The Evolution of Multistreaming Events in the Formation of Large Scale Structures

Uliana Popov, Katrin Heitmann, James Ahrens, Salman Habib, and Alex Pang

Abstract—
This paper describes the analysis and application of visualization techniques to identify, track, and characterize multistreaming events and flows in the evolution of the Universe. Multistreaming is associated with the formation of visually striking large scale structure (LSS) comprised of elements such as halos, filaments, and sheets which have been theoretically predicted and observed in cosmological surveys. Many aspects in LSS theory still remain to be understood; it is therefore of great interest to study the role of multistreaming in the formation and evolution of cosmic structure. This problem is now being attacked with the aid of high accuracy cosmological simulations. In this paper, we describe new methods of identifying multistreaming regions based on various velocity based feature extractors and perform particle and region tracking of multistreaming events. We find that incorporating particle velocity information in the analysis reveals new insights about the evolution of LSS.

Index Terms—Cosmology, multistreaming, feature detection, particle tracking, region tracking, velocity field.

1 Introduction

This paper describes the analysis and application of visualization techniques to an interesting and challenging problem in the theory of cosmological structure formation – the identification, tracking and characterization of multistreaming events and flows. Multistreaming plays an essential role in determining the large scale structure (LSS) of the Universe at the current epoch. The dynamical mechanism for the formation of structure is the action of the gravitational instability on very small initial density perturbations taken to be a Gaussian random field. In the cold dark matter model (CDM), which is very successful observationally, the initial condition for structure formation has essentially no velocity dispersion, thus the proliferation of multistream flows with time, and the generation of velocity dispersion, is due to the (nonlinear) action of gravity. The use of multistreaming as a probe of cosmic evolution, and hence of gravity itself, is the focus of this paper.

The paper is organized as follows: We first provide a short overview of the problem in its cosmological context. Next, we re-examine the different ways in which multistreaming may be described and identify velocity based criteria for detecting it. We also provide a brief review of previous (and related) work in cosmological visualization and detection of the onset of multistreaming. Next we describe the data sets obtained from cosmological simulations, how they are prepared for analysis and visualization, the velocity based feature extractors, the scientific basis for selecting length scale based thresholds, and two complementary methods for tracking the evolution of multistreaming events. Finally, we present some of the interesting results obtained during our investigation.

Over the last two decades cosmology has made extremely rapid progress. There now exists a cosmological "Standard Model" that is in very good agreement with a large number of observational datasets at better than the $5-10\%$ level of accuracy. A key feature of the model is the existence of a "dark" sector that is not directly observable by emission or absorption of light but may be inferred via effects such as gravitational lensing and by its dynamical effects, especially in the formation of cosmic structure. Observations indicate that 70% of the Universe consists of a mysterious dark energy, 25% of a yet unidentified dark matter component (CDM), and only 0.4% of the remaining 5% of ordinary (atomic) matter is visible. Understanding the physics of the dark sector is the foremost challenge in cosmology today.

The large scale structure of the Universe has a complex topology consisting of sheets, filaments, and clumps. In the Standard Model, the formation of this structure is due to the amplification of tiny density fluctuations in the very early Universe that grow under the influence of gravity as the Universe expands. As the density fluctuations grow in amplitude, eventually becoming nonlinear, structure forms via a hierarchical process, in which the dark matter collapses into localized clumps called *halos* which attract and collect baryonic matter (gas) that eventually lights up as galaxies. The galaxies, being visible, can be used to track the dynamics and distribution of the underlying dark matter [15].

The evolution and dynamics of the dark matter distribution can be investigated by following the formation of LSS as observed in the distribution of galaxies today, and in the past. Tracers of LSS such as galaxy clusters ("0-D"), filaments ("1-D"), and surface-like pancakes ("2-D") can be considered to correspond to nodes, edges, and faces respectively, in a tessellation of the topology of the universe [4, 13]. The complex geometry and topology of cosmic structure is illustrated in Figure 1.

How structure forms in the Universe, and how it grows, can be used to test different dark energy models, including alternative theories of gravity. Such studies will be essential to interpret results from major ongoing and up-coming large-scale structure surveys, such as the Dark Energy Survey (DES), the Large Synoptic Survey Telescope (LSST), the Baryon Oscillation Spectroscopic Survey (BOSS), and space probes such as Euclid. Therefore, a deep understanding of cosmic structure formation is not only interesting in of itself but is an essential component of modern cosmology. This understanding can be used to guide and analyze numerical and analytical studies both in the nonlinear (the realm of numerical simulations) and quasi-linear (where one can still apply perturbation theory techniques) regimes of structure formation.

Precision dark matter simulations are a key foundation of cosmological studies. These simulations track the evolution of the dark matter with very high resolution in time, force, and mass. At the scales of interest to structure formation, a Newtonian approximation in an expanding universe is sufficient to describe gravitational dynamics. The evolution is given by a collisionless Vlasov-Poisson equation [8], a six-dimensional partial differential equation. Thus, on memory grounds alone, a brute force approach is impractical. Having made their first appearance in plasma physics, N-body codes now form a

- *Uliana Popov is with UCSC, E-mail: uliana@soe.ucsc.edu.*
- *Katrin Heitmann is with LANL, E-mail: heitmann@lanl.gov.*
- *James Ahrens is with LANL, E-mail: ahrens@lanl.gov.*
- *Salman Habib is with LANL, E-mail: habib@lanl.gov.*
- *Alex Pang is with UCSC, E-mail: pang@cse.ucsc.edu.*

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

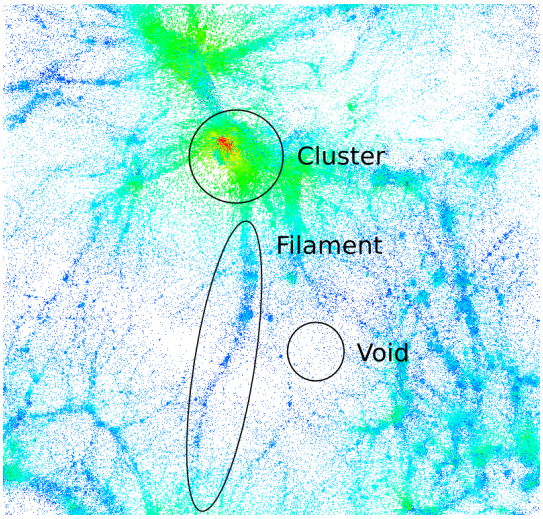


Fig. 1. Large scale cosmological structures of the universe.

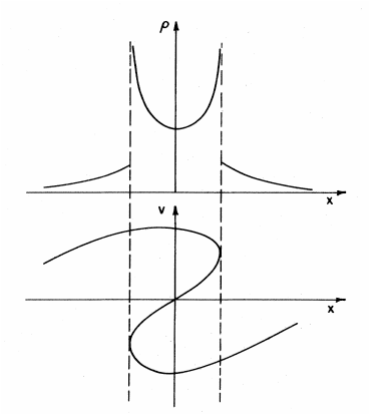


Fig. 2. 1-D illustration of multi-stream flow. Top panel: Over-dense region with three-stream flow confined between the dashed lines. Bottom panel: The corresponding phase space plot showing the different stream regions [17].

standard approach for dealing with this problem. In the N-body approach, the six-dimensional phase space distribution is sampled by "tracer" particles and these particles are evolved by computing the inter-particle gravitational forces.

The starting point of the simulations is a Gaussian random density field which imprints small perturbations on a uniform density, isotropic universe. The simulations start in the linear regime of the density fluctuations which then evolve under the influence of gravity. At any given length scale, during the early stages, the evolution remains linear but as time progresses, evolution first enters the quasi-linear regime (where perturbation theory can be applied) before finally reaching the fully nonlinear regime at which point all analytic descriptions break down. There is substantial interest in determining and characterizing the transitions between linear, quasi-linear, and nonlinear dynamics in the simulations by tracking the dynamics of dark matter tracer particles. At the start of the simulation, the velocity dispersion is initially zero, and the phase-space distribution is a three-dimensional sub-manifold of the phase space (only one velocity direction at a given spatial point). As the 3-hypersurface evolves, it folds, leading to the occurrence of singularities in the density field corresponding to the appearance of regions with multistream flow.

Finding multistreaming regions in cosmological simulations is an important endeavor for several reasons. The onset of multistreaming and the evolution of multistreaming regions as part of the theory of nonlinear structure formation is certainly interesting in of itself. Additionally, it is becoming an increasingly important aspect in understanding the formation of galaxy clusters where several "cold flows" combine. As mentioned above, different cosmological models and theories of structure formation will make different predictions for multistreaming.

The determination of the onset of multistreaming with respect to time and length scale is important in predicting the validity of approximate methods such as perturbation theory. Since running large cosmological simulations is very costly, cosmologists are always searching for methods that provide accurate answers at certain scales that do not require expensive simulations. For example, consider a key cosmological statistic measured from simulations: the density fluctuation power spectrum, P(k). The power spectrum can be predicted over a range of (large) length scales by perturbation theory. Multistreaming, however, cannot be described within perturbation theory. Thus, it is important to study the relationship of the breakdown of perturbation theory and the onset of multistreaming, especially with regard to being able to predict the associated time and length scales.

Finally, a robust method to capture the onset of multistreaming across multiple scales will help to set the initial cosmological time for starting cosmological simulations. The initial conditions for cosmological simulations are based on the Zel'dovich approximation, which is only valid if the paths of tracer particles do not cross (i.e., before multistreaming). Therefore, the simulations have to be started sufficiently before the occurrence of multistreaming events in order to guarantee accurate results.

Traditionally, LSS is investigated primarily by considering the distribution of dark halos. Although there are differences between methods, halos are typically identified by thresholding on the density of tracer particles (for a description of halo finders, see, e.g., [12]). In this paper, we are concerned not so much with density, but how the velocity information of tracer particles can find and characterize multistreaming regions. This work is carried out in close collaboration between cosmologists and computer scientists.

2 Multistreaming

Multistreaming is said to occur when there are multiple velocities at a given spatial point. A simple example is illustrated in Figure 2 for a one-dimensional cold and collisionless medium [17]. In the phase space plot (bottom panel), the boundary between a three-stream flow and a single stream is denoted by the dashed lines. At the boundaries, there is a shell-crossing singularity (caustic) in the density field because the mapping from phase space to physical space becomes multi-valued. This picture generalizes to higher dimensions.

At early times, the universe is close to homogeneous and the velocity dispersion – in the CDM model – at any space point is essentially zero. At this early stage, leading order Lagrangian perturbation theory (the Zel'dovich approximation) is applicable; under this assumption, particle motion is inertial to a good approximation:

$$x(t,q) = q + t \cdot v(q). \tag{1}$$

The particle position in Eulerian space x is a function of time t and particle position in the (initial) Lagrangian space is q. $v(q)$ is the initial velocity field. The cosmological analog of this equation is

$$r(t,q) = a(t)[q - b(t)\nabla\Phi(q)], \tag{2}$$

the particle position in Eulerian space r is a function of cosmic time t and particle position in Lagrangian space q. $a(t)$ is the cosmological expansion scale factor, and $b(t)$ is the growth rate of linear density fluctuations. In linear theory, the gravitational potential field $\Phi(q)$ describes the initial velocity field via $-\nabla\Phi(q)$, which is conservative and irrotational. In Equation 2, the first term is the unperturbed particle position, and the second term is the spatial perturbation. Equation 1 reduces to Equation 2 through the following simple substitutions:

$$r(t,q) = x(t,q)a(t), \quad b(t) = t, \quad v(q) = -\nabla\Phi(q)$$

The Zel'dovich approximation provides a good description of the evolution of density perturbations until the onset of multistreaming where it breaks down. In the one-dimensional case above, the approximation correctly describes the shell-crossing phase space structure in the early stages of evolution, but fails as the number of streams continues to grow (from three to five to seven ...).

As particles move under gravity, they form structures such as clusters and super-clusters, filaments, and pancakes, or voids. Multistreaming happens in these structures: particles from different position $q_1, q_2, q_3, \ldots, q_n$ in Lagrangian space congregate at a given point x in Eulerian space having different velocities $v_1, v_2, v_3, \ldots, v_n$ [17]. In other words, a multistreaming region consists of heterogeneous particle flows. (Here we would like to distinguish between multiple "cold" flows which are the focus of attention here and the velocity dispersion due to late-time virialization of particle velocities in dense regions.)

Given a general structure formation scenario, we wish to investigate all of the three conditions associated with multistreaming in a given spatial region:

- Particle flows have different speed and direction.
- Particles flows have the same speed but different direction.
- Particles flows have different speeds but the same direction.

Aside from these conditions, one can search for other possible metrics for extracting multistreaming regions. For example, regions with high velocity variance [18], or regions with high shear can account for any of the three situations above. Furthermore, the association of multi-streaming with high-density regions can suggest new methods as well as provide new avenues for understanding the formation of structure [18].

To summarize, there are a number of potential identifiers for caustics or multistreaming events e.g. high density regions, exploiting singularities in phase-space, high velocity variance, and high shear regions. Other possibilities include looking for changes in the velocity field — e.g. see if the flow remains curl-free as well as linear. We describe those used in this paper in Section 4.3.

3 Previous Work

The visualization of cosmological data sets has received considerable attention recently. Since all cosmological simulations are particle-based, one of the popular tools for visualizing particles directly is Partiview [10]. Within the visualization community, there are also a handful of recent contributions dealing with astrophysics data sets. These include the work of Li et al. [11], focusing on how to display positional and trajectory uncertainties in astrophysical data sets. In the same year, Navratil et al. [14] described their visualization approach for a data set that studies the formation and effects of the radiation from the first stars and the impact on subsequent star formation. Haroz et al. [7] investigated particle-based simulation data sets of the evolution of the universe and studied the relationship of different variables especially in the face of uncertainty arising from the different code settings, e.g. the starting time for the simulation. A similar data set was also investigated by Ahrens et al. [1] where the focus was on comparison rather than on uncertainty visualization.

There have also been several papers on multistreaming events. For example, Yano et al. [20] studied the distribution of caustics (see Section 2) in the expanding universe, while Gouda [6] investigated the relationship between catastrophe theory and gravitational clustering leading to caustics. In these studies, the models describe continuous matter density fields, e.g. density perturbations, singularities of density, etc. Most recently, Shandarin [18] proposed a new numerical technique to identify the cosmic web based on locating multistream flows. This work uses the velocity information of particles and not just the location information. The results reported in this paper is similar in that we also examine the velocity information in addition to the positional information of particles. A good starting point for reading more about about multistreaming can be found in [5].

4 Analysis Methods

In this section, we describe the data from cosmological simulations, the preprocessing steps, the local feature extractors and the associated physics based thresholds, and both particle and region tracking methods.

4.1 Cosmological Simulation Data

In this paper, the analysis and results are based on a cosmological simulation consisting of 512^3 particles within a box that is 256 h^{-1}Mpc long on each side. Each megaparsec (Mpc) is approximately 3.26 million light years. Other parameters used in the simulation include a grid of $n_g = 1024^3$ that is used to resolve the forces, a length scale of L =256 h^{-1}Mpc, and a starting value for the scale factor of $a = 5 \cdot 10^{-3}$. h is a parameter in the range [0.5,0.75] reflecting the uncertainty in the value of the Hubble constant H for the rate of expansion of the universe: $h = \frac{H}{100 \text{ km/s/Mpc}}$. 500 snapshots are collected until $a = 1$. An N-body Mesh-based Cosmology Code on the Cell (MC^3) was used in generating cosmological simulations [16]. The code simulates the physics of tracer particles under gravitational influences. Each particle has a unique tag, position, and velocity at each time frame. More importantly, particles are collisionless. They can occupy the same space without physically colliding into each other (at the scales that clustering is being represented and simulated).

4.2 Preprocessing

All our analysis methods are based on continuous fields while the simulations themselves are particle based. In order to obtain the continuous fields from the particles, we need to calculate the field quantities at any location in space. For practical and physical considerations, we impose a grid over the spatial domain and considered two different methods of calculating field quantities at the grid vertices. Field quantities of interest are particle density and aggregate velocity. We will use particle density to explain the process, but other field variables can be substituted. Commonly used methods for assigning values to different grid locations are based on nearest-grid-point (NGP) and cloud-in-cell (CIC) methods. NGP simply assigns each particle to the nearest grid point while CIC [9] uses a weight factor to account for the distance of the particle to its closest grid points. That is, it looks at the cells containing the grid point. CIC sums the weight of each particle and normalizes by the number of particles. The weight that each particle contributes to a grid point is calculated as a ratio between two volumes. The numerator is the volume formed by multiplying the differences between the influence radius (set to be the same as the grid spacing by default) and the distance of the particle to the grid point along the three principal axes. The denominator is the size of volume over which a particle exerts any influence (set to be the same as the cell size by default). Figure 3 illustrates how the weights are calculated for the 2D case. The results shown in this paper are based on CIC generated fields, we confirmed that the NGP algorithm leads to similar results.

The choice of grid resolution is quite important. If the grid is too coarse, the resampling process will smooth out the data too much and we may miss the multistreaming event. In addition, the grid size has to be small enough to resolve the features of interest at certain length scales. On the other hand, if the grid is too fine, it would result in a low particle count and confidence, not to mention the extra computational expense. For our investigation, we choose a grid resolution such that on average there are 64 particles contributing to each grid point. For the 512^3 particle data set, this goal is achieved by a regular grid with 256^3 cells. As we explain later, this grid size also allows us to find multistreaming regions early on in the evolution. At the start of the simulation, each grid cell contains 8 particles on average. Therefore, the 8 cells sharing a grid point contain 64 particles on average. The simulation uses periodic boundaries. Note that as time progresses, some regions become more dense while others become more sparse or even empty. Empty cells as well as those in their immediate vicinity must be treated with care and are specially marked so that they do not produce erroneous results in the analysis.

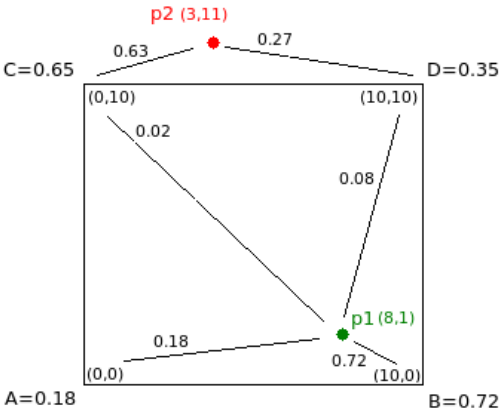


Fig. 3. Illustration of how weights are calculated using cloud-in-cell for the 2D case. The 10 × 10 cell ABDC contains a particle p1 within it. The weight of p1 on grid point C is the area bounded by p1 and B over the area of the cell (assuming the influence radius is set to the cell size). That is, $W(C,p1) = (10-|8-0|)(10-|1-10|)/(10*10) = 0.02$. The weight of a point p2 on grid point C is calculated in a similar fashion – $W(C,p2) = (10-|3-0|)(10-|11-10|)/(10*10) = 0.63$.

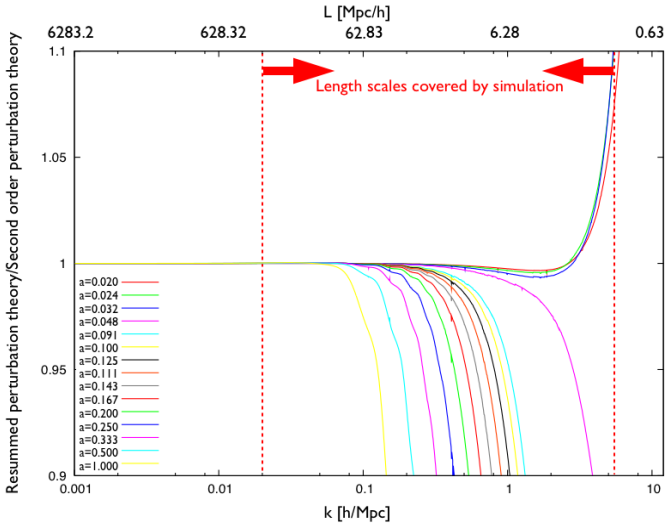


Fig. 4. Breakdown of perturbation theory at different scale factors and different length scale. The curves are the ratio of two different perturbation theories. As the ratios deviate from one, perturbation theory is not valid anymore. Each curve shows the result for one time snapshot. At the top of the plot we indicate length scales, at the bottom we indicate wave numbers. The dashed lines in red show the scales that can be resolved by the simulation data.

4.3 Local Feature Extractors

The continuous density and vector fields obtained from the preprocessing step above are used by several local feature extractors to identify potential multistreaming sites. These local feature extractors are based on dot products, velocity variance, vorticity, linearity tests, divergence, and maximum shear [3]. In this paper, we present our finding of multistreaming behavior based on a combination of particle density, dot product and vorticity.

An intuitive way of measuring agreement (or lack thereof) of particle velocities is to measure their dot products. The motivation of using this as a feature extractor is that it can capture differences in particle trajectories and speed. We first calculate an average velocity at each grid point using the CIC method. Then we obtain the dot product of each particle velocity against the average velocity. These are summed for all particles and the result normalized by the number of particles.

Vorticity measures the tendency of vector field elements to spin. In cosmological simulations the velocity field is irrotational (zero curl) prior to multistreaming. We hypothesize that vorticity may be used as an indicator for multistreaming. The vorticity at a point is a vector and is defined as the curl of the velocity. That is, vorticity is $\nabla \times \vec{V}$ where $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$. Since we are primarily interested in detecting the presence of regions with rotational motions and not their particular orientations, we look at the vorticity magnitudes in the simulations. In Section 5, we use the velocity based information in conduction with a density based method to find and track multistreaming regions.

4.4 Perturbation Theory

In order to estimate at what time and on what length scales multistreaming should be seen we investigate the validity of perturbation theory. The perturbative treatment of gravitational clustering should break down in regions where multistreaming events occur. To predict these events, we make the following simple argument based on an internal check within the perturbative analysis. To do this, we note that perturbation theory can be carried out at different orders in the density perturbation. In the regimes where perturbation theory works, higher-order corrections serve to improve the lower-order results. However, once the fluctuations are too large, consistency between orders no longer exists, and different order results can disagree strongly. By investigating at what scales two different approaches at different orders diverge from each other, we can estimate the scale where perturbation theory fails, and hence produce a candidate scale for the onset of multistreaming. Following Carlson et al. [2], we calculate the matter power spectrum for second order perturbation theory and a re-summed scheme with a code provided by the authors. We then take the ratio of these power spectra at different epochs. The results are shown in Figure 4 for scale factors between $a = 0.02$ and $a = 1.0$.

An estimate of when and at what length scales multistreaming will occur can be obtained by measuring the scales at which the curves deviate from unity in Figure 4. The figure indicates that these scales vary with time. Multistreaming regions that are relevant to the breakdown of perturbation theory start out as small structures which grow bigger over time. The dashed line on the right indicates the resolution limits due to smoothing from the density calculation. It can be easily varied by reducing or increasing the grid size for the CIC (an increase moves the cutoff lower and a reduction moves it higher), although one cannot increase it beyond a certain point set by particle spacing limits in the simulation. For the data set being presented in this paper, the smallest wave-number ($k = 2\pi/L$) is k $\approx 0.02\ h^{-1}$Mpc, and the corresponding smallest length scales we can resolve is $0.256\ h^{-1}$ Mpc. Using a grid of 256^3 cells, we can resolve length scales of $1\ h^{-1}$ Mpc. However, when coupled with CIC with window size equal to one cell, our resolution drops to length scales of $2\ h^{-1}$Mpc.

Table 1 is created based on the predictions from Figure 4. It lists the expected size of the multistreaming scales for different snapshots in the simulation data. The time stepping unit is measured with respect to the scale factor a. Given that there are 500 time steps in the simulation, $\Delta a = 0.002$ from one frame to the next.

This table is instrumental in determining the threshold values used by both the dot product and vorticity feature extractors. As can be seen in this table, multistreaming regions grow over time. We therefore use the information from Table 1 to guide us in determining an appropriate threshold value to use. For example, if we are searching for regions of interest at frame 250, we expect these regions to have length scale of about 37 h^{-1}Mpc. Therefore, we want to find a threshold value that will produce regions of this expected size. Since the regions may come

a	Frame #	L scale 10%	L scale 5%
0.500	248	37 h^{-1}Mpc	43 h^{-1}Mpc
0.333	165	30 h^{-1}Mpc	34 h^{-1}Mpc
0.250	123	24 h^{-1}Mpc	27 h^{-1}Mpc
0.200	95	18 h^{-1}Mpc	24 h^{-1}Mpc
0.167	80	14 h^{-1}Mpc	19 h^{-1}Mpc
0.111	70	10 h^{-1}Mpc	12 h^{-1}Mpc
0.125	60	8 h^{-1}Mpc	10 h^{-1}Mpc
0.111	52	6 h^{-1}Mpc	8 h^{-1}Mpc
0.100	47	5.8 h^{-1}Mpc	7 h^{-1}Mpc
0.091	18	5.6 h^{-1}Mpc	6 h^{-1}Mpc
0.045	21	2 h^{-1}Mpc	3 h^{-1}Mpc
0.033	13	1 h^{-1}Mpc	1.05 h^{-1}Mpc
0.020	7	0.9 h^{-1}Mpc	0.98 h^{-1}Mpc

Table 1. This table shows the relationship between the scale factor a and the frame number of the simulation. It also shows length scale for two different tolerances at which perturbation theory breaks down. The tolerances are at 10 and 5 percent from the ratio of one between the two perturbation calculations seen in Figure 4. When choosing the grid size for calculating the continuous fields it is important that the smallest length scale of interest is resolved. For example, in frame 30 at a tolerance of 10 percent, the scales of interest are at 1 h^{-1}Mpc. With a box size of 256 h^{-1}Mpc the grid size has to be at least 256^3 to resolve these scales. If the grid is coarser, the length scale that can be resolved increases and therefore multistreaming events could only be resolved at a later time step.

in a variety of shapes, and because the length scale itself does not fully capture shape information, we use it as an indicator of a region size rather than a strict length scale. In this regard, region size is taken to mean the number of connected grid points that are above the current threshold. To determine the appropriate threshold for a given frame, the initial threshold threshold$_0$, is set to a value that will result in all points being classified as multistreaming according to the feature extractor. For the dot product we set threshold$_0$ $=$ 0, while for vorticity, threshold$_0$ $=$ *maxint*. If the extracted regions are not at the predicted length scales, we adjust the current threshold to threshold$_i$ automatically and restart the scanning process. The adjustment to the current threshold value depends on the range of values for a particular feature extractor e.g. dot product's range is [-1..1] while vorticity's range is [0..maxint]. Once we find at least one region with the expected feature size, we finalize the threshold value for that frame. Because the growth of region size is fairly well understood, we can use the final threshold value of the current frame as the initial guess for the next frame. The resulting regions extracted for every frame are each assigned a region ID and size. The number of regions in a frame is also recorded.

To find an overdense region, we use the following equation: $\delta = \frac{\rho - \widetilde{\rho}}{\widetilde{\rho}}$, and setting $\delta = 1$. This will set the density threshold ρ to twice the mean particle density in the entire volume. Since we use a 256^3 grid to hold the 512^3 particles, the average density is 8. The density threshold is held constant and does not vary from one frame to another like the vorticity and dot product thresholds.

Since we have a number of possible criteria for classifying a point as either multistreaming or not, they can be combined using the boolean operators *and* and *or* e.g. density and dot product criteria are both met. Such regions can be easily obtained from the multistreaming region list based on individual feature extractors.

4.5 Region Tracking

Once a multistreaming region is found, perhaps using a combination of density and velocity based methods, it is tracked both forward and backward in time. Similar to volume tracking where a region of interest is represented by a collection of voxels [19], in region tracking, a region of interest is represented by a collection of grid points. However, in contrast to volume tracking, we do not use volume differencing to track a region from one frame to the next. Instead, our region tracking method uses a region growing strategy to determine whether regions from one frame to the next have merged, continue, or bifurcated. For each region that needs to be tracked, we examine the neighborhood of connected grid points in the adjacent time frame to see if the region has grown or shrunk. In addition, if another region is visible within the neighborhood of the region being tracked, we determine if the new region can be merged. And if so, the neighborhood grows to encompass the entirety of the other region as well. To determine how large a neighborhood we need to examine is also physically determined. The average particle speeds do vary over time, but they are bounded. That is, from one frame to the next, a particle (and hence the multistreaming region) can only travel so far. We exploit this fact by choosing an appropriately sized neighborhood search window to track and look for possibly connected regions.

Tracking may start at any frame in the simulation. This is useful, for example, to determine the onset of multistreaming; or to investigate where the different contributions of a large multistreaming region came from.

4.6 Particle Tracking

While region tracking is useful for showing the evolution of multistreaming regions, it only provides a "zoomed out" picture and does not provide enough information about what exactly happened to a particles inside a region. As such, we consider it as a higher level tracking technique. The following scenarios illustrate some of the inadequacies of region tracking. When two multistreaming regions get close together and form a single region, did all the particles become part of a new region or were some of them left out? As regions disappear (fall below the threshold) or appear suddenly (meet the threshold), what particles contributed to their change in status? Likewise, given a multistreaming region, what is the behavior of the particles within that region? How do those particles evolve over time? Do they stay with the same multistreaming region or mix and migrate to other multistreaming regions?

To address such questions, we track individual particles. Since each particle has a unique tag, particle tracking is straightforward since there is a one-to-one correspondence and we can find its location in any time frame. It provides information at a much finer level of detail, but it does not have the ability to look at the whole picture. For example, it would be very difficult to understand why a particle moved in a certain fashion without seeing its environment e.g. there could be a dense multistreaming region that is exerting a strong gravitational pull on this particle.

In analyzing our data set, we use both region tracking and particle tracking to confirm previous results and test new hypothesis. We discuss these in the next section.

5 RESULTS

Using different types of information derived from particle velocity, and combined with particle and region tracking strategies, we can study LSS formation in a way that is different from density thresholding. Although, some aspects will be similar, others are different and can become independent probes of structure formation [18]. As an example, we have been able to confirm the standard scenario of LSS formation within which denser and larger multistreaming regions cannibalize smaller regions in the local vicinity. However, this process is not as straightforward as it seems. An isolated region may grow slowly until another multistreaming region passes nearby and attracts some of its particles. The original region may therefore decrease in size. It may also increase in size if it is denser or larger relative to the other region. These effects can even cause a complete disappearance of a region when a sufficient number of its particles are absorbed by other regions. In addition, the interaction among the particles in such encounters exhibit very interesting dynamical behaviors e.g. spirals, flocking, etc., which would not be seen by an analysis relying solely on the density. These complex dynamical effects possibly reflect the many different types of multistreaming region. This presents an interesting area for future work: In addition to classifying multistreaming

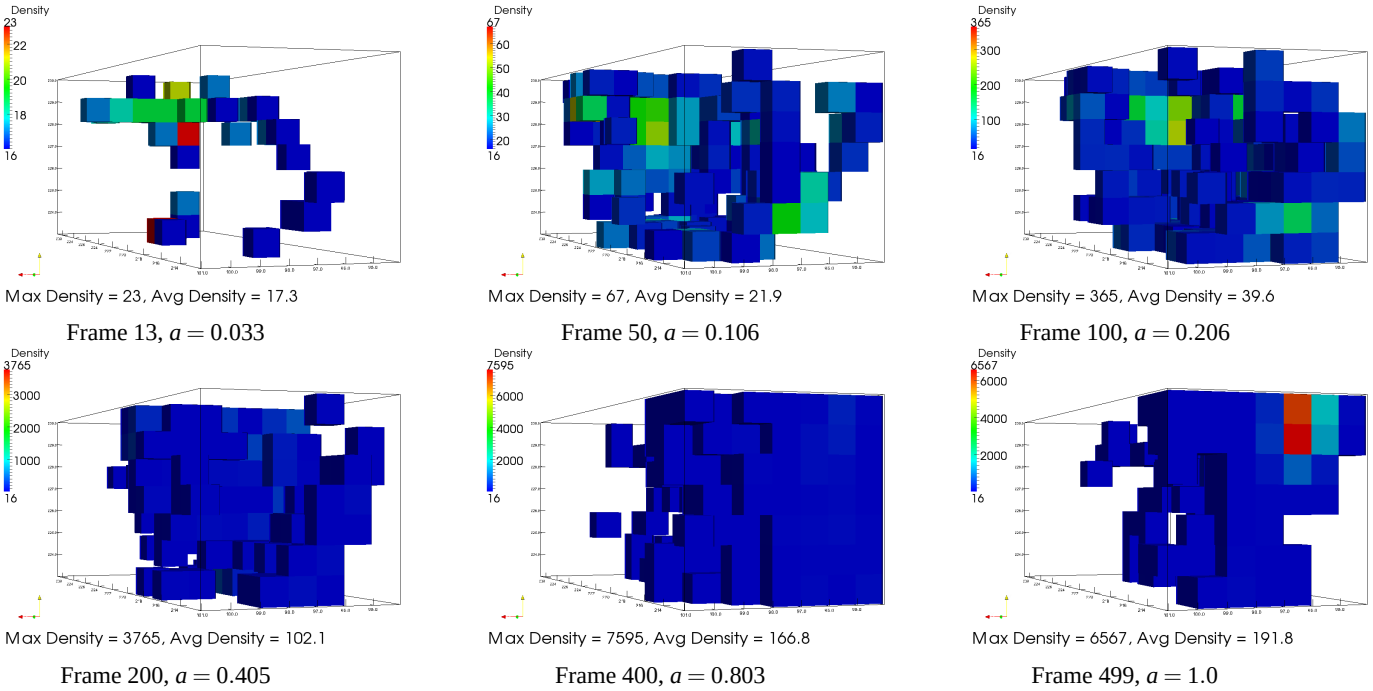


Fig. 5. Overdensity regions found with density threshold set to twice the mean density. We can observe that over time, the average density in this region is increasing. The regions extracted with density as the only criterion will be larger than those that use density and vorticity (Figures 6 to 8) and density and dot product (Figure 9).

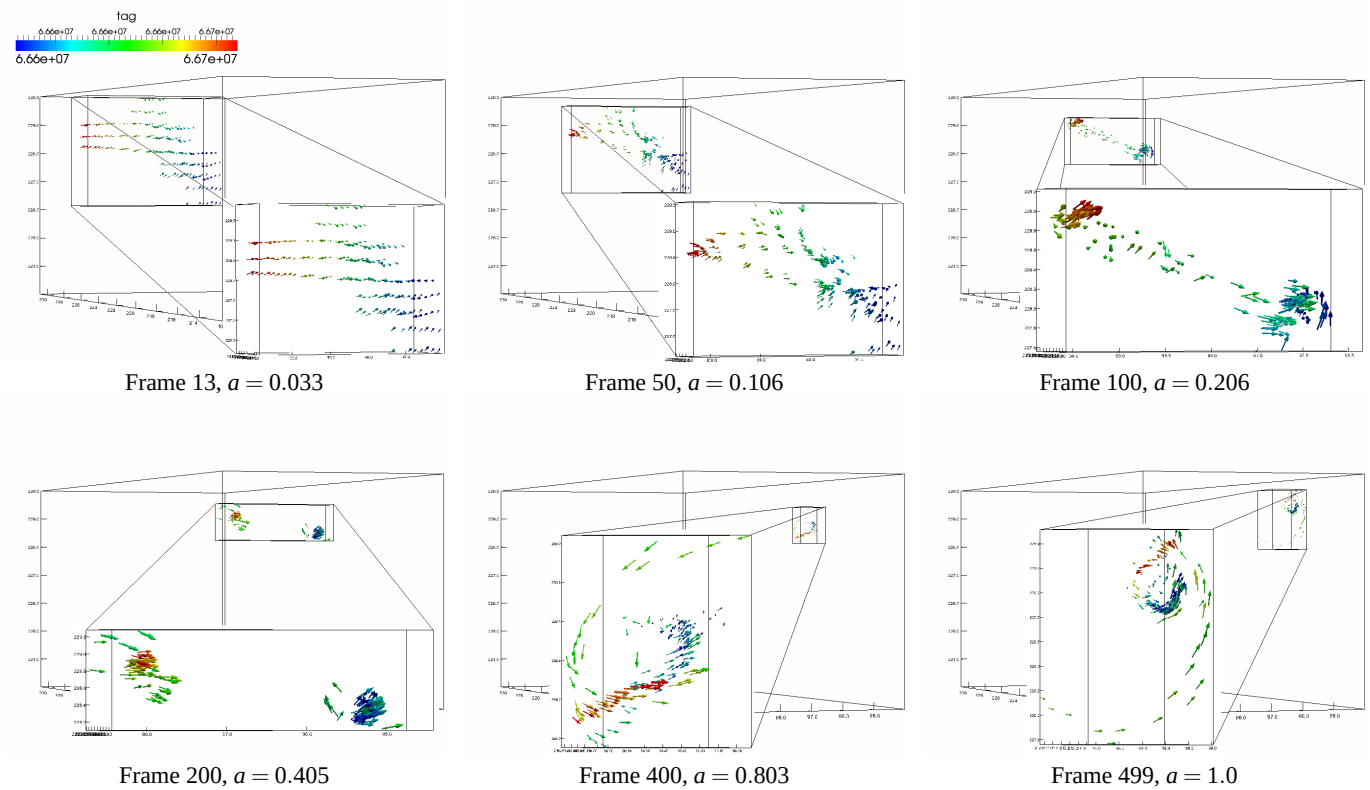


Fig. 6. Particle tracking: Using a combination of vorticity and density criteria, we pick a region in frame 13 and track the evolution of the particles contained in that region. In this case, the density threshold is at 16 particles per cell (twice the average density) and the vorticity threshold is 150. The particles are colored by their tag ID. The inner box is the bounding box of the particles in the region. At the starting frame for tracking (frame 13) the inner box has two purposes: (i) it bounds the particles to be tracked, (ii) it bounds the multistream region we pick using the density and vorticity threshold. After frame 13 it only serves as a bounding box for tracking purposes. The outer box displays the region over which the inner box travels during the simulation. We can see that these particles move towards the center of the outer box (frame 50), then split into two different groups (frame 200), and later recombine into a single group again (frame 400). Interestingly, they reform in a spiral pattern.

regions by their size and structural morphology (cluster-like, filament-like, or sheet-like), we may be able to further classify them by their behavioral characteristics. We illustrate these in the examples below using different combinations of feature extractors: density and vorticity (Figures 6-8) and density and dot product (Figure 9).

Figure 6 demonstrates particle tracking. Here, the particles forming the feature of interest are identified by a combination of vorticity and density criteria. The largest such feature from frame 13 is selected for tracking. This time frame corresponds to the expected onset of multistreaming based on the prediction from Figure 4. We follow all the particles that are inside this region at time frame 13 using particle tracking. Figure 6 tells us an interesting story. Particles first start to move towards the center of the region, then, around frame 200, split into two groups, and later, after frame 350 they combine into one region again. The particles move in a spiral fashion.

Figure 9 also shows an example for particle tracking. This time the features are extracted using a combination of density and dot product criteria. We pick a region at a later time frame in the simulation and track the particles forward in time. Similar to Figure 6, the particles split into two groups, and rejoin into a single group towards the end of the simulation. However, unlike the features extracted using vorticity, the dot product features do not exhibit strong spiraling behavior. Another difference in the behavior of these two types of multistreaming regions is that the vorticity based one traveled significantly farther (more than $20h^{-1}Mpc$) than the one based on dot product. However, it is not clear that such behaviors can be generalized since what happens in the vicinity of these particles can greatly influence their behavior.

Next we turn to the region finding algorithm and analyze the region shown in Figure 6 in more detail. The results are shown in Figure 7. The multistreaming regions extracted at each time frame are rendered as grey transparent boxes. There can be multiple multistreaming regions present. Figure 7 provides a glimpse of why the particles in Figure 6 split into two groups and later rejoined into a single group.

Finally, Figure 8 illustrates region tracking. Unlike Figure 7 where the multistreaming regions are extracted, the pink transparent multistreaming region is tracked from frame 13 onwards. The pink regions are always a subset of the grey regions, except for the portion that went outside of the larger box.

Existing methods like halo finders primarily rely on region finding using overdensity only. As can be seen in Figures 6 to 8, relying exclusively on these techniques may lead to missing some very interesting behaviors.

6 Conclusions and Future Work

This paper presented the results of our investigation into the evolution of multistreaming regions in the universe. There were several components needed to make this possible: derivation of velocity based feature extractors based on the various descriptions found in literature, determination of a physically meaningful, time-varying threshold for finding multistreaming regions, and development of region tracking to follow the evolution of multistreaming regions and particle tracking to obtain a better understanding of the local behavior of particles. In our investigation of using particle velocity information to find multistreaming regions, we were able to reproduce results using density information alone. We have also learned that: (i) Once multistreaming starts, regions do not necessarily increase in size in a monotonous fashion, but can grow and shrink over time. (ii) Multistreaming regions are not static and migrate around. (iii) There are very interesting dynamic behaviors within multistreaming regions as well as leading to the formation of multistreaming regions. The last result suggests there may be more than one type of multistreaming region, i.e., not simply classified by size and shape, but also characterized by dynamical behavior.

The evolution of multistreaming regions in the formation of LSS is a subject of great interest in cosmology. While this paper created and applied analysis and visualization tools to identify and track multistreaming regions, it has also unearthed new questions that we would like to pursue. Some of these are: (i) Can the different velocity based feature extractors be used to identify different types of multistreaming regions? (ii) What types of behavior can one find in a multistreaming region? Do these behaviors evolve over time? Does one type of behavior become more dominant? (iii) Are certain types of behavior associated with different types of large scale structures e.g. filaments versus clusters? We believe that the tools we have created and the insights we have gained so far will be instrumental in answering these questions.

References

[1] James Ahrens, Katrin Heitmann, Salman Habib, Lee Ankeny, Patrick McCormick, Jeff Inman, Ryan Armstrong, and Kwan-Liu Ma. Quantitative and comparative visualization applied to cosmological simulations. *Journal of Physics*, Conference Series 46, SciDAC:526–534, 2006.

[2] Jordan Carlson, Martin White, and Nikhil Padmanabhan. Critical look at cosmological perturbation theory techniques. *Physics Review D*, 80(4):043531, Aug 2009.

[3] Eddy Chandra, Katrin Heitmann, James Ahrens, Salman Habib, and Alex Pang. Exploring multistreaming in the universe. Technical report, University of California, Santa Cruz, CA 95064, 2009. UCSC-SOE-09-37.

[4] Peter Coles. Gravitational instability and the formation of the supercluster-void network in the universe. *Chaos, Solitons and Fractals*, 16:513–525, 2003.

[5] Uriel Frisch and Roland Triay. Caustics and cosmological structures, 2005. `http://www.cpt.univ-mrs.fr/~cosmo/NLCP_2005/NLCP2005_Scope.html`.

[6] Naoteru Gouda. Morphology in cosmological gravitational clustering and catastrophe theory. *Progress of Theoretical Physics*, 99(1):55–68, 1998.

[7] Steve Haroz, Kwan-Liu Ma, and Katrin Heitmann. Multiple uncertainties in time-variant cosmological particle data. In *IEEE Pacific Visualization Symposium*, pages 207–214, 2008.

[8] K. Heitmann, P.M. Ricker, M.S. Warren, and S. Habib. Robustness of cosmological simulations I: Large scale structure. *Astrophysics Journal Supplement*, 160(28), 2005.

[9] R. W. Hockney and J. W. Eastwood. *Computer simulation using particles*. Taylor & Francis, Inc., Bristol, PA, USA, 1988.

[10] Stuart Levy. Partiview@ncsa. http://dart.ncsa.uiuc.edu/partiview/.

[11] Hongwei Li, Chi-Wing Fu, Yinggang Li, and Andrew Hanson. Visualizing large-scale uncertainty in astrophysical data. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1640–1647, 2007.

[12] Zarija Lukic, Darren Reed, Salman Habib, , and Katrin Heitmann. The structure of halos: Implications for group and cluster cosmology. *The Astrophysical Journal*, 692(1):217–228, 2009.

[13] R. Mohayaee, S. Colombi, B. Fort, R. Gavazzi, S. Shandarin, and J. Touma. Caustics in dark matter haloes. *EAS Publications*, Series 20:19–24, 2006.

[14] Paul Navratil, Jarrett Johnson, and Volker Bromm. Visualization of cosmological particle-based datasets. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1712–1718, 2007.

[15] P.J.E. Peebles. The standard cosmological model, 1998. http://nedwww.ipac.caltech.edu/level5/Peebles1/paper.pdf.

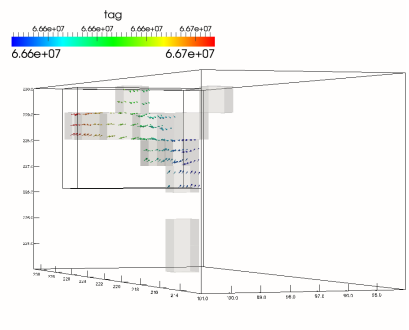
[16] Adrian Pope, Salman Habib, Zarija Lukić, David Daniel, Patricia Fasel, Nehal Desai, and Katrin Heitmann. The accelerated universe. *Computing in Science & Engineering*, 12(4):17–25, 2010.

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

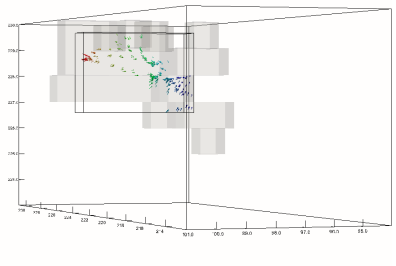
[18] Sergei Shandarin. The multi-stream flows and the dynamics of the cosmic web, 2010. http://arxiv.org/abs/1011.1924.

[19] D. Silver and X. Wang. Volume tracking. In *Visualization '96. Proceedings.*, pages 157 –164, 27 1996-nov. 1 1996.

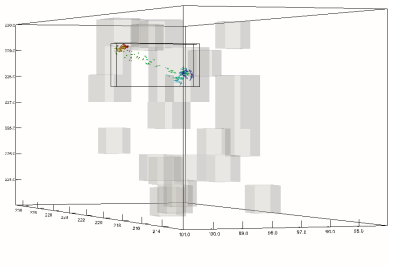
[20] Taihei Yano, Hiroko Koyama, Thomas Buchert, and Naoteru Gouda. Universality in the distribution of caustics in the expanding universe. *The Astrophysical Journal Supplement Series*, 151:185–192, 2004.



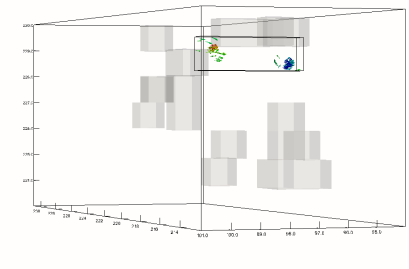
Frame 13, $a = 0.033$



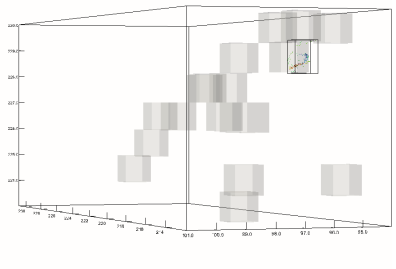
Frame 50, $a = 0.106$



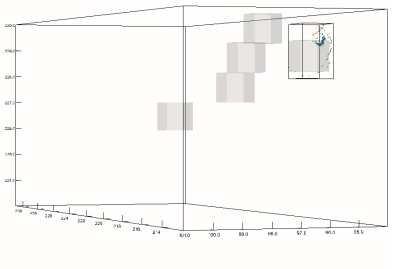
Frame 100, $a = 0.206$



Frame 200, $a = 0.405$

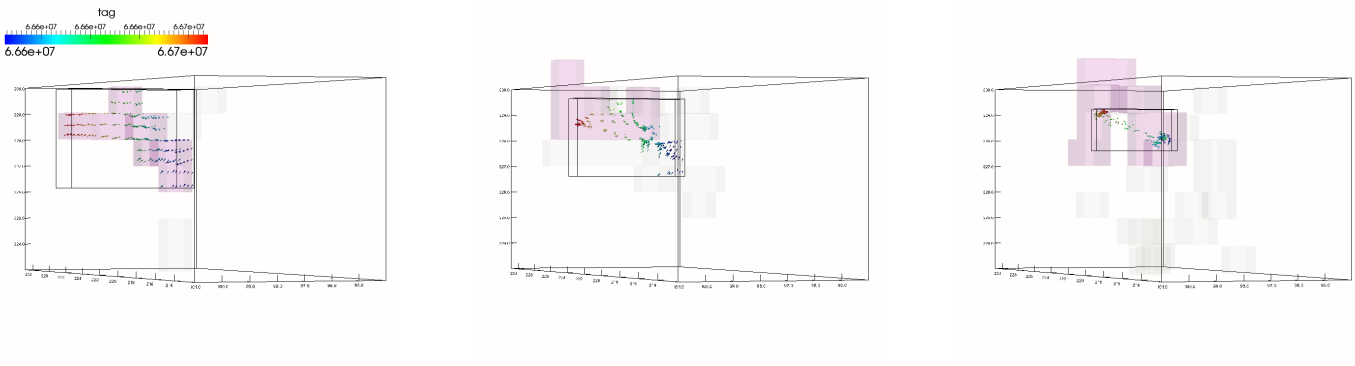


Frame 400, $a = 0.803$

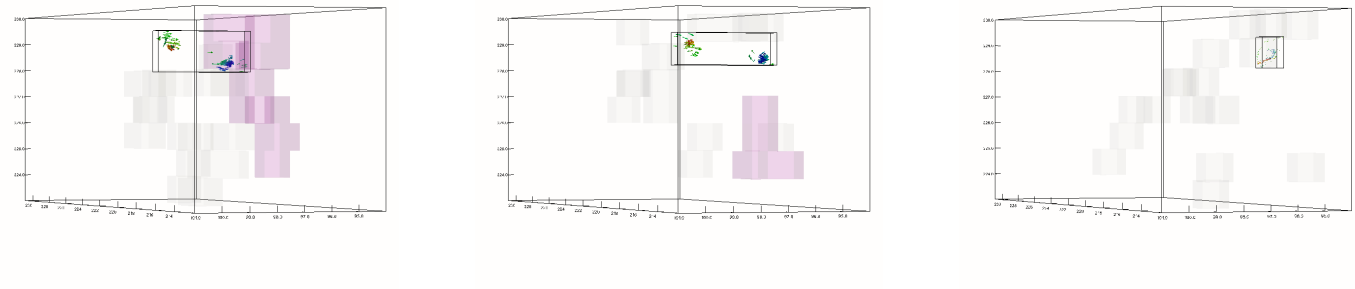


Frame 499, $a = 1.0$

Fig. 7. Region finding: In this sequence, we overlay the results of region finding performed at each frame. The regions that are found to be multistreaming are represented by transparent grey boxes. In frame 13, we see that in addition to the region that contains the particles that are traced in Figure 6, there is another smaller region below it. In frame 50, there is only one large multistreaming region. In subsequent frames, there are more regions. Note that these regions are not tracked. Instead they are extracted using a combination of vorticity and density criteria. Two interesting observations can be made from this sequence. First, the particles are not necessarily contained within multistreaming regions beyond frame 13. For example, in frame 200, the particles do not meet the combined criteria for multistreaming. Secondly, the behavior of the particles are influenced by neighboring multistreaming regions. For example, when the multistreaming region that contained the particles in frame 100 disappeared, the particles were “orphaned”. However, they were eventually adopted and absorbed by other nearby multistreaming regions.

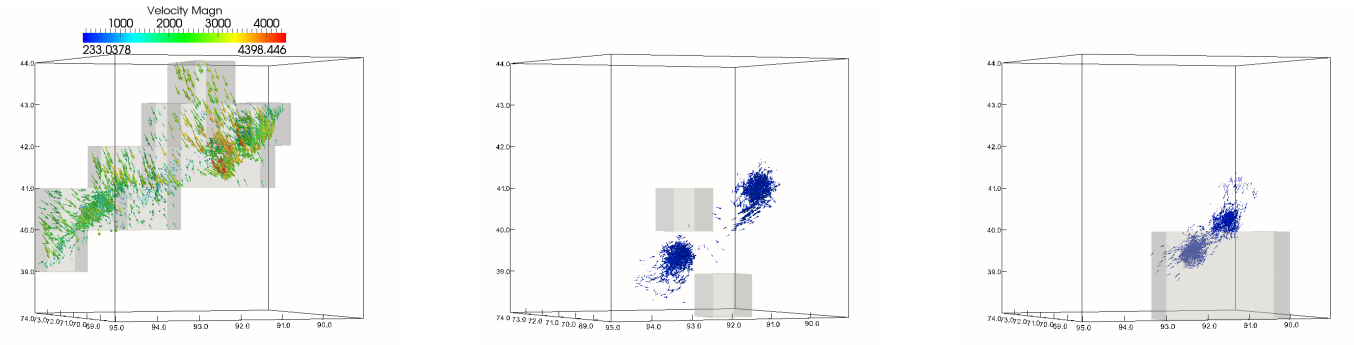


Frame 13, $a = 0.033$ Frame 50, $a = 0.106$ Frame 100, $a = 0.206$



Frame 150, $a = 0.306$ Frame 200, $a = 0.405$ Frame 400, $a = 0.803$

Fig. 8. Region tracking: In this sequence we track the grey region that contained the particles in frame 13 and display it in pink. The results from region finding (Figure 7) and particle tracking (Figure 6) are kept for reference. The evolution of this pink region is tracked by examining its neighborhood as described in Section 4.5. What we can observe from this sequence is that particles in a multistreaming region do not always stay with the same region. In fact, the particles joined a different multistreaming region as can be seen in frame 400. We also note that after frame 200, the tracked multistreaming region in pink, eventually moved out of the outer bounding box initially established in Figure 6.



Frame 101, $a = 0.208$ Frame 350, $a = 0.703$ Frame 499, $a = 1.0$

Fig. 9. Using a combination of dot product and density criteria, a multistreaming region was extracted in frame 101. The active multistreaming regions at each time frame were extracted and represented by the grey transparent boxes. 2006 particles from the multistreaming region in frame 101 were tracked. We can observe that they split into two groups but later reformed into a single group which also happen to coincide with a multistreaming region in frame 499. Unlike the sequence from Figure 6, we do not observe strong spiral behavior.