table 3), which is also the first stage of non-linearity.

An important feature of our halo detection method is that the detected  $\lambda_3$ -haloes do not have a global threshold on  $n_{str}$  or mass density values. The local conditions may





**Figure 9.** Multistream environment of a  $\lambda_3$ -halo. The contours represent regions with 3 different multistream values: Outermost  $n_{str} \geq 25$  (gray) is tubular, The blue region has  $n_{str} \geq 75$ . The inner region is highly non-linear with  $n_{str} \geq 200$ . The black dots represent the mass particles belonging to a  $\lambda_3$ -halo, as detected by our algorithm.

be more suited in identifying haloes in multistream fields, since the multistream environments around haloes are very diverse. For instance, regions with  $n_{str} \geq 25$  are tubular around one of the large haloes in Figure 9. Even the region with more than 75 streams does not enclose a convex multistream region. Whereas, for  $n_{str} \geq 200$  the region is convex and the particles detected by our method reside mostly within. We detect closed regions in the multistream field as long as they are not in void, and have at least three foldings in the Lagrangian sub-manifold.

However, the  $\lambda_3$ -halo boundary is different from any constant multistream contour. That is, from the function  $n_{str}(\mathbf{x})$ , convex regions in the four-dimensional function space  $(-n_{str}, x, y, z)$  are projected onto three-dimensional co-ordinate space using eigenvalues. This is different from projecting ‘iso-multistream’ slice onto three-dimensional co-ordinate space. Appendix A illustrates the difference in the two approaches for a one-dimensional function.

The multistream field usually has concentric shells in the regions of successive gravitational collapses (as explained in Section 2 and Appendix in ?). In a specific scenario of Figure 9, regions of lower number-of-streams ( $n_{str} = 25$  and lower) is are tubular and have regions of higher  $n_{str}$  inside ( $n_{str} = 200$  and higher) that is closed. However, this transition from concavity to convexity of the multistream field does not occur at a constant value of  $n_{str}$  throughout the field. Instead, it is a local geometrical change that occurs at  $\lambda_3 = 0$ . For the  $\lambda_3$ -haloes in our simulation ( $N_p = 128^3$ ), minimum of multistream values  $n_{str}^{low}(H_i)$  within each halo has a range of values shown in Table 3 and Figure 8 – this varies between  $3 \leq n_{str}^{low}(H_i) \leq 459$ . Hence a global condition on  $n_{str}$  does not guarantee that the region is convex.

The particles in a massive  $\lambda_3$ -halo shown in Figure 9 form a spheroidal structure. The number of particles in similar massive haloes are in the order of  $10^3 - 10^4$  particles. For instance, the most massive halo in the simulation (with

**Table 4.** Number of haloes,  $N_H$  detected by the three halo finder algorithms in the two simulations of  $L = 100h^{-1}\text{Mpc}$  with different mass resolutions. Values shown for  $\lambda_3$ -haloes are done in the multistream fields with refinement factor of 2, and smoothing scale equal to the diagnostic grid size.

$N_p$	$N_H^{\text{AHF}}$	$N_H^{\lambda_3}$	$N_H^{\text{FOF}}$
$128^3$	3374	4492	5440
$256^3$	24710	27929	35765

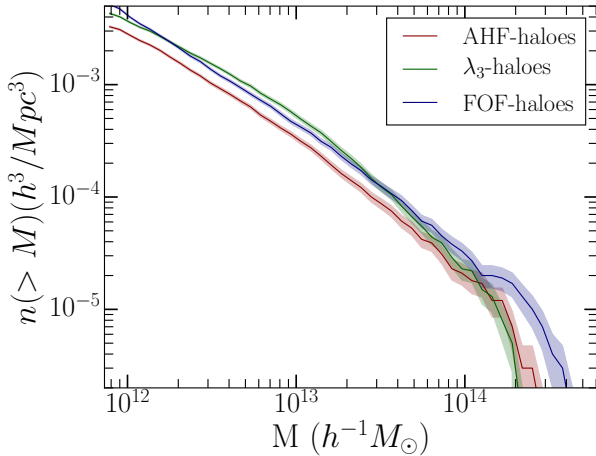
$N_p = 128^3$ ) has 5593 particles. We have chosen a minimum threshold of 20 particles, which is an artificial parameter (may be cooked up or ad hoc) used by most halo finder methods. Majority of the  $\lambda_3$ -haloes have low number of particles; median of number of particles per halo is 49. Each particle in this simulation is approximately  $3.65 \times 10^{10} h^{-1} M_\odot$ . Hence the halo mass range varies in the order of  $10^{11} M_\odot$  to  $10^{14} M_\odot$ . Combined mass of all the  $\lambda_3$ -halo candidates is about 31 per cent of the total mass in the simulation. In contrast, the haloes occupy just 0.3 per cent of the total volume. Thus the  $\lambda_3$ -haloes are extremely dense structures. Further analysis of halo mass function of  $\lambda_3$ -haloes and comparison with AHF- and FOF-haloes is done in the Section 6.

## 6 CORRESPONDENCE WITH OTHER HALO FINDERS

Comparison of haloes obtained from AHF and FOF method, along with our geometric analysis of the multistream field reveals several interesting features. The number of haloes ( $N_H$ ) with at least 20 particles that were detected by all the algorithms is shown in Table 4. For both the simulations, FOF detects the highest number of haloes and AHF detects the least. By applying the Hessian algorithm on multistream fields smoothed at the scale of diagnostic grid size,  $l_d$ , we detected around 4500 and 28000 haloes in simulations with  $128^3$  and  $256^3$  particles respectively. The number of  $\lambda_3$ -halo is close to the mean of AHF- and FOF- haloes in each simulation – around 2 per cent for the  $N_p = 128^3$  and 8 per cent for the  $N_p = 256^3$  simulation. Multistream field both the simulations we calculated at a refinement factor of  $l_l/l_d = 2$ .

The halo mass functions from all three finders are shown in Figure 10. For smaller haloes of order of  $10^{13} M_\odot$ , our method predicts a slightly higher number of haloes than FOF and AHF. For the most massive haloes of mass around  $10^{14} M_\odot$ , number of  $\lambda_3$ -haloes is fewer than the other 2 methods, albeit around the range of error of AHF-haloes.

By observing some of the massive haloes, like the one in Figure 11, we find that the  $\lambda_3$ -halo particles are within AHF- or FOF-halo region. This is generally observed in other massive haloes too: the large  $\lambda_3$ -haloes have fewer particles than corresponding AHF- or FOF-haloes. For haloes greater than  $10^{14} M_\odot$ ,  $\lambda_3$ -haloes have boundaries slightly within the AHF virial radius. Without unbinding, the FOF-haloes can be very large compared to other methods, as seen in Figure 11. This results in a deviation in the  $\lambda_3$ -halo mass function (Figure 10) from the other two methods over halo mass of  $10^{14} M_\odot$ . Further discussion of size of the detected



**Figure 10.** Halo Mass functions from AHF-, FOF- and  $\lambda_3$ -haloes. The AHF-haloes are fewer than FOF- or  $\lambda_3$ -haloes. The number of haloes above a mass threshold are binned and taken along vertical axis, normalized to simulation box volume. Error of  $1\sigma$  is shown in shaded region.

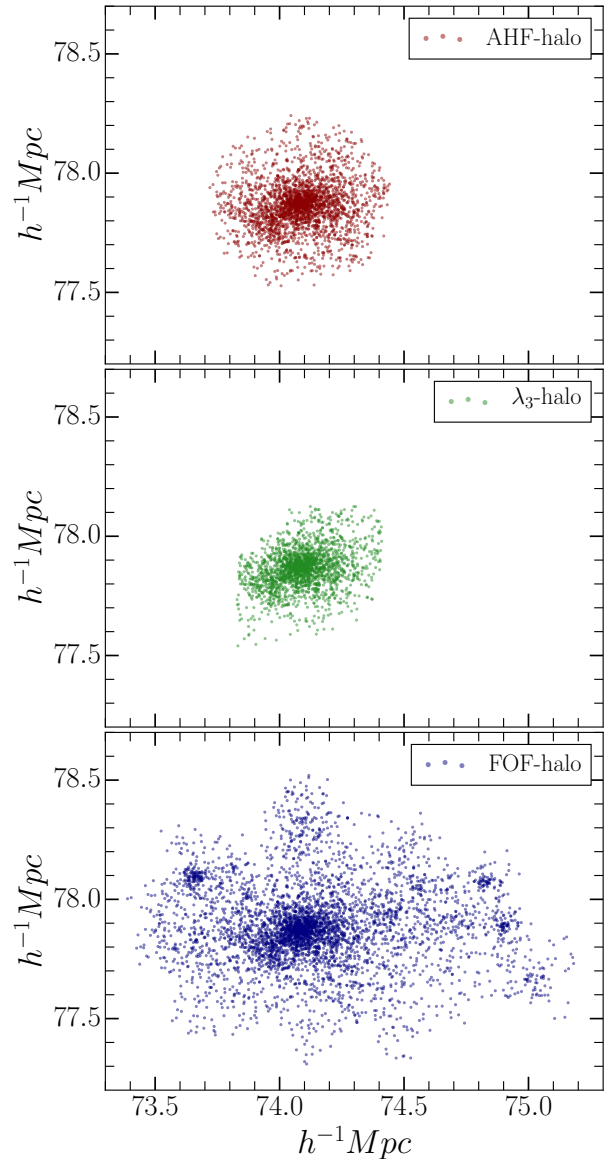
$\lambda_3$ -haloes in the context of smoothing of the multistream is done in Section 4.3.

The particles identified by the AHF as belonging to haloes form spherical structures due to a series of processes (including virialization) applied to unbind the particles. Inherently, the iso-density contours at virial levels are not spherical or spheroidal. The virialized AHF-haloes on the web are shown in the top panel of Figure 12. However, the spherical AHF-haloes are fewer in number compared to the other methods.

The popular choice of linking length of  $b = 0.2$ , although corresponding to virial density, does not ensure that the haloes have positive curvature. Most algorithms based on the FOF method re-define the halo boundaries by unbinding the particles outside a truncation radius. This truncation radius maybe the distance from the centre of mass of the halo to the farthest particle, rms distance, or an inflection point in the density field (For details on these methods, see ? and references therein). Some halo finders define the virial radius,  $r_{vir}$  at the distance from halo center where the density is  $\Delta_{vir}$  times the background density. In the middle panel of Figure 12, the FOF-haloes are shown without any of the above post-processing schemes. Without any unbinding, the FOF-haloes are generally larger in size than  $\lambda_3$ -haloes in the centre panel of Figure 12. For a specific case of a massive halo, Figure 11, FOF identifies more particles as bounded, than AHF or our algorithm.

In contrast to the AHF and FOF algorithms, our halo method directly detects a closed, convex surface (approximately the largest one, since  $\lambda_3 > 0$ ) for each of the haloes. There is no unbinding procedure on the particles identified within the halo site. The boundaries of  $\lambda_3$ -haloes themselves are not spherical or of any regular structure, but they are closed **convex surfaces**, as seen in **the middle panel** of Figure 12.

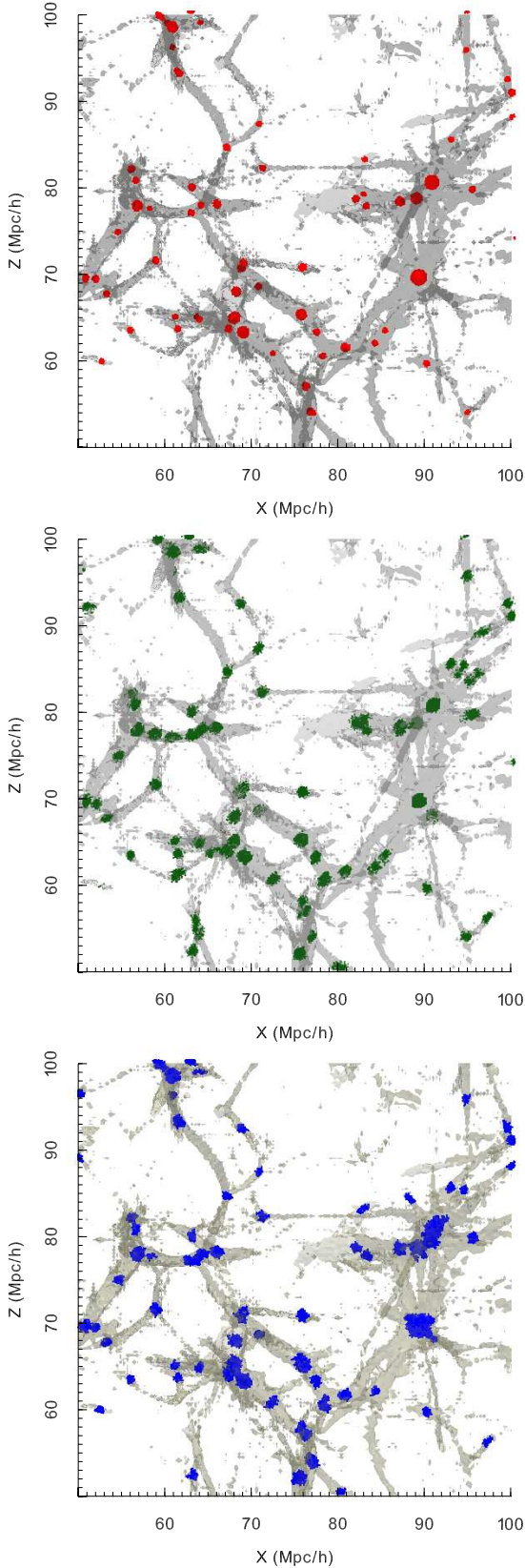
Haloes from the three finders in Figure 12 also show other differences in the halo boundaries. While all the AHF-haloes are spherical **by definition**, the FOF-haloes are irregular. The boundaries of the  $\lambda_3$  haloes are not spheri-



**Figure 11.** A large halo that was detected by AHF (top, red), our geometric analysis in the multistream field (centre, green) and FOF (bottom, blue). Halo boundary differs for each halo finder method. AHF detects particles within a sphere of virial radius. FOF-halo is irregularly shaped.  $\lambda_3$ -halo particles are in a non-spherical, yet compact structure.

cal either, but are more compact than FOF-haloes **and in addition they are convex surfaces by design**. At some junctions of the filaments, FOF identifies a large region as belonging to single halo, whereas AHF and our algorithm detect multiple isolated haloes. Each isolated  $\lambda_3$ -halo region enclose one maximum of multistream field, thus guaranteeing that multiple haloes are always resolved. On the other hand, a linking length cut-off or a constant threshold on scalar fields may enclose regions with multiple local maxima ( For one-dimensional fields, an illustration of this is shown in Appendix A).

**For a simulation box with  $N_p$  particles, each of mass  $m_p$ ,**

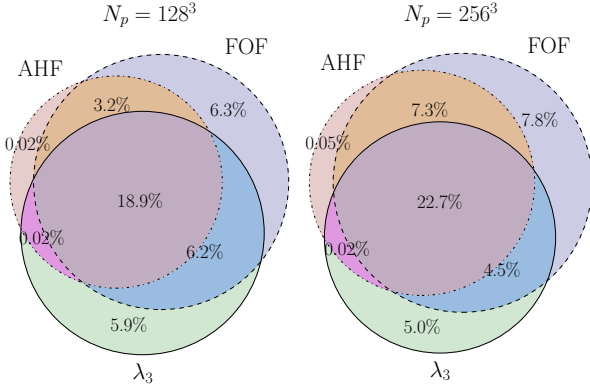


**Figure 12.** Potential haloes detected by AHF (top), our analysis (centre) and FOF (bottom). Most of the haloes are embedded in a percolating filament with  $n_{str} \geq 9$ . AHF-haloes are spherical by definition. FOF-haloes without any post processing are elongated along the filament.  $\lambda_3$ -halo candidates are neither spherical, nor elongated. Boundaries of  $\lambda_3$ -haloes are well resolved.

the halo-mass fractions,  $f_h = (\sum_{i=1}^{N_H} m_H(i)) / (m_p N_p)$  (where  $m_H$  is the mass of each halo and  $N_H$  is total number of haloes) are shown in a Venn diagram in Figure 13. For the simulation with  $N_p = 128^3$  particles, AHF-,  $\lambda_3$ -, and FOF- haloes occupy 22, 31 and 35 per cent of the total mass respectively. Nearly 19 per cent of the total mass are concurrently detected as belonging to haloes by all the three algorithms. FOF (with highest halo mass fraction) detects virtually all the haloes that AHF (with least halo mass fraction). About 3 per cent of the particles classified as belonging to haloes by both AHF and FOF are not classified as multistream halo particles. Our method also detected nearly 6 per cent of mass particles as haloes, which neither FOF nor AHF classify as haloes. For simulation with  $N_p = 256^3$  particles, the corresponding halo mass fractions  $f_h^{AHF} = 30$  per cent,  $f_h^{\lambda_3} = 32$  per cent, and  $f_h^{FOF} = 42$  per cent respectively. Thus the mass fraction  $f_h^{\lambda_3}$  remains fairly consistent over increasing mass resolution, as opposed to AHF and FOF. However, large fractions of these mass particles, nearly 23 per cent of the  $N_p = 256^3$  (increased from 19 per cent for low mass resolution simulation), are simultaneously detected as belonging to haloes by different methods, as shown in the right panel of Figure 13. For the simulation with  $N_p = 256^3$ , we also see increase in agreement between any two pairs for halo finders. That is, the mass fraction of haloes simultaneously detected by  $\lambda_3$  and FOF jumps from 25 per cent to 27 per cent. This correspondence increases from 19 to 23 per cent for  $\lambda_3$ -AHF, and 22 to 30 per cent for FOF-AHF pairs. For the same pair (in the  $N_p = 256^3$  simulation), 12.3 per cent of particles are detected by FOF but not AHF, whereas almost all the particles ( $> 99.9$ ) for the AHF particles were also detected by FOF. For the pair  $\lambda_3$ -AHF, 9.5 per cent of particles are detected by  $\lambda_3$  but not AHF, and 7.3 per cent of particles were detected by AHF but not by  $\lambda_3$ . Finally, for the  $\lambda_3$ -FOF pair, 5 per cent of particles are detected by  $\lambda_3$  but not FOF, and 15 per cent of particles were detected by FOF but not by  $\lambda_3$ .

On the other hand, looking at the mass particles that were only detected as haloes by one method, but not by other two, we see that only the multistream haloes improve (i.e., the disagreement reduces from 5.9 to 5 per cent) with mass resolution. FOF detects 6.3 and 7.8 per cent of haloes in simulations of  $128^3$  and  $256^3$  particles respectively, that were not classified as haloes by the other two methods. AHF-halo particles, being sub-set of FOF-haloes for the most part, show less than 0.1 per cent disagreement with other finders.

The discrepancies may have to be addressed on a case-by-case basis. One of the major difference between the haloes detected by isolating local multistream maxima regions and AHF/FOF is shown for a large halo in Figure 11. Without any unbinding procedure, FOF may detect very large irregular sized haloes, often consisting of multiple sub-haloes as shown in the bottom panel. On the other hand, the corresponding AHF-halo (top panel) is smaller spherical subset of FOF-halo. Furthermore, the  $\lambda_3$ -halo in the middle panel is smaller than both. Our multistream field detection technique selects strictly one  $n_{str}$  maxima within them. The sub-haloes detected by FOF (or



**Figure 13.** Mass fraction of haloes  $f_h$  (in per cent) as detected by the three finders. Each circle represents fractions of mass of all halo particles (AHF, FOF or  $\lambda_3$ ) in the total mass of the simulation box. The particles concurrently detected as belonging to haloes by different frameworks are shown in the intersections.

AHF), may be detected as separate  $\lambda_3$ -haloes. Nevertheless, some of the mass particles between the two neighboring haloes (like ones along saddle regions of multistream fields) will not be included as belonging to the halo. This effect is seen in halo mass functions (Figure 10 for large haloes of mass more than  $10^{14} M_\odot$  – number density of large  $\lambda_3$ -haloes is smaller than FOF. Similarly it causes a few discrepancies in mass fractions of potential haloes as well.

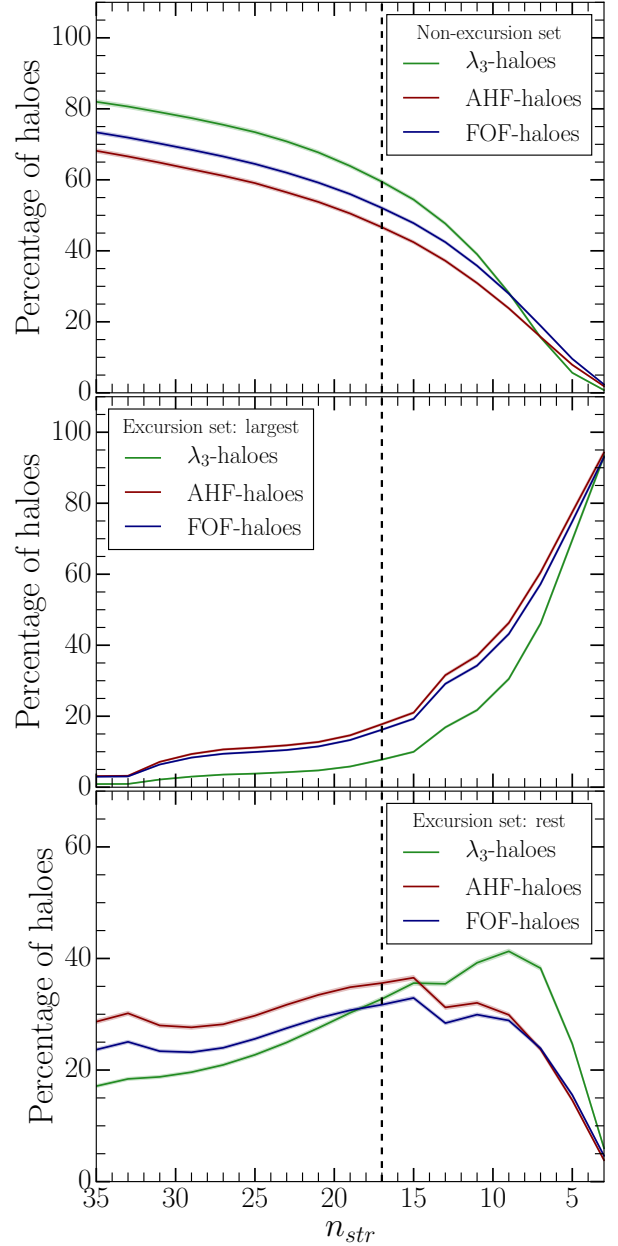
Other cause for differences in mass fraction is also rooted in the definition of haloes. Single-streaming regions are excluded from our halo search completely. Whereas, FOF and AHF employ no such mechanism to check for number of gravitational collapses. ? showed that a significant fraction (nearly 35 per cent) of FOF-haloes are in contact with single-streaming voids. Particles within these regions would not be considered as potential  $\lambda_3$ -halo particles. This also contributes to the discrepancy in halo mass fraction by different halo finders.

### 6.1 Haloes in the percolating filament

SECTION 6.1 that has been removed from paper-1, added here

The excursion set of multistreams above an  $n_{str}$  threshold hosts a varying number of haloes. We compare the multistream halo candidates from our geometric method with the AHF and the FOF method in Figure 14 for the simulation with  $N_p = 256^3$ . The regions in the co-ordinate space are classified into excursion set and non-excursion set regions based on whether the multistream is over or under the  $n_{str}$  threshold. In the excursion set we also distinguish the largest structure from the rest of the structures because the largest region of the excursion set plays the crucial role in detecting the transition to percolation. Percolation takes place at thresholds  $n_{str} \leq 17$  (?) to the right from the vertical dashed dashed line. Based on the coordinates of the halo particles, we check if a halo is in contact with the largest region of the excursion set or with rest of the excursion set.

The fraction of haloes in the non-excursion set are shown at various  $n_{str}$  thresholds in the top panel of Figure 14. At thresholds greater than 17 streams (i.e. in



**Figure 14.** Percentage of haloes detected (in the simulation with  $N_p = 256^3$ ) that are embedded in the non-excursion set (top panel), largest excursion set segment (center panel) and the rest of excursion set regions (bottom panel). Multistream haloes, AHF-haloes and FOF-haloes are shown in green, red and blue respectively. Vertical dashed line at  $n_{str} = 17$  is where percolation transition occurs.

non-percolating regime), a large fractions the AHF-, FOF- and  $\lambda_3$ -haloes are in the non-excursion set, as shown in the top panel of Figure 14. The fraction of  $\lambda_3$ -haloes is slightly higher than FOF or AHF in this regime. At relatively high threshold of, say,  $n_{str} = 35$ , about 65% of the AHF-haloes, about 75% of the FOF-haloes and about 80% of the  $\lambda_3$ -haloes are in the non-excursion set.

For multistream values lower than the percolation threshold of  $n_{str} \leq 17$  (i.e. in percolating regime) the fractions of AHF-,  $\lambda_3$ - and FOF-haloes in the largest (i.e.



percolating) region quickly grow with decreasing values of the threshold and surpass both the corresponding fractions in the non-percolating regions of the excursion set and that in the non-excursion set at  $n_{str} \approx 10$ .

The majority of the haloes belong to the single percolating structure (shown for the simulation with  $N_p = 128^3$  in Figure 12. Similar spatial distribution of SUBFIND haloes ? in the multistream regions is shown by ?) and at  $n_{str} = 3$ , all the haloes are attached to the web.

At  $n_{str} = 3$ , the filling fraction  $f_1/f_{ES}$  is almost close to unity ?. Most halo candidates from all three algorithms are at least in contact with the percolating cosmic web. Due to the threshold on  $n_{str}$  in our halo detection method, the  $\lambda_3$ -haloes are not only in contact, but completely within this structure.

## 7 DISCUSSION

The multistream field can have only integral values, moreover these values must be odd numbers practically at every point, because the number of streams can be even only on a set of measure zero. It means that in numerical simulations even values may occur on extremely rare occasions. We have analysed functional variation of the scalar field  $-n_{str}(\mathbf{x})$  using Hessian eigenvalues. The Hessian analysis is generally done for inherently continuous fields since it requires evaluation of the second derivatives. Geometries of web structures unveiled by Hessian signatures of smoothed density fields (such as ?, ?, ?, ? and many more), tidal shear or velocity shear tensor (?, ?, ?, ?, ?, ? etc.) and observational data from surveys (?, ?, ?, ? etc.).

Although the multistream field has discrete values by definition, it may be smoothed for numerical analysis at some scale, typically the scale of grid length of the field. The resulting Hessian eigenvalues approximately characterize the geometry in a four-dimensional hyper-space of  $(-n_{str}, x, y, z)$ . Our only assumption about the shape of the boundary of a virialized halo is that it is a convex surface. Therefore the boundary of a halo can be defined as a region with  $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$  since it is a closed convex contour in the  $(-n_{str}, x, y, z)$  hyper-space, and thus its projection onto the three-dimensional  $(x, y, z)$  space is also closed and convex.

Differentiating a smoothed  $n_{str}(\mathbf{x})$ -field may still pose a problem in the regions where  $n_{str}(\mathbf{x})$  is close to a constant and therefore the eigenvalues represent noise about the zero level. Fortunately in the most of such regions the unsmoothed  $n_{str}(\mathbf{x}) = 1$  therefore they can be easily eliminated.

Our algorithm for detecting potential dark matter haloes is unique due to two important factors: the geometrical attribute and the choice of field. Local geometrical analysis on the multistream field conveniently delineates the non-void structures without any free parameters. The dark halo candidates have compact surfaces that enclose local maxima of the multistream field. We do not employ non-local thresholds that several halo finders use (see ?, ? and ? for comparisons of various halo finders). Global thresholds (like a constant  $n_{str}$  cut-off) might be unsuitable for detecting halo candidates since the halo multistream environments are generally not categorical. Secondly, the  $n_{str}$  field enables us

to mask out the regions belonging to mono-stream regions without a heuristic criteria. Our method guarantees that none of the  $\lambda_3$ -halo particles belong to void region.

We note that present halo finders employ a variety of physical and numerical processes to identify dark matter halo candidates. Furthermore, there is no consensus in the definition of haloes itself (see discussion in ?). This is also the cause for the few differences between FOF-, AHF- and multistream haloes: FOF and AHF haloes only use Eulerian co-ordinates  $\mathbf{x}(z)$  – either raw positions or in the form of mass density. On the other hand, we utilize a mapping on the Lagrangian sub-manifold  $\mathbf{x}(\mathbf{q}, z)$  to define the multistream field  $n_{str}(\mathbf{x})$ . The boundaries of haloes in FOF and AHF are defined by the free-parameter thresholds of linking length and density, and the halo-center is usually defined as the center of mass of the particles. Conceptually, the center of a  $\lambda_3$ -halo is the location of the local multistream maximum, and the boundary of the halo is the convex region surrounding it. At least for large haloes like the one in Figure 11, this convex region is well within the FOF boundary.

The applicability of non-local thresholds in detecting haloes deserves deeper investigation. Lower bounds on over-density or linking-length thresholds traditionally define halo regions in several halo finders. Values such as  $\Delta_{vir} \approx 180$  or  $b \approx 0.2$  correspond to virial theorem applied to isolated spherical collapse models. Recently ? demonstrated that depending on halo environment, cosmology and redshift the over-densities corresponding to  $b = 0.2$  have different values. The virial theorem itself is a good measure of equilibrium of a system. However, the global thresholds empirically derived from it may not be pertinent to diverse environment of dark matter haloes.

The algorithm prescribed in Section 4 lists out a set of physically motivated steps that filter out the noisy  $\lambda_3 > 0$  regions that cannot be identified as haloes. The analysis in the simulation of  $100h^{-1}$  Mpc side length and  $128^3$  particles, with the multistream calculated on  $256^3$  diagnosis box, approximately 40000 labelled segments satisfying  $\lambda_3 > 0$  criterion in the non-voids were filtered out by a lower limit on multistreaming regions. One of the possible improvements in our algorithm would be to use information of number of flip-flops of each particle (For instance, using methods prescribed by ??). Such methods involving the Lagrangian sub-manifold may reveal rich sub-structure in the haloes.

The requirement that each halo should have closed convex surfaces with a multistream maximum inside may identify sub-haloes in large haloes but is too demanding because a halo with sub-haloes must have saddle points in the  $n_{str}$  field. This may explain the shortage of massive haloes shown in Figure 10. Applying a filter for smoothing the  $n_{str}$  field increases the number of massive haloes but reduces the number of low mass haloes. Although our present method does not currently perform an analysis simultaneously on multiple smoothing scales, such approaches done in density, lognormal density, tidal, velocity divergence or velocity shear fields (see MMF by ? and NEXUS+ by ?) have shown interesting multi-scale features of the cosmic web. Applying a more sophisticated procedure for linking  $\lambda_3$ -sub-haloes in a more massive halo will be done in the follow-up paper.

Dark matter haloes, being localised structures, are

uniquely convenient for our local Hessian analysis. Conditions of  $\lambda_1 > 0 > \lambda_2 \geq \lambda_3$  and  $\lambda_1 \geq \lambda_2 > 0 > \lambda_3$  also give information about curvature. Hessian eigenvalue analysis at high resolution of multistream fields may be very interesting in understanding the tubular edges of filaments and surfaces of walls at smaller scales. However, in this study, Hessian analysis is only applied to haloes. Walls and filaments span large volumes in the dark matter simulations, and we employ topological tools to investigate them.

## 8 SUMMARY

We studied certain geometrical of the multistream field in the context of halo formation. Findings from our analysis are summarized as follows:

(i) Several aspects of halo formation in the Lagrangian sub-manifold are considerably different than that of toy models of spherical top-hat collapse and ellipsoidal collapse. Successive formations of caustics (and consequently multiple velocity streams) play a crucial role in the process of clustering.

(ii) We present a novel halo detection algorithm for identifying dark matter halo candidates in the multistream field. Conditions on the local geometric indicators of the field are used to ensure that each closed halo boundary hosts a local multistream maxima. The positive signs of all principal curvatures (please note that we use curvatures of  $-n_{str}(\mathbf{x})$  field) inside the boundary also guarantee that the boundary is convex. Bounds on  $n_{str}$  guarantee that all the halo particles are in the non-void structure. We also ensure that the halo regions have foldings in the Lagrangian sub-manifold in more than one direction.

(iii) The multistream field within the halo boundaries may be very diverse. We do not detect halo candidates from a global lower bound on  $n_{str}$ . Instead, we look for closed convex regions in the multistream field. For the simulation with  $128^3$  particles, minima of  $n_{str}$  in each halo vary from 3 to nearly 450. Maxima of  $n_{str}$  in the halo vary from 7 to about 2800.

(iv) Our multistream halo candidates had a reasonably good correspondence with haloes from AHF and FOF catalogues. One notable difference was found with massive haloes. Our algorithm predicted fewer particles than the FOF method. This is likely to be caused by the requirement that the multistream field in the regions of the  $\lambda_3$ -halo candidates is convex which may be a reasonable approximation for simple haloes (i.e. having no sub-haloes) but massive haloes are more likely to have sub-haloes and therefore the  $n_{str}$  field in the corresponding regions must have saddle points and therefore cannot be entirely convex. Our study of the smoothing effects has shown that the number of massive sub-haloes tend to increase with growing smoothing scale which seems to agree with the above explanation. We will address this problem in the following study.

(v) Halo candidates were mostly embedded on the excursion set of the multistream field after percolation transition ( $n_{str} = 17$  in the simulation with  $256^3$  particles). At lower thresholds (around  $n_{str} = 5$  to 11), the largest percolating structure in the excursion set hosts most of the haloes.

In conclusion, the Lagrangian sub-manifold contains dy-

namical information of structure formation. We analysed the multistream field that contains the information of foldings in the sub-manifold. In addition, we demonstrated the use of geometrical features of the multistream field in identifying potential dark matter halo candidates in cosmological N-body simulations.

## ACKNOWLEDGEMENTS

This work has been funded in part by the University of Kansas FY 2017 Competition General Research Fund, GRF Award 2301155. This research used resources of the Advanced Computing Facility, which is part of the Center for Research Computing at the University of Kansas. We thank Mark Neyrinck and Mikhail Medvedev for discussions and suggestions. We also thank the anonymous referee for insightful comments on improving this manuscript and the editor for help with the submission.

## REFERENCES

- Abel T., Hahn O., Kaehler R., 2012, *MNRAS*, 427, 61
- Albareti F. D., et al., 2016, The Thirteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-IV Survey MAPPING Nearby Galaxies at Apache Point Observatory ([arXiv:1608.02013](https://arxiv.org/abs/1608.02013)), <http://arxiv.org/abs/1608.02013>
- Angulo R. E., Hahn O., Abel T., 2013, *MNRAS*, 434, 3337
- Angulo R. E., Hahn O., Ludlow A., Bonoli S., 2016, Earth-mass haloes and the emergence of NFW density profiles ([arXiv:1604.03131](https://arxiv.org/abs/1604.03131)), <http://arxiv.org/abs/1604.03131>
- Aragon-Calvo M. A., Jones B. J. T., Van De Weygaert R., M J., Der Hulst V., 2007, *A&A*, 474, 28
- Aragon-Calvo M. A., Van De Weygaert R., Jones B. J. T., Aragon-Calvo M. A., Van De Weygaert R., Jones B. J. T., 2010, *MNRAS*, 408, 2163
- Aragon-Calvo M., Neyrinck M., Silk J., 2016, preprint ([arXiv:1607.07881](https://arxiv.org/abs/1607.07881))
- Arnold V., Shandarin S., Zeldovich I., 1982, *Geophys. Astrophys. Fluid Dyn.*, 20, 111
- Ascasibar Y., 2010, *Comput. Phys. Commun.*, 181, 1438
- Bagla J., Padmanabhan T., 1997, *Pramana*, 49, 161
- Behroozi P. S., Wechsler R. H., Wu H.-y., 2013, *ApJ*, 762, 20pp
- Bertschinger E., 1985, *Astrophys. J. Suppl.*, 58, 39
- Bertschinger E., Edmund 1998, *Annu. Rev. Astron. Astrophys.*, 36, 599
- Bertschinger E., Gelb J., 1991, *Comput. Phys.*, 164, 164
- Bond J. R., Kofman L., Pogosyan D., 1996, *Nature*, 380, 603
- Bond N. A., Strauss M. A., Cen R., 2010a, *MNRAS*, 406, 1609
- Bond N. A., Strauss M. A., Cen R., 2010b, *MNRAS*, 409, 156
- 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
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- Diemand J., Kuhlen M., Madau P., 2006, *ApJ*, 649, 1
- Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
- Eisenstein D. J., Hut P., 1998, *ApJ*, 498, 137
- Evrard A. E., et al., 2002, *ApJ*, 573, 7
- Falck B., Neyrinck M. C., 2015, *MNRAS*, 450, 3239
- Falck B. L., Neyrinck M. C., Szalay A. S., 2012, *ApJ*, 754, 126
- Fillmore, J. A., Goldreich P., 1984, *ApJ*, 281, 1
- Forero-Romero J. E., et al., 2009, *MNRAS*, 396, 1815

Frenk C. S., White S. D. M., Davis M., Efstathiou G., 1988, *ApJ*, 327, 507

Gill S. P. D., Knebe A., Gibson B. K., 2004, *MNRAS*, 351, 399

Giocoli C., Tormen G., Sheth R. K., van den Bosch F. C., 2010, *MNRAS*, 404, 502

Gottlober S., Klypin A., Kravtsov A. V., 1999, *APS Conf. Ser.*, 176

Gregory S. A., Thompson L. A., 1978, *ApJ*, 222, 784

Gurbatov, S.N., Saichev, A.I., Shandarin S. F., 1989, *MNRAS*, 236, 385

Habib S., et al., 2009, *J. Phys. Conf. Ser.*, 180, 2019

Hahn O., Abel T., 2011, *MNRAS*, 415, 2101

Hahn O., Angulo R. E., 2016, *MNRAS*, 455, 1115

Hahn O., Carollo C. M., Porciani C., Dekel A., 2007, *MNRAS*, 381, 41

Hahn O., Abel T., Kaehler R., 2013, *MNRAS*, 434, 1171

Han J., Jing Y. P., Wang H., Wang W., 2012, *MNRAS*, 427, 2437

Heitmann K., et al., 2015, *Astrophys. J. Suppl. Ser.*, 219, 34

Hidding J., Shandarin S. F., van de Weygaert R., 2014, *MNRAS*, 437, 3442

Hoffman Y., Metuki O., Yepes G., Gottl S., Forero J. E., Libeskind N. I., Knebe A., 2012, *MNRAS*, 425, 2049

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, *MNRAS*, 321, 372

Klypin A., Shandarin S., 1983, *MNRAS*, 204, 891

Klypin A., Gottlober S., Kravtsov A. V., Khokhlov A. M., 1999, *ApJ*, 516, 530

Knebe A., et al., 2011, *MNRAS*, 415, 2293

Knebe A., et al., 2013, *MNRAS*, 435, 1618

Knollmann S. R., Knebe A., 2009, *Astrophys. J. Suppl. Ser.*, 182, 608

Kofman L., Pogosyan D., Shandarin S. F., Melott A. L., 1992, *ApJ*, 393, 437

LSST Science Collaboration et al., 2009, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))

Lacey C., Cole S., 1994, *MNRAS*, 271, 676

Lee J., Yepes G., 2016, *ApJ*, 832, 185

Libeskind N. I., Hoffman Y., Forero-Romero J., Gottlöber S., Knebe A., Steinmetz M., Klypin A., 2013, *MNRAS*, 428, 2489

Lin C., Mestel L., Shu F., 1965, *ApJ*, 142, 1431

Maciejewski M., Colombi S., Springel V., Alard C., Bouchet F. R., 2009, *MNRAS*, 396, 1329

Melott A. L., Shandarin S. F., 1989, *ApJ*, 342, 26

More S., Kravtsov A. V., Dalal N., Gottlöber S., 2011, *Astrophys. J. Suppl. Ser.*, 195, 4

More S., Diemer B., Kravtsov A. V., 2015, *ApJ*, 810, 36

Neyrinck M. C., 2016, *MNRAS*, 460, 816

Neyrinck M. C., Gnedin N. Y., Hamilton A. J. S., 2005, *MNRAS*, 353, 1222

Onions J., et al., 2012, *MNRAS*, 423, 1200

Pahwa I., et al., 2016, *MNRAS*, 457, 695

Planelles S., Quilis V., 2010, *A&A*, A94, 16

Press W. H., Schechter P., 1974, *ApJ*, 187, 425

Ramachandra N. S., Shandarin S. F., 2015, *MNRAS*, 452, 1643

Ramachandra N. S., Shandarin S. F., 2017, *MNRAS*, 467, 1762

Rasera Y., Alimi J. M., Courtin J., Roy F., Corasaniti P. S., Fuzfa A., Boucher V., 2010, *AIP Conf. Proc.*, 1241, 1134

Shandarin S. F., 1983a, Evolution of Potential Perturbations after Decoupling (The Adiabatic Scenario). Springer Netherlands, Dordrecht, pp 171–178, [doi:10.1007/978-94-009-7939-0\\_9](https://doi.org/10.1007/978-94-009-7939-0_9)

Shandarin S. F., 1983b, *Pis'ma Astron. Zh.*, 53, 195

Shandarin S., Medvedev M., 2014, preprint ([arXiv:1409.7634](https://arxiv.org/abs/1409.7634))

Shandarin S. F., Medvedev M. V., 2016, The features of the Cosmic Web unveiled by the flip-flop field ([arXiv:1609.08554](https://arxiv.org/abs/1609.08554))

Shandarin S. F., Zeldovich Y. B., 1989, *Rev. Mod. Phys.*, 61, 185

Shandarin S., Habib S., Heitmann K., 2010, *Phys. Rev. D*, 81, 1

Shandarin S., Habib S., Heitmann K., 2012, *Phys. Rev. D*, 85, 1

Skory S., Turk M. J., Norman M. L., Coil A. L., 2010, *Astrophys. J. Suppl. Ser.*, 191, 43

Sousbie T., Pichon C., Colombi S., Novikov D., Pogosyan D., 2008, *MNRAS*, 383, 1655

Springel V., 2005, *MNRAS*, 364, 1105

Springel V., Yoshida N., White S. D., 2001a, *New Astron.*, 6, 79

Springel V., White S., Tormen G., Kauffmann G., 2001b, *MNRAS*, 328, 726

Sutter P. M., Ricker P. M., 2010, *ApJ*, 723, 1308

Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, *A&A*, 506, 647

Vogelsberger M., White S. D. M., 2011, *MNRAS*, 413, 1419

Vogelsberger M., et al., 2014, *MNRAS*, 444, 1518

Weinberg D. H., Hernquist L., Katz N., 1997, *ApJ*, 477, 8

Zeldovich Y. B., 1970, *A&A*, 5, 84

Zeldovich Y. B., Einasto J., Shandarin S. F., 1982, *Nature*, 300, 407

de Lapparent V., Geller M. J., Huchra J. P., 1986, *ApJ*, 302, L1

van Kampen E., 1995, *MNRAS*, 273, 295

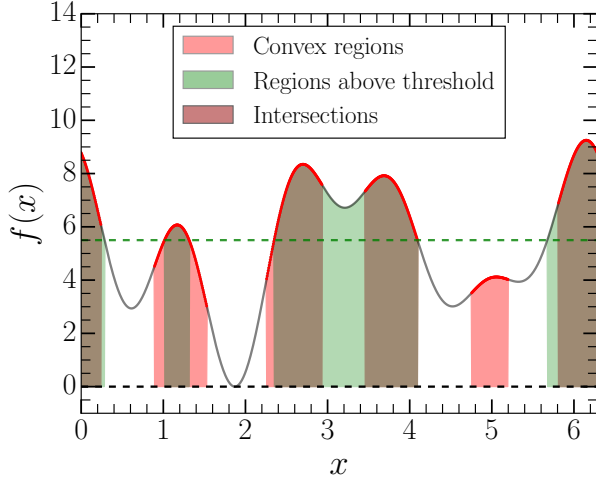
van de Weygaert R., Bond J., 2008, in Plionis M., López-Cruz O., Hughes D., eds, *Lecture Notes in Physics*, Berlin Springer Verlag Vol. 740, A Pan-Chromatic View Clust. Galaxies Large-Scale Struct.. p. 24, [doi:10.1007/978-1-4020-6941-3\\_11](https://doi.org/10.1007/978-1-4020-6941-3_11)

van de Weygaert R., Schaap W., 2009, *Lect. Notes Phys.*, 665, 291

## APPENDIX A: HESSIAN SIGNATURES OF THE MULTISTREAM FIELD

Second-order local variations of a scalar field  $f$  is described by a Hessian matrix, whose element in a three-dimensional domain is given by Equation 1. The geometry of the scalar field is classified by the Eigenvalues of the Hessian. The convex regions have at-most one maxima within the (3+1)-dimensional functional space. Projection of this closed region onto three-dimensional coordinate space also gives a closed surface in coordinate space. An illustration of the projection is shown in Figure A1 for a simpler function  $f(x)$  in one-dimensional domain. The eigenvalue criteria for regions are simplified: for instance,  $\frac{\partial^2 f}{\partial x^2} < 0$  for convex region. Projection of these regions onto coordinate space is shown in the shaded regions. This is different from regions within a contour, which is the projection of the curve along which the function has a constant value. Boundaries of these two regions may, but not necessarily, intersect.

In the case of cosmic fields, thresholds like  $\Delta_{vir}$  are equivalent to the green dotted line in Figure A1. The over-dense regions (green shaded regions) are not constrained to be convex. Similarly structures selected based on  $n_{str}$  thresholds do not universally result in convex structures either. Local geometry can be probed from the eigenvalue criteria instead, as shown by the red line on the curve and corresponding shaded area. The projected structures, albeit convex, may have very small values of  $f(x)$  (like the red shaded area around  $x = 5$ ). In the framework of identifying potential haloes in multistream field, multistream thresholds are devised in so that some of these small peaks detected by the Hessian are not considered as potential halo sites.



**Figure A1.** Projections of regions of  $f(x)$  from (1+1)-dimensional function space onto one-dimensional coordinate space. Convex regions and regions above a threshold of an arbitrary function  $f(x)$  are shown. Both the regions intersect around a few maxima, but not universally.