

wordt
NIET
uitgeleend

Topology of galaxy models

Master's thesis Computer Science,
Rijksuniversiteit Groningen

Bob Eldering

February 15, 2006

Supervisors:

Prof.Dr. Gert Vegter
Institute for Mathematics and Computing Science
University of Groningen

and

Prof.Dr.Rien van de Weygaert
Kapteyn Institute
University of Groningen

Co-supervisor:
Drs. Nico G.H. Kruithof
Institute for Mathematics and Computing Science
University of Groningen

Rijksuniversiteit Groningen
Bibliotheek FWN
Nijenborgh 9
9747 AG Groningen

Acknowledgement

I would like to thank everybody who helped making this project possible. Next to all their other assistance, I would especially like to thank Erwin Platen for providing the data sets, Nico Kruithof especially for the brainstorming sessions and Gert Vegter and Rien van de Weygaert for all their advise and suggestions.

Contents

1 Introduction	5
2 Related work	9
2.1 Shape analysis	9
2.2 State of the art in astrophysical shape analysis	9
3 Astrophysical background	15
4 Voronoi diagram and Delaunay triangulation	19
4.1 Voronoi diagram	19
4.2 Delaunay triangulation	20
4.3 Duality	22
5 Alpha shapes	24
6 Betti numbers	28
6.1 Definition	28
6.2 Algorithm	29
7 Betti numbers for periodic data	32
7.1 2-cycle detection	32
7.2 Proof of correctness	33
8 Heuristic and kinematic Voronoi models	37
9 Results	52
10 Conclusions and future work	71
10.1 Astrophysical	71
10.2 Topological	71
A Graphs	77

B Implementation	87
B.1 alpha_shapes.C	87
B.2 defines.h	87
B.3 make_periodic.h	87
B.4 output_betti.h	88
B.5 output_components.h	88
B.6 read_data.h and read_voronoi.h	88
B.7 typedefs.h	88

Abstract

Galaxy Redshift survey's have shown us that galaxies are not distributed evenly in our universe. They form a web-like structure, called the cosmic web. A large number of models exist to explain this structure. To compare these models one has to quantify the information content hidden in the cosmic web predicted by the model. A new method to quantify this information will be discussed: with the Betti numbers of alpha shapes. The alpha shape of a point set describes the intuitive notion of the shape of that point set and can be defined as a subcomplex of the Delaunay triangulation of the point set. The Betti numbers count the number of components, tunnels and holes of the alpha shape. To apply this method to data sets representing galaxy distributions, an existing algorithm to compute Betti numbers of alpha shapes has been generalized to be able to handle periodic data. This new algorithm will be presented together with the results of applying it to galaxy distribution data sets.

1 Introduction

Observations of the universe reveal that matter within it clusters on a variety of scales [27]. On scales between 10 and 100 Mpc (1 pc = 1 parsec = 3.2616 light years), the universe is spanned by a network of superclusters interspersed with large and almost empty regions, called voids. The morphology, the structure and form, of the supercluster-void network defines a complex pattern and has inspired evocative descriptions such as being 'honeycomb-like', 'bubble-like', 'a filamentary network', 'Swiss cheese', 'cosmic web', etc [27], see Figure 1 for an impression.

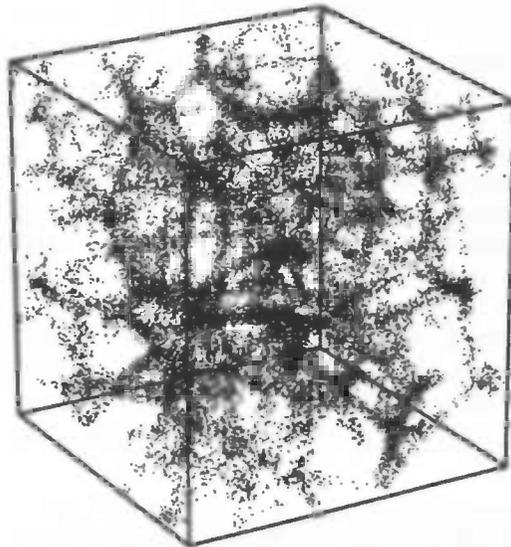


Figure 1: A distribution of galaxies, showing the cosmic web [22]. This distribution is made by a computer model for the evolution of the distribution of galaxies as described in Section 8.

The Megaparsec scale structure has been revealed by galaxy redshift surveys [27], where the distances to the galaxies are estimated on the basis of their redshift. See Figure 2 from [25] for an impression. This structure is called the large scale structure. One of the consequences of the Hot Big Bang theory (see Section 3) is that the

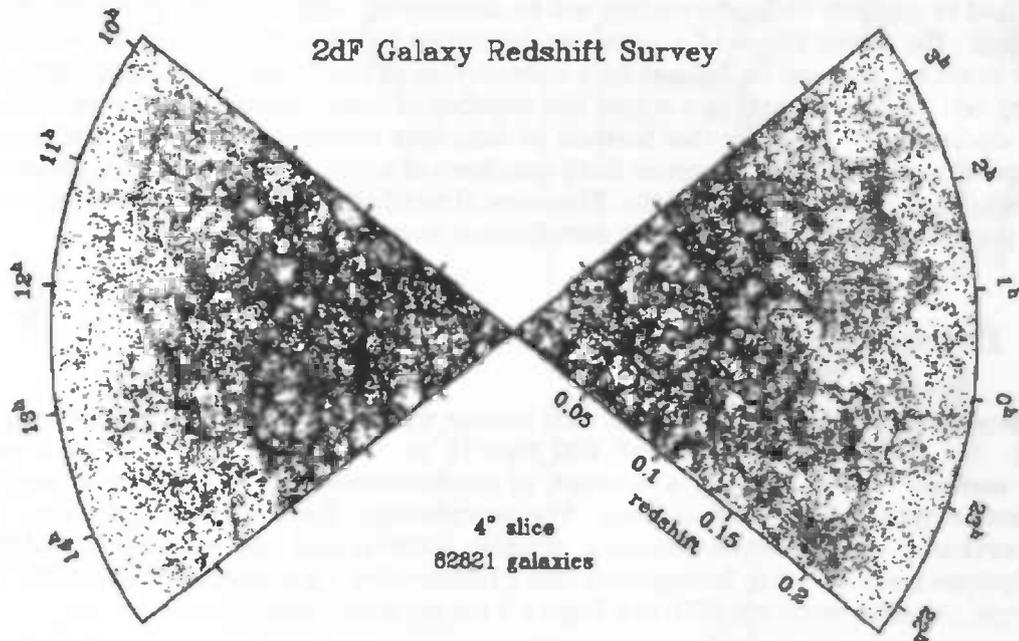


Figure 2: Slices of the 2dF galaxy redshift survey in northern and southern galactic hemisphere are shown. The supercluster-void network visually stands out.

matter distribution in the universe is homogeneous and isotropic when averaged over very large scales (the average density is the same everywhere and in every direction). However on smaller scales we can see structure in the distribution of galaxies, a large number of different cosmological scenarios exist to explain this structure [30]. Each of these scenarios predicts a different evolution from the initial conditions. The patterns in the cosmic matter distribution are very sensitive to the underlying scenario. This allows us to test whether the observed large scale clustering properties of galaxies are consistent with the predictions of the preferred cosmological models. To do this, one has to quantify the information content hidden in the distribution data of galaxies [25].

In this paper we will quantify the topological features of the large scale structure using the Betti numbers of alpha shapes. The alpha shape of a finite point set is a polytope that is uniquely determined by the set and a real number α [8]. Taking the galaxy distribution to be the point set, the alpha shape expresses the intuitive notion of the shape of the structure of the galaxy distribution, see for example Figure 13. Figure 53 shows more clearly how alpha shapes capture the intuitive notion of the shape of a point set. The α value will control the level of detail reflected by the polytope, Figure 3 shows how alpha shapes reveal the shape of a galaxy distribution. In the most intuitive

terms, the Betti numbers for 3D data count the number of components, the number of tunnels and the number of holes. Note that the Betti numbers are only concerned with the topology of the shape; they will only represent how the shape is connected. No information is hidden within the Betti numbers about the geometry of the shape. A common narrative in this context is that a coffee cup and a donut have the same Betti numbers, however they clearly have different geometry.

Therefore, next to the Betti numbers we will also calculate the volume and surface area of the alpha shapes. Functions of these values will also be investigated to see how we can use these values to discriminate between models. For example:

- the dimensionless value $\frac{V}{S^{\frac{3}{2}}}$ will be plotted against the α value, where V and S are the volume and surface area respectively,
- a log log plot will be made of β_0 and α , since that makes it easier to discriminate between the models,
- the volume will be plotted against the surface area,

All these calculations will be done for galaxy distributions generated in a simulation of a simplified computer model for the evolution of the distribution of galaxies, based on the Voronoi diagram [31] as described in Section 4. This model has the advantage that its process of creating distributions is well understood; we know what the large scale structure of the resulting distribution looks like. Figure 1 displays a distribution made by this model. Because we have this extensive knowledge of the formation process used by this model, using this model will allow us to get a better insight into the meaning and significance of Betti numbers of alpha shapes as characteristics of the patterns in the cosmic web.

A particular focus of our investigation of alpha shapes and Betti numbers concerns the issue of their calculation in the case of periodic data sets. An algorithm by Delfinado and Edelsbrunner for non-periodic data sets [7] will be generalized to be able to handle periodic data, since the computer models of cosmic structure formation usually produce galaxy distributions with periodic boundary conditions.

The structure of the article will be as follows:

- the current scientific situation in astrophysics and computational geometry, relevant to quantifying the structure of the cosmic web, will be discussed in Sections 2 and 3,
- the necessary subjects from computational geometry will be introduced in Sections 4, 5 and 6,
- the the adjusted algorithm to calculate Betti numbers for periodic data sets will be proposed in Section 7,
- the computer model producing the used data sets will be explained in Section 8
- the remaining sections, 9 and 10, will be used to evaluate the usefulness of Betti numbers.

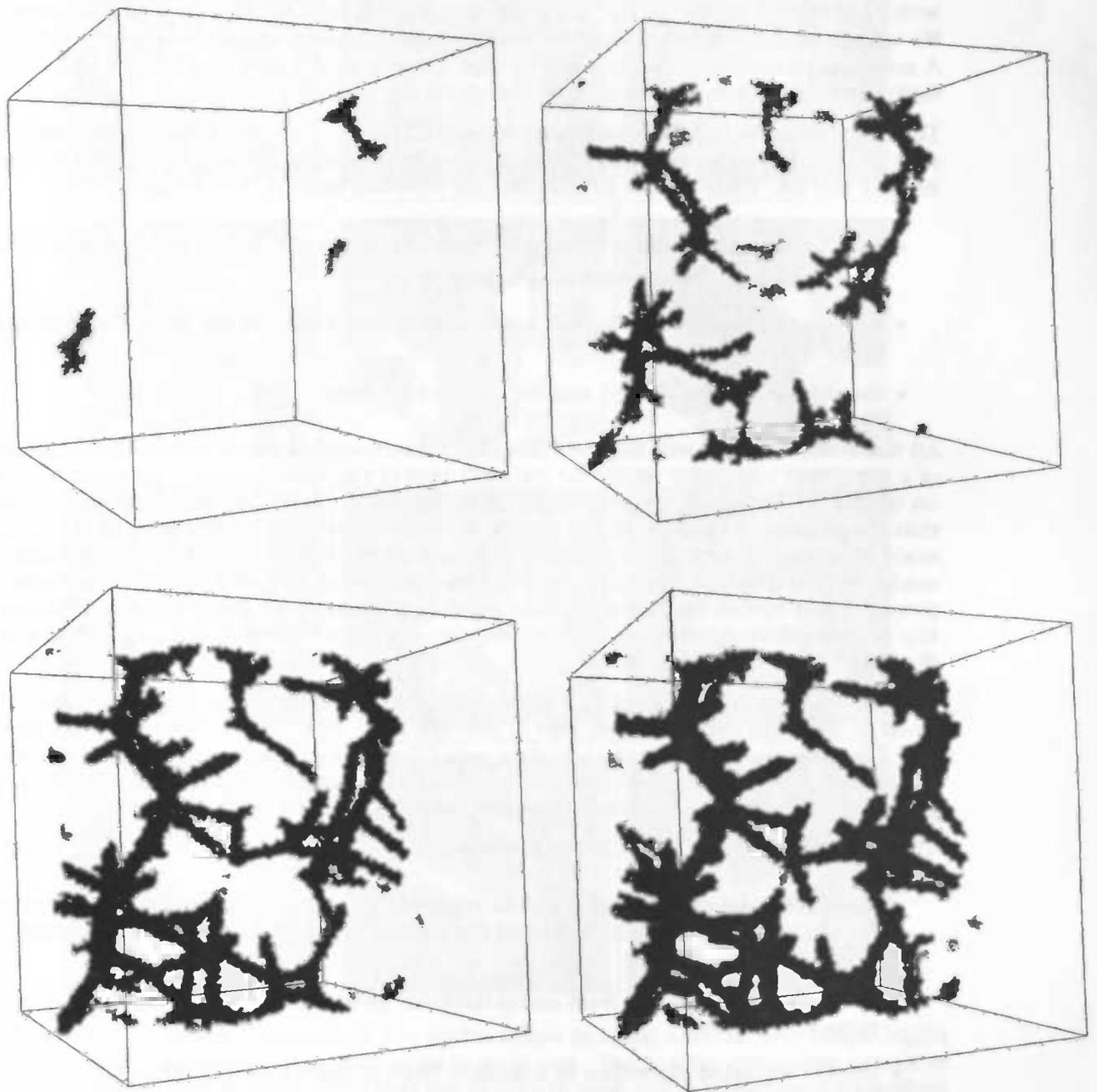


Figure 3: Four alpha shapes of the same data set, for increasing value of the alpha parameter

2 Related work

2.1 Shape analysis

Sample based shape (re-)construction has gained popularity in recent years because of its wide applicability in science and engineering [11]. A fair amount of related work has been done for point sets in \mathbb{R}^2 , and some for point sets in \mathbb{R}^3 [9]. Jarvis (1977) was one of the first to consider the problem of computing the shape as a generalization of the convex hull of a planar point set. A mathematically rigorous definition of alpha shape was later introduced by Edelsbrunner et al. (1983). Two-dimensional alpha shapes are related to the dot patterns of Fairfield (1979; 1983) and the circle diagrams used in bivariate cluster analysis (see for example Moss (1967)). Different graph structures that serve similar purposes are the Gabriel graph (Matula and Sokal 1980), the relative-neighborhood graph (Toussaint 1980), and their parameterized version, the β -skeleton (Kirkpatrick and Radke 1985). For \mathbb{R}^3 , Boissonnat (1984) suggested the use of Delaunay triangulations in connection with heuristics to "sculpture" a single connected shape of a point set. More recently, Veltkamp (1992) also generalized the above-mentioned two-dimensional graph structure to three dimensions, calling them γ -graphs. Finally, note the superficial similarity between alpha shapes and isosurfaces in \mathbb{R}^3 . The latter is a popular concept in volume visualization (see for example Drebin et al. (1988) and Lorensen and Cline (1987)).

Recently, flow shapes introduced by Giesen and John [12] and also by Edelsbrunner [10] provide another means of creating shapes out of a point set. In fact both alpha shapes and flow shapes can be used to define a hierarchy of shapes from a set of points [11]. Alpha shapes put a ball of radius $\sqrt{\alpha}$ around each point and construct a simplicial complex that respects the intersections among these balls, see Section 5 for details. The underlying space of this simplicial complex is defined as the alpha shape. As α changes, new simplices are added or deleted as intersections among the balls appear or disappear. This means a hierarchy of shapes can be defined using α as a scale parameter. Flow shapes, on the other hand, are defined as a cell decomposition of the ambient space of the sample points, see [12]. The decomposition is based on the gradient flow of the distance function to the point set. The cells of this decomposition can be ordered by some distance values giving a hierarchy of shapes. While both alpha shapes and flow shapes define a hierarchy of shapes using some scale parameter, they are quite different geometrically as can be seen in Figure 4.

The two hierarchies have a certain topological similarity, namely they are homotopy equivalent. Specifically, both alpha shapes and flow shapes change their topology only at discrete critical levels in the hierarchy. These critical levels turn out to be the same in both hierarchies. This means that calculating Betti numbers for both shapes will give the same results.

The alpha shape and Betti numbers will be further explained in the following sections.

2.2 State of the art in astrophysical shape analysis

However, all these methods from computational geometry are not generally used in astrophysics. In cosmology, it is a conventional practice to quantify the clustering of

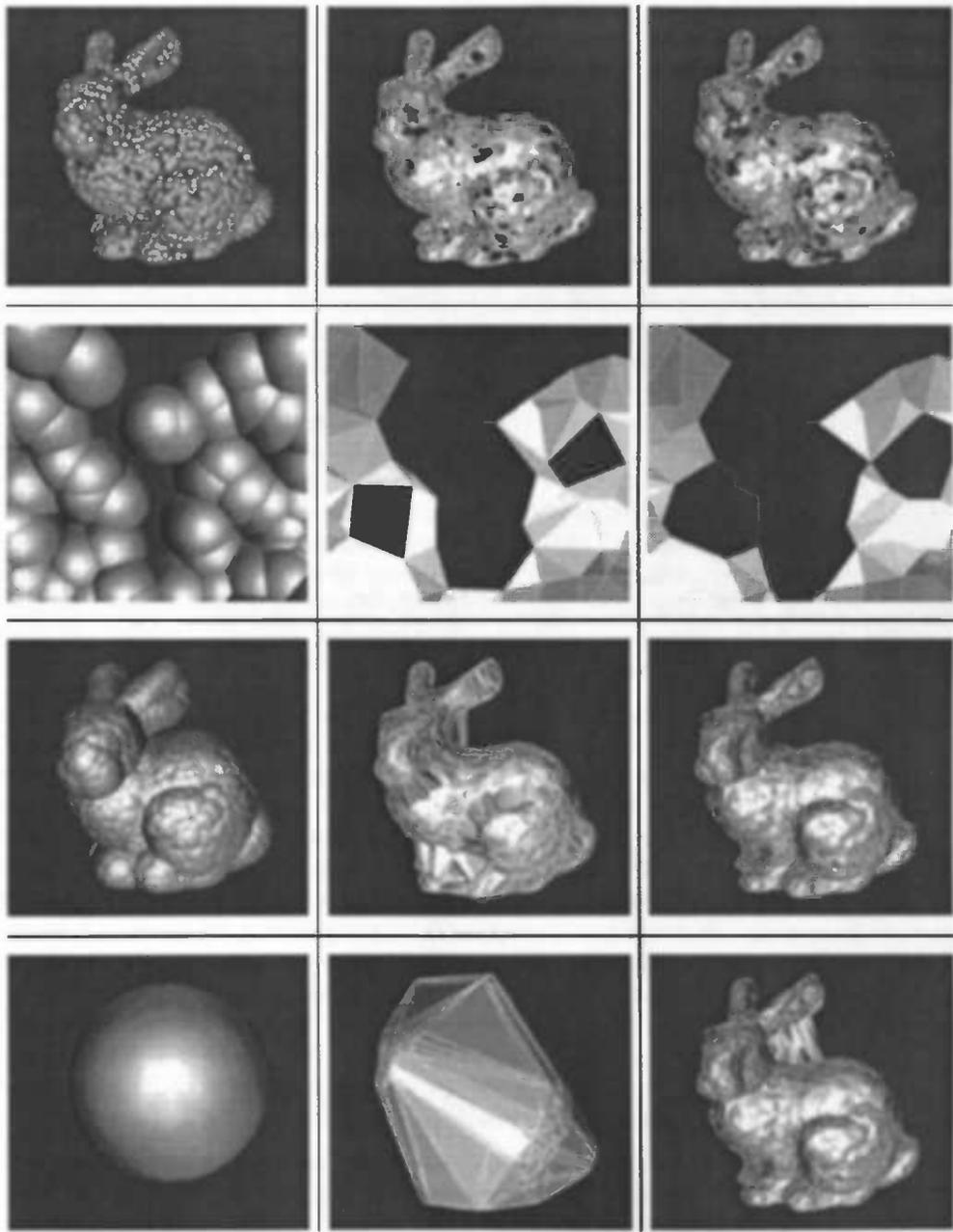


Figure 4: The union of balls (left), the alpha shape (middle) and the flow shape (right) for increasing values of α (top to bottom) in each row . The second row shows a zoom of the pictures in the first row. Note that the shapes in each row are homotopy equivalent [11].

structure on large scales using the two-point correlation function or, equivalently, the power spectrum [25]. The following definition is often used for the two-point correlation

function: Given a random galaxy in a location, the correlation function quantitatively describes the probability that another galaxy will be found within a given distance [20]. It can be thought of as a lumpiness factor: the higher the value for some distance scale, the more lumpy the universe is at that distance scale.

However, the observed cosmic weblike structures are clearly non-linear patterns and for a full characterization higher order clustering measures are necessary in order to effectively compare the theoretical predictions for galaxy distribution with the data from redshift surveys [25].

A number of statistical measures have been proposed to quantify the pattern made by galaxies as they cluster in our universe. However, these measures are heuristic in nature. Prominent among these measures are Percolation Analysis (Zeldovich, Einasto & Shandarin 1982; Shandarin & Zeldovich 1983), Counts in Cells (Janes & Demarque 1983; de Lapparent, Geller & Huchra 1991), Minimal Spanning Trees (Barrow, Sonoda & Bhavsar 1985), the Genus measure (Gott, Melott & Dickinson 1986) etc [27].

Recently (1993) Mecke, Bucherf and Wagner have introduced the Minkowski functionals to cosmology in [3]. In this paper they point out that both topological and geometrical descriptors, respectively characterizing the connectivity and the shape of figures, are required to specify the full morphology of spatial patterns like the cosmic web [3]. The authors of [3] argue that integral geometry supplies a suitable family of such descriptors, the Minkowski functionals (what the Minkowski functional describe will be discussed in detail below).

To calculate the Minkowski functionals a shape needs to be constructed from the galaxy distribution data, as discussed in the previous subsection. In [27] the authors of this paper discuss how the Minkowski functionals can be calculated using a method called SURFGEN to create a shape from the point set, see Figure 6. However to create this shape the SURFGEN method requires a density function approximated from the data points. The density function described in [27] is created by smoothing the density of the particles over cubic lattice using a cloud-in-cell method (see Figure 5 from [15] for an explanation of the cloud-in-cell method). With this density function SURFGEN

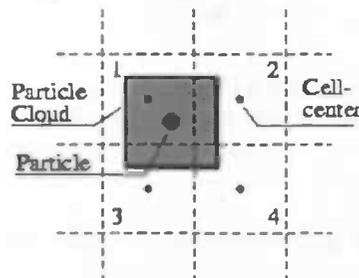


Figure 5: The mass of the particle is distributed over the cells. The amount of mass contributed to a cell is proportional overlapping area (or volume in 3D)

creates a polyhedral approximation to an isodensity surface using the marching cubes algorithm [17]. Next to the density function, a cube size has to be chosen for the cloud-in-cell and marching cube method. The geometric and in particular the topologic

properties of the constructed surface are sensitive to this cube size, while this value is picked on an ad hoc basis. Of course the isodensity surface construction requires an isodensity value to be chosen. In [27] this isodensity value is varied to obtain a series of measurements, just like we will vary the α value of the alpha shape in this paper.

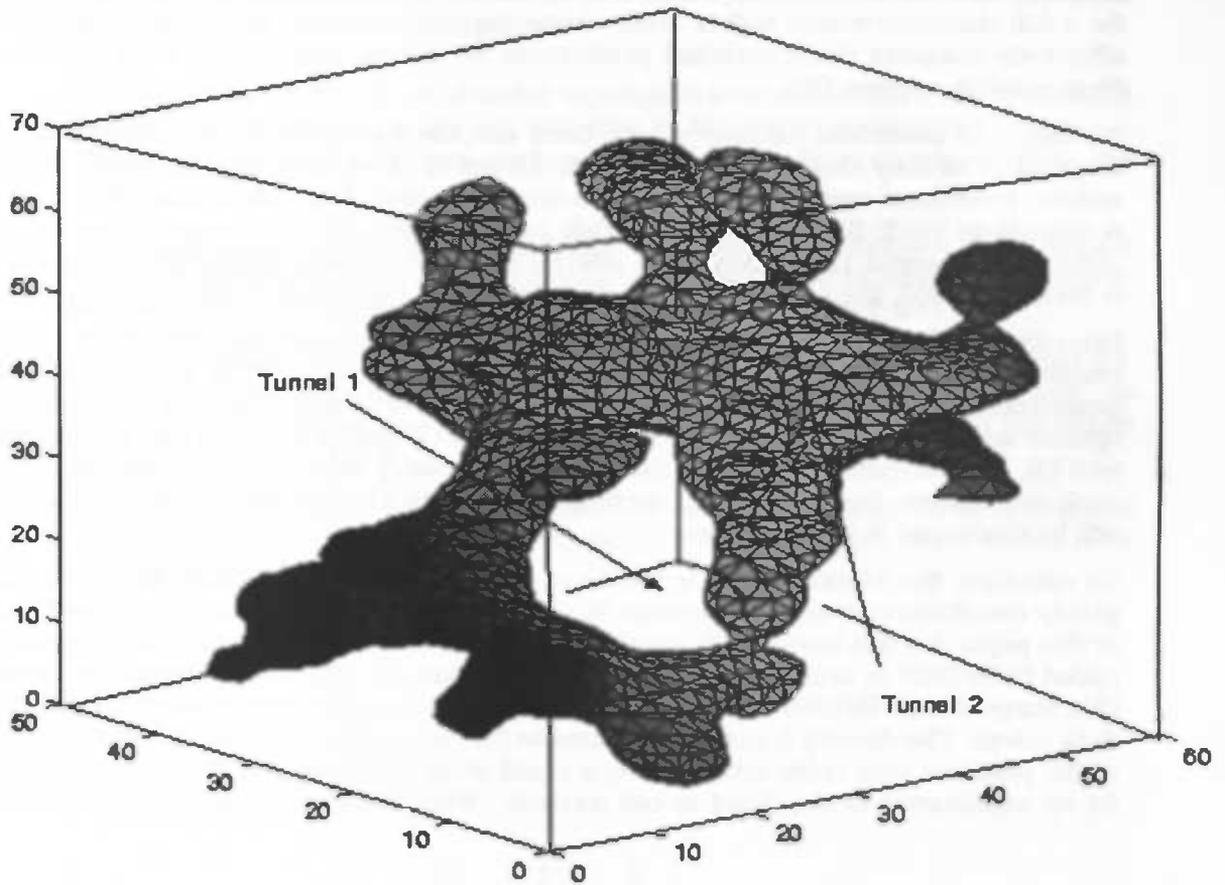


Figure 6: A cluster appearing in the Λ CDM simulation is triangulated using SURFGEN (from [27]). The tunnels pointed out in the figure are detected by hand. We will detect them automatically.

The following four Minkowski functionals describe some geometric and topological properties of the shape constructed by SURFGEN:

1. area S of the surface
2. volume V enclosed by the surface
3. integrated mean curvature C of the surface

$$C = \frac{1}{2} \oint \left(\frac{1}{R_1} + \frac{1}{R_2} \right) dS,$$

where R_1 and R_2 are the principal radii of curvature at a given point on the surface

4. integrated intrinsic curvature χ of the surface, also called the Euler characteristic

$$\chi = \frac{1}{2\pi} \oint \left(\frac{1}{R_1 R_2} \right) dS$$

The Euler characteristic can also be defined using the Betti numbers (see Section 6) of the same shape in an alternating sum:

$$\chi = \beta_0 - \beta_1 + \beta_2 - \dots$$

While the Euler characteristic and the Betti numbers give information about the connectivity of the shape, the other three Minkowski functionals are sensitive to local surface deformations. The Minkowski functionals give information about geometric and topological properties of the shape, as described above, while Betti numbers will only give information about the topological properties. The Betti numbers, on the other hand, are more detailed in their description of the topology; the Euler characteristic will only give a "summary" of the topology. Homeomorphic surfaces have the same Euler characteristic, the other way around is not generally true however. Here one can intuitively describe two surfaces to be homeomorphic if a continuous stretching and bending of one of the shapes into the other shape is possible.

As demonstrated in Sahni et al. (1998) and Sathyaprakash et al. (1998) the ratios of Minkowski functionals called "Shapefinders" provide an excellent measure of morphology [27], that is they quantify the information content hidden in the shape very well. On the basis of the Minkowski functionals the authors of [27] also define the "Shapefinders", T (Thickness), B (Breadth) and L (Length) which are defined as follows:

$$T = \frac{3V}{S}, \quad B = \frac{S}{C}, \quad L = \frac{C}{4\pi(G+1)},$$

where G is the genus. The genus of a triangulated closed polyhedral surface is given by the expression:

$$G = 1 - \frac{\chi}{2},$$

where χ is the Euler characteristic of the triangulated surface. The genus G of a connected surface is an integer representing the maximum number of cuttings along closed simple curves without rendering the resultant manifold disconnected. The three Shapefinders (describing an object) have dimensions of length and provide an estimate of the "extension" of the object along each of the three spatial directions. For simple shapes the Shapefinders are sometimes interpreted as follows: T describes the characteristic thickness of the object, L characterizes the length of the object; B can be associated with the breadth of the object. For example, a triaxial ellipsoid has the values of T , B and L closely related to the lengths of the three principal axes: shortest, intermediate and the longest respectively. However the shapefinders are just estimates and the relation $T \leq B \leq L$ is not generally true.

An excellent indicator of "shape" is provided by the dimensionless Shapefinder statistic:

$$P = \frac{B - T}{B + T}, \quad F = \frac{L - B}{L + B},$$

where P and F are measures of Planarity and Filamentarity respectively ($P, F \leq 1$). A sphere has $P = F = 0$, an ideal filament (that is a quasi one-dimensional object, not necessarily straight) has $P = 0, F = 1$ while $P = 1, F = 0$ for an ideal pancake (a quasi two-dimensional object that can be curved) [26].

To summarize the differences between the SURFGEN approach and the Betti numbers of alpha shapes and their characteristics:

- Betti numbers contain more detailed topological information than the Minkowski functionals, however they are not sensitive to the actual geometry of the shape. That is, as long as two shapes are homeomorphic, the Betti numbers will be the same.
- The Shapefinders provide means to directly relate visual inspection of the shapes to the Minkowski functionals of the shapes. For the Betti numbers this relation is hard to find and needs further investigation (see Sections 9 and 10).
- The alpha shape construction method is mathematically rigorously defined [9]. It does not use heuristic parameters like a density function or cube size, as the SURFGEN method does. Instead the alpha shape method creates the shape directly from the point set.
- Both methods produce a serie of shapes representing different levels of detail of the structure hidden in the point set. Both methods actually analyse this serie instead of one shape.

3 Astrophysical background

In this overview of the astrophysical background I have based myself in particular on [28]:

The story of structure formation begins more than 13 billion years ago, roughly 380,000 years after the Big Bang, when the universe was very different from today. There were no stars, galaxies or webs yet, just a uniform soup of free-floating protons and electrons. In fact, the gas was so evenly distributed that its peak densities differed by only 1 part in 100,000 from the cosmic average. These tiny differences can now be seen in the cosmic microwave background radiation as measured by the WMAP satellite in 2003 [30].

As the universe expands it also cools down. Around 380,000 years after the Big Bang, the universe was cool enough for the plasma to form hydrogen atoms. The matter formed was still distributed nearly homogeneous and isotropic. Currently the matter distribution in the universe is still homogeneous and isotropic when averaged over very large scales. However on scales of dozens of Megaparsecs we can see structure in the distribution of galaxies; it is clearly anisotropic.

This transition from almost smooth to clearly anisotropic can be explained with gravitational instability. Minimal fluctuations in the density field of the mass existed in the initial conditions. These fluctuations cause the gravity in high density regions to be slightly higher than in low density regions. Therefore the mass will be attracted towards the high density regions, causing an even larger density difference, which will enhance the process even further. This self enhancement is the reason this process is called gravitational *instability* [30].

A second ingredient needed to explain the observations of the structure formation process is dark matter. Dark matter is quite different from the ordinary matter that makes up stars, planets and people. Not only does dark matter not shine, it interacts with "our" kind of matter only through the force of gravity. It is largely believed to consist of particles that have no other effects on ordinary atoms and molecules. Furthermore, dark matter appears to outweigh normal matter throughout the universe by a factor of four to one. This notion is indeed odd, and it has met with resistance since it was first suggested by the eccentric astronomer Fritz Zwicky in the 1930s. However, cosmologists have now grown to accept its existence as nearly certain in the face of overwhelming evidence from a variety of observations. Although we may not understand exactly what dark matter is, we do understand what it does - it holds galaxies together, bends light, slows down the universe's expansion and drives the formation of intergalactic structure.

To understand this last point, we need to return to the early history of the universe. During the first 380,000 years, the relic heat from the Big Bang kept the universe so hot (greater than 3,000 kelvins) that electrons and protons in the primordial soup could not combine to form neutral hydrogen atoms. Such ionized gas, in this case consisting of dissociated electrons and protons, is known as a plasma. When plasma particles are in their free-floating state, they can interact with light, exchanging energy and momentum. In the early universe, this scattering increased the gas pressure within the cosmic soup. So, when gravity tried to collapse the first density perturbations, the gas pressure pushed back. As long as the electrons and protons were separated, the

gas could not form larger structures. Instead, the potential structures churned and oscillated as the inward pull of gravity fought the outward push of gas pressure.

Then, when the universe was 380,000 years old, a major event took place. As the universe was expanding, it was also cooling, and at this point it became cold enough for electrons and protons to combine, forming hydrogen atoms. Suddenly, these new atoms became decoupled from the photons - they no longer interacted so strongly with light - which drastically reduced the pressure that had kept gravity at bay. With gravity free to work on all the newly formed hydrogen atoms, structures could form in earnest.

While the protons, electrons and photons were oscillating under the competing influences of gravity and pressure, the dark matter followed a different storyline. Because dark matter interacts with normal matter only through gravity, the pressure that kept the normal gas from collapsing could not act on it. Particles of dark matter enjoyed an unimpeded assembly into large structures long before the normal gas could begin to get organized. By the time normal matter decoupled from the photons, the dark matter had already grown into a primitive web-like network. As soon as the normal matter lost its support from the photon pressure, the gravity from the pre-existing dark-matter structures quickly pulled normal gas into the web. In this way, normal matter was given a gravitational "head-start" by the dark matter; dark matter is the key to the structure formation process.

Once this process was set in motion, the gravitational building blocks of the intergalactic medium were in place. Normal and dark matter continued to free-fall toward concentrations of mass until the rising gas pressure slowed the infall. The cosmic web, as we can see it now in redshift survey's, was taking shape.

The forming of the objects in this cosmic web takes place in a number of stages. Since the universe as a whole is expanding, as discovered by Edwin Hubble in 1929 and is therefor called the Hubble expansion H_0 [30], the contractions of matter in higher density regions have to slow down this global expansion in the first stage. Slowly the density fluctuations will be large enough for mass to actually contract, against the global expansion. In the final stage the mass totally collapses and tries to find a steady state by exchanging energy [30]. This last process is called virialization

An important aspect of this gravitational formation process is the inclination to progress via stages in which the cosmic matter distribution settles in striking anisotropic patterns [31]. Aspherical overdensities, on any scale and in any scenario, will contract such that they become increasingly anisotropic, as long as virialization has not yet set in. At first they turn into a flattened pancake, possibly followed by contraction into an elongated filament. Such evolutionary stages precede the final stage in which a virialized object, for example a galaxy or cluster, will emerge. Together these pancakes, filaments and clusters form the large scale structure known as the cosmic web. The tendency to collapse anisotropically finds its origin in the intrinsic primordial flattening of the overdensity, augmented by the anisotropy of the gravitational force field induced by the external matter distribution, that is by tidal forces. Naturally, the induced anisotropic collapse has been the major agent in shaping the cosmic foamlike geometry. An impression of this generic evolutionary scheme can be obtained from Figure 7 which illustrates the process by means of an N-body computer simulation.

This all explains the cosmic web we observe today. Now we have to look at what we can learn from the cosmic web. Within this theory of structure creation a large

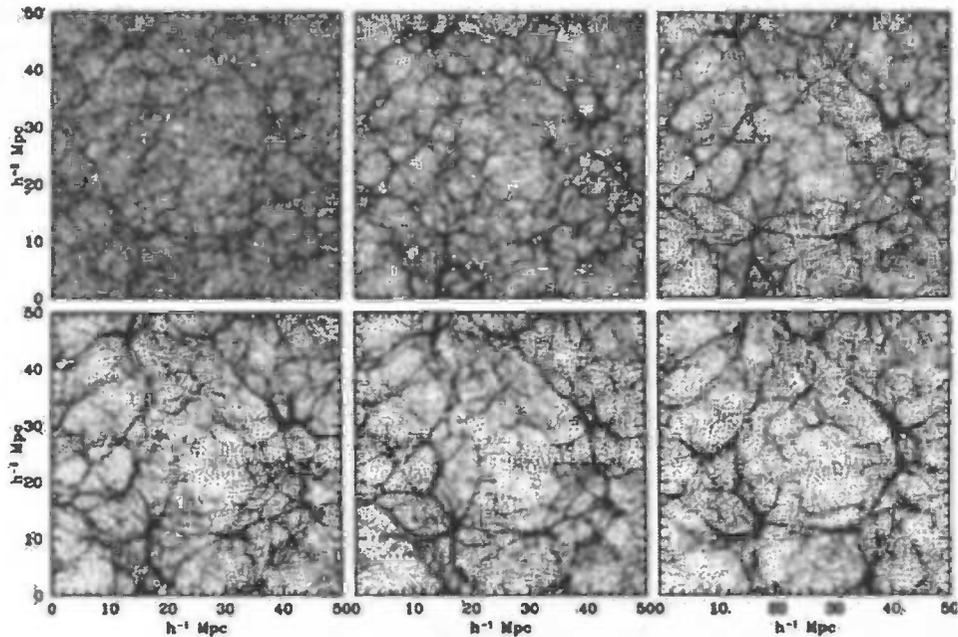


Figure 7: A computer simulation of the forming of structure in the universe. The evolution of the web-like structure of the anisotropic filaments, walls and clusters stands out [30].

number of cosmological scenarios exist [30]. Each of these scenarios predicts a different evolution from the primordial density field [30]. Important to these scenarios are global cosmological properties, such as the value of the cosmological constant Λ and the expansion speed of the universe, expressed by the Hubble constant H_0 . The cosmological constant is an extra term in Einstein's equations of general relativity which physically represents the possibility that there is a density and pressure associated with "empty" space [19].

Even more important to the structure evolution are the statistical properties of the primordial density fluctuations. They will, for example, determine whether structure appears at large scales first followed by structure showing on repeatedly smaller scales (this will happen if density fluctuations are relative strong on large scales and is called the top-down view) or exactly the opposite: small scales structure first, followed by larger scale structure (the bottom-up view). During the 1970s, perhaps fueled by Cold War politics, these two competing theories of structure formation emerged, the top-down view devised by Yakov Zel'dovich at the School of Russian Astrophysics in Moscow, and the other by James Peebles and his collaborators at Princeton University. Observations have proved the bottom-up variant to be correct for the "normal" baryonic matter. However for dark matter this is still an object of study.

Another statistical property of the primordial density field important to the evolution of the cosmic web is the average mass density of the universe, since this density will determine whether a density fluctuation is large enough to develop itself against the global expansion of the universe.

So we can gain important knowledge about global cosmological parameters from the cosmic web. An important tool to extract this information, vital to the forming of cosmic web structure, from the cosmic web, is to quantify the information content hidden in this cosmic web structure. With such a tool we could compare models, representing different cosmological scenarios, and even validate them against redshift surveys. In this paper we will discuss a method to quantify galaxy distributions, made by computer implementations of the models, using the Betti numbers of alpha shapes.

4 Voronoi diagram and Delaunay triangulation

4.1 Voronoi diagram

A Voronoi diagram of a set of sites (points) is a collection of regions that subdivides the space. Each region corresponds to one of the sites, and all the points in one region are closer to the corresponding site than to any other site. Given a set of sites $P = \{p_1, p_2, \dots, p_n\}$ in a space S the region $V(p_i)$ corresponding to site p_i is defined as:

$$V(p_i) = \{x \in S \mid \forall 1 \leq j \leq n, j \neq i : d(p_i, x) \leq d(p_j, x)\}$$

Where $d(x, y)$ is the Euclidian distance between the points x and y . See Figure 8 for an example of a Voronoi diagram. Another way to define the Voronoi regions is by looking

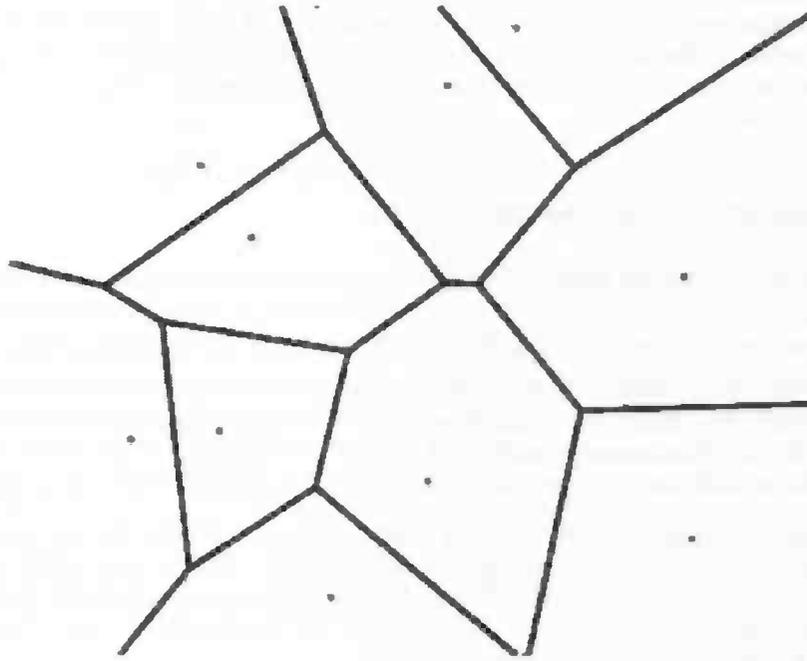


Figure 8: A Voronoi diagram in 2D space

at the half-spaces defined by each pair of sites:

$$HS(p_i, p_j) = \{x \in S \mid d(p_i, x) \leq d(p_j, x)\}$$

Using these half-spaces, the Voronoi regions are defined by:

$$V(p_i) = \bigcap_{1 \leq j \leq n, j \neq i} HS(p_i, p_j)$$

Because the Voronoi regions are intersections of half-spaces, the regions are convex polyhedra.

The Voronoi diagram is usually introduced as a tool to solve the so called "post office" problem [1]. This problem is defined as: given a set P of n sites (post offices), report the site closest to a given query point (the location of a person). This problem could be solved in $O(n)$ time by simple computing the distance to each site from the query point. However, after constructing the Voronoi diagram, algorithms exist that compute the Voronoi region of the query point in $O(\log n)$ time and thus the closest site.

Voronoi diagrams are often attributed to Dirichlet - hence the name Dirichlet is sometimes used - and Voronoi [18]. They can already be found in Descartes's treatment of cosmic fragmentation in Part III of his *Principia Philosophiae*, published in 1644. Also in the 20th century the Voronoi diagram has been re-discovered several times. In biology this even happened twice in a very short period. In 1965 Brown studied the intensity of trees in a forest. He defined the *area potentially available to a tree*, which was in fact the Voronoi cell of that tree. One year later Mead used the same concept for plants, calling the Voronoi cells plant polygons. By now there is an impressive amount of literature concerning Voronoi diagrams and their applications in all kinds of research areas. The book by Okabe et al., *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, contains an ample treatment of Voronoi diagrams and their applications.

4.2 Delaunay triangulation

A triangulation of a set of points $P \subset \mathbb{R}^d$ is a partition of the convex hull of P into d -dimensional simplices, whose vertices are the points P . The d -dimensional simplices are such that their intersection is either a i -dimensional simplex or empty, where $i < d$.

This paper will only work with three dimensional data. The two dimensional case is sometimes used for illustration purposes. In this three-dimensional case a triangulation consists of tetrahedra (3-simplices, these will also be called cells), triangles (2-simplices, these will also be called faces), edges (1-simplices) and vertices (0-simplices).

The Delaunay triangulation has the empty sphere property, that is, the circumscribing sphere of each cell of such a triangulation does not contain any other vertex of the triangulation in its interior [21]. These triangulations are uniquely defined except in degenerate cases where five or more points are cospherical. See Figure 9 for an example of a Delaunay triangulation.

The empty sphere property of the Delaunay triangulation ensures that it minimizes the maximum of the radii of the circumscribing spheres of the d -simplices of the triangulation [23]. This means that the Delaunay triangulation will be of high quality. To determine the quality of a triangulation in this sense, one usually looks at the highest aspect ratio of all d -simplices of the triangulation. The aspect ratio of a d -simplex is the radius of the circumscribing d -sphere divided by the radius of inscribed d -sphere [16]. An alternative but weaker quality measurement is to use the radius-edge ratio. It is the circumradius divided by the shortest edge length of the d -simplex.

In 2D the empty sphere property also means that the Delaunay triangulation maximizes the minimum angle in all triangles of the triangulation. However for higher dimensions, no such property can be guaranteed. In these dimensions slivers can exist. Slivers are the elements of the triangulation that have small radius-edge ratio, but large aspect ratio. To avoid slivers in higher dimensions extra points have to be

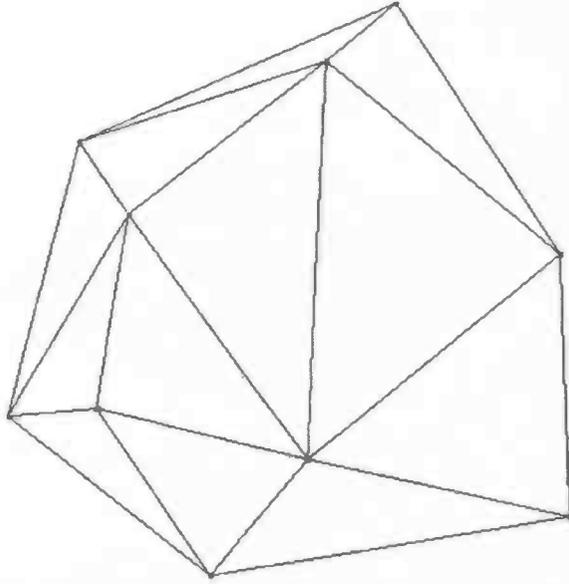


Figure 9: A Delaunay triangulation in 2D

added to the point set generating the triangulation. Many algorithms exist to do this, see for example [16].

Because of this property, the Delaunay triangulation is, for example, used for interpolation. A common example for using the Delaunay triangulation for interpolation is the terrain height interpolation problem. This problem is defined as: the height of the terrain is known at a given set of points P in \mathbb{R}^2 . Using these heights, approximate the height at a query point q . This problem can be solved by creating the Delaunay triangulation of the point set P and locating the Delaunay triangle containing the query point. The height can then be approximated by linear interpolation using barycentric coordinates. The query point q can be expressed uniquely as a linear combination of the vertices x_i ($i = 1..3$) of the triangle [14]:

$$q = \sum_{i=1}^3 \alpha_i x_i \quad \text{with the constraint} \quad \sum_{i=1}^3 \alpha_i = 1$$

Define $f(x_i)$ to be the height at the point x_i , then the linearly interpolated height will be $f(q) = \sum_{i=1}^3 \alpha_i f(x_i)$. See Figure 10 for an impression of the Delaunay triangulation.

This interpolation scheme is also used in astronomy to estimate the density throughout the whole space given a galaxy distribution. This method first estimates the density at the point set representing the distribution using the volume of all incident Delaunay tetrahedra. To estimate the density at any point, the Delaunay cell containing the point is found and linear interpolation using the density at four vertices of the cell is used.

The Delaunay triangulation is also used in finite element mesh generation as a way of yielding good meshes [23]. A good mesh is loosely defined as the one whose elements are of uniform size and shape.

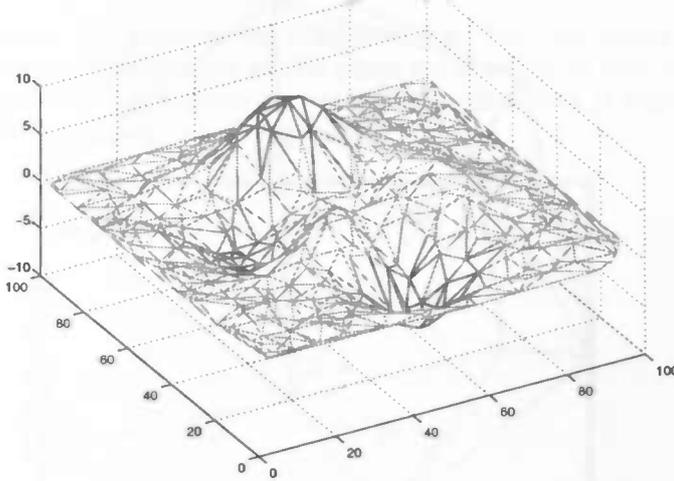


Figure 10: The Delaunay triangulation of a set of points in \mathbb{R}^2 , the height at each point is mapped to the z -axis

4.3 Duality

The Delaunay triangulation is the dual of the Voronoi diagram. They are dual in the sense that, whenever the points are in general position, the simplices of the Delaunay triangulation can be created from the Voronoi diagram. To do this take a set Q of sites, $Q \subseteq P \subset \mathbb{R}^d$. If $\bigcap_{p_i \in Q} V(p_i) \neq \emptyset$, the convex hull of the sites in Q is a simplex of the Delaunay triangulation. For the 3D case this duality means that:

- the site of a Voronoi region is a vertex in the Delaunay triangulation
- whenever Voronoi regions share a face, the corresponding Delaunay vertices have an edge between them
- whenever Voronoi regions share an edge, the Delaunay triangulation contains a face between the corresponding vertices
- whenever Voronoi regions share a point, the Delaunay triangulation contains a cell between the corresponding vertices

See Figure 11 for an example of the duality between the Voronoi diagram and the Delaunay triangulation.

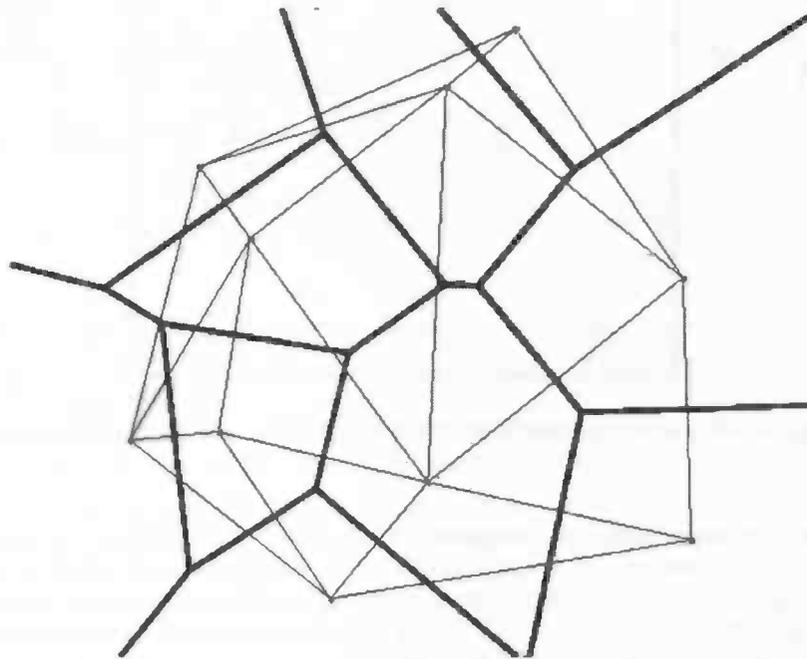


Figure 11: A Delaunay triangulation and Voronoi diagram of the same point set in 2D

5 Alpha shapes

Alpha shapes are used to describe the intuitive notion of the shape of a point set. When we look, for example, at a slice of the point set created by one of the models for the distribution of galaxies in Figure 12, we can imagine that it represents a shape similar to one of the shapes shown in Figure 13.

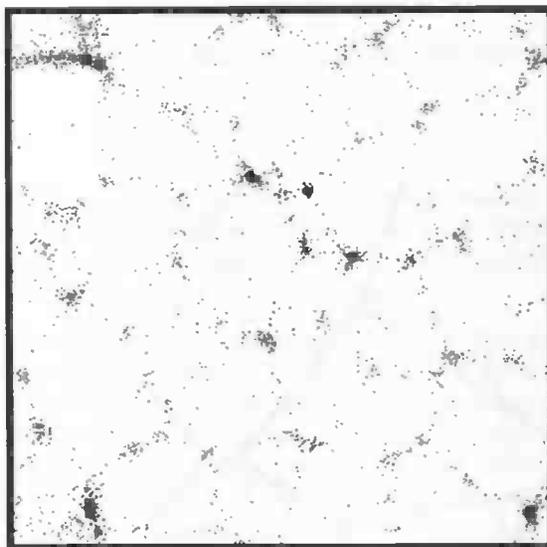


Figure 12: A z -slice of a galaxy distribution, only x - and y -coordinates are used

The alpha shape formalizes this intuitive notion, by constructing a polytope representing this shape. The definition of the alpha shape of a finite point set S is based on the Delaunay triangulation of S . We will first define the alpha complex of the set of points S . For this we will need to define the terms simplex, simplicial complex and subcomplex. A simplex σ is the convex hull of a sequence of $k + 1$ vertices, $\sigma_k = [u_0, u_1, \dots, u_k]$, where u_0, u_1, \dots, u_k are affine independent points in some fixed Euclidean space \mathbb{R}^m ($m \geq k$) and k is the dimensionality of the simplex (if k is not specified a simplex of any dimensionality is meant). In this context, σ' is a face of σ if the set of vertices of σ' is a subset of the set of vertices of σ . A simplicial complex K is a collection of simplices, with two properties [7]:

- if a simplex σ' is a face of a simplex $\sigma \in K$ then $\sigma' \in K$
- if $\sigma^1, \sigma^2 \in K$ then $\sigma^1 \cap \sigma^2$ is either empty or a face of σ^1 and σ^2

A subset $L \subseteq K$ is a subcomplex of K if it is a complex itself.

The alpha complex is a subcomplex of the Delaunay triangulation [9]. For a given non-negative value of α , the alpha complex consists of all the simplices in the Delaunay triangulation which have an empty circumsphere with squared radius less than or equal to α . Here “empty” means that the open sphere does not include any points of

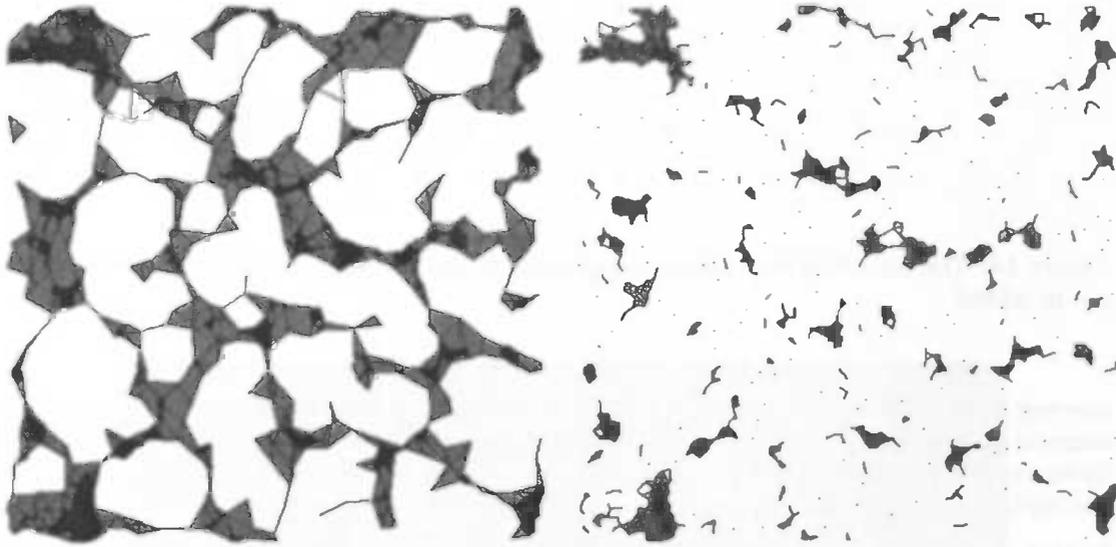


Figure 13: Two different shapes associated with the point set

S. The alpha shape is the union of all simplices of the alpha complex. This means that although the alpha shape is defined for all $0 \leq \alpha < \infty$, only a finite number of different alpha shapes exist for one point set. At the extremes $\alpha = 0$ the alpha complex is formed by just the vertices of the point set and a value α_{max} exists, such that for $\alpha \geq \alpha_{max}$ the alpha shape is just the convex hull of the point set.

From the definition of the alpha complex we can see that for each simplex σ there is a alpha value $t(\sigma)$ such that σ belongs to the alpha complex iff $t(\sigma) \leq \alpha$ [9]. We can therefore sort the Delaunay simplices as $\sigma_1, \sigma_2, \dots, \sigma_m$, such that $t(\sigma_i) < t(\sigma_j)$ implies $i < j$. Such an ordering is a filter of a simplicial complex, if $K_j = \{\sigma_1, \sigma_2, \dots, \sigma_j\}$ is a complex for $1 \leq j \leq m$. The sequence K_0, K_1, \dots, K_m is called a filtration of K . The sequence of alpha complexes for increasing α is almost a filtration of the underlying Delaunay triangulation. The only problem is that two or more simplices may appear in the alpha complex at the same α value, even if the point set is in general position. See for example Figure 14. In this figure the last edge and the face appear at the same α value. This problem can be solved by sorting simplices which appear at the same α value on their dimension, so vertices before edges, edges before faces and faces before cells with remaining ties broken arbitrarily. Sorting the simplices in this fashion will ensure that the ordering results in a filter. The advantage of the sequence of alpha complexes being a filter of the Delaunay triangulation is that it allows for a relatively easy calculation of the Betti numbers of all alpha complexes as will be shown in the next section.

The alpha shape of the point set S is often intuitively described by imagining a mass of ice-cream with a boundary equal to the convex hull of S and containing the points as "hard" chocolate pieces [4]. Using one of those sphere-shaped ice-cream spoons we carve out all parts of the ice-cream mass we can reach without touching chocolate pieces, thereby even carving out holes in the inside (parts not reachable by simply

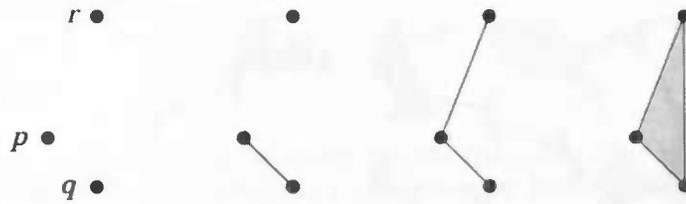


Figure 14: The four different alpha complexes, in the last step the edge qr and the face pqr is added

moving the spoon from the outside). We will eventually end up with a (not necessarily convex) object bounded by cavities, arcs and points. If we now straighten all "round" faces to triangles and line segments, we have an intuitive description of what is called the alpha shape of S .

In this description, the α value is the squared radius of the spoon. In this way, the α value controls the level of detail shown in the shape. A small α value will carve out many simplices, so we will only see small simplices, which appear in general in high density regions. Figure 13 shows the alpha shape for a large alpha value on the left and for a small alpha value on the right. In some cases the desired shape is known and good α values may be defined by looking at the difference between the alpha shape and the desired shape, see for example Figure 15, taken from [9]. However in general this is not the case and the interpretation of the best α value is left to the user.

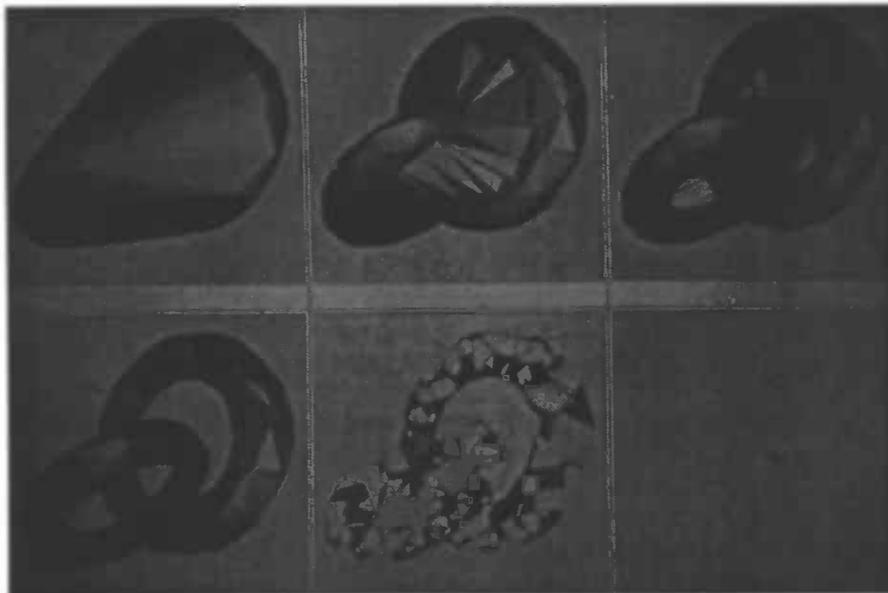


Figure 15: The points are randomly generated on the surface of two linked tori. Six different α -shapes for values of α decreasing from top to bottom and left to right are shown. The first shape is the convex hull, the last shape is the point set itself. The α -value used in the fourth frame neatly separates the two tori. Further decreasing α disassembles the shape.

6 Betti numbers

6.1 Definition

The Betti number β_k of a topological space counts the number of k -dimensional holes of that topological space. For simplicial complexes the Betti numbers can be defined using oriented k -simplices. The orientation of a k -simplex σ_k is defined by an arbitrary but fixed ordering of the vertices x_1, \dots, x_k , $\sigma_k = [x_0, x_1, \dots, x_k]$ [29] and [5]. Even permutations of this ordering give the same orientation and odd permutations reverse it. The simplicial complex, K , is given a group structure by defining the addition of k -simplices in a similar manner to addition in a vector space. The resulting group is called the chain group, C_k . Its elements consist of k -chains c_k , the sum of a finite number of oriented k -simplices: $c_k = \sum_i a^i \sigma^i$ where $a^i \in \mathbb{Q}$. Now we will take a look at how the k -chains are related to the $(k-1)$ -chains. This is done using the boundary operator, the boundary operator transforms a k -simplex in a collection of multiples of $(k-1)$ -simplices:

$$\partial_k \sigma_k = \sum_{i=0}^k (-1)^i [u_0, u_1, \dots, u'_i, \dots, u_k]$$

Where u'_i means that u'_i is omitted and $(-1)\sigma$ is σ with a reversed orientation. So, for example, a face $\sigma_2 = [u_0, u_1, u_2]$ would be transformed in a sum of edges:

$$\partial_2 \sigma_2 = [u_1, u_2] - [u_0, u_2] + [u_0, u_1]$$

We can extend the definition of the boundary operator to work on k -chains:

$$\partial_k \left(\sum_i a^i \sigma_k^i \right) = \sum_i a^i \partial_k \sigma_k^i \text{ where } a_i \in \mathbb{Q}$$

See Figure 16 for an example of the boundary operator ∂_2 . The action of the boundary

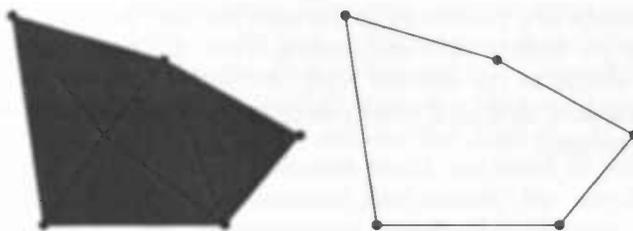


Figure 16: On the left the triangulation X and on the right $\partial_2 X$

operator on the chain groups leads to the definition of three more groups. Firstly, the image of ∂_{k+1} is a subgroup $B_k \subseteq C_k$ called the boundary group, $B_k = \{c \mid \exists z : z \in C_{k+1}, \partial_{k+1} z = c\}$. Secondly, the set of all k -chains that have empty boundary forms the group of k -cycles, Z_k , (i.e. Z_k is the kernel or null space of ∂_k), $Z_k = \{c \in C_k \mid \partial_k c = 0\}$. These two groups are related by the fact that the boundary of a boundary is empty. This is a fundamental property of the boundary operator, $\partial_k \partial_{k+1} = 0$. It implies that B_k

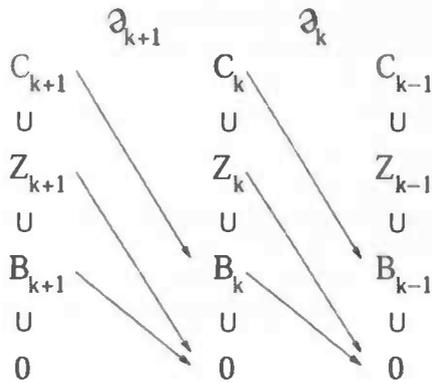


Figure 17: The mapping of k -chains and subgroups by the boundary operator

is a subgroup of Z_k . The relationships between the chain, cycle, and boundary groups are illustrated in Figure 17 from [24].

The structures we are really interested in are the k -cycles that do not bound anything, since these indicate the presence of a k -dimensional hole. Now we can define the quotient group $H_k = Z_k/B_k$, which is called the homology group. The dimension of H_k , $|H_k|$, is the k th Betti number β_k . The Betti numbers of a space, in our case defined by the union of all simplices of a simplicial complex, are independent of the actual underlying triangulation. However the underlying triangulation (which is the Delaunay triangulation for alpha shapes) does define the evolution of the shape for growing values of α and thereby influence the evolution of the Betti numbers for growing values of α .

The k th Betti number effectively counts the number of independent k -dimensional holes in (our case) the simplicial complex. The first Betti number β_0 gives the number of connected components. For subsets of \mathbb{R}^3 , we can interpret β_1 as the number of independent tunnels, and β_2 as the number of independent enclosed voids. For example, the solid torus has $\beta_0 = 1$, $\beta_1 = 1$ and $\beta_2 = 0$ (the same as a circle), whereas the hollow torus has $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$, see Figure 21 for a torus.

6.2 Algorithm

With the definition of the k -cycles we can take a look at the algorithm from [7] for calculating the Betti numbers of the alpha shapes. Defining T to be the underlying alpha complex for which we will calculate the Betti numbers, this algorithm can be summarized as shown in Algorithm 1.

The simplices on line 4 are processed in the same order as they appear in the alpha complex, when alpha is increased. Lines 6 to 10 of the algorithm mean:

- When a vertex is added to the alpha complex, a new component is created. Vertices always form a 0-cycle.

Algorithm 1 The abstract incremental method

```
1: for  $l = 0$  to 3 do
2:    $\beta_l = 0$ 
3: end for
4: for all  $\sigma \in T$  do
5:    $k = \dim \sigma$ 
6:   if  $\sigma$  belongs to a  $k$ -cycle in the subcomplex composed of all processed simplices
       then
7:      $\beta_k = \beta_k + 1$ 
8:   else
9:      $\beta_{k-1} = \beta_{k-1} - 1$ 
10:  end if
11: end for
```

- When an edge is added, the number of tunnels is increased by one if it creates a new cycle, else two components are connected and the number of connected components is decreased by one.
- When a face is added, the number of holes is increased by one if it creates a new cycle, else a tunnel is filled so β_1 has to be decreased by one.
- When a cell is added, a hole is filled up.

All that is left now is to find a way to detect if a k -simplex belongs to a k -cycle. As said, this is trivial for vertices and cells, so we need an algorithm to detect 1- and 2-cycles. Detecting 1-cycles can be done with a union-find structure [7]. A union-find data structure represents a collection of elements partitioned into a system of pairwise disjoint sets [2]. In this case the vertices are the elements and the edges connect the components. If the vertices incident to an edge belong to different components these components are joined and it is clear that the edge does not belong to a 1-cycle, else the edge belongs to a 1-cycle.

For 2-cycles we can use a similar idea. However now we need to look at a graph in which nodes represent the cells of the triangulation and arcs between the two nodes represent faces defined by the intersection of the two corresponding cells. This graph is called the dual graph, Alexander duality ensures that we can use this graph in our argumentation, see [7] for details. To ensure the dual graph represents a correct structure, meaning it is a simplicial complex itself, we need to process the simplices in reversed order (else an arc may connect two nodes, for which the cells are not present in the complex yet). We can represent this dual graph again with a union-find structure. Now the elements of the structure represent the nodes of the dual graph (which in turn represent cells in the triangulation) and an arc connects two components. When we encounter a face while processing the simplices of the triangulation in reversed order, we check whether the elements representing the cells incident to the added face, belong to different components. If they do, the components are joined and the face is part of a cycle. The interaction between the main algorithm, which processes the simplices in forward direction, and the 2-cycle detection algorithm is done by marking the faces which belong to a cycle during the 2-cycle detection algorithm. After 2-cycle detection algorithm has checked each face, the main algorithm is run. See [7] for details.

Intuitively, we can describe this algorithm for detecting 2-cycles by looking at what would happen if we process the simplices in normal order (the order in which the simplices appear in the alpha complex). Now we need the graph with nodes representing all cells of the complete triangulation. Adding a face in reversed order adds an arc to the dual graph, so processing the simplices in normal order would remove an arc from the graph. Now, if removal of an arc splits a component in the graph in two, we have found a face which is part of a 2-cycle, because it is part of a collection of faces which is the boundary of a collection of cells. However, detecting whether a component splits after removing an arc from the graph is harder than detecting merging of two components, so the algorithm described above is used.

7 Betti numbers for periodic data

7.1 2-cycle detection

The method described in Section 6.2 works for non-periodic data, but fails for periodic data. This detection fails, because with periodic data a 2-cycle does not need to separate two components in the dual graph. We can illustrate this easier by looking at the detection of 1-cycles in 2D using the dual graph. Note that we could detect 1-cycles in 2D using the same method as described in Section 6.2, we only use the dual graph to illustrate what happens with periodic data, see Figure 18.

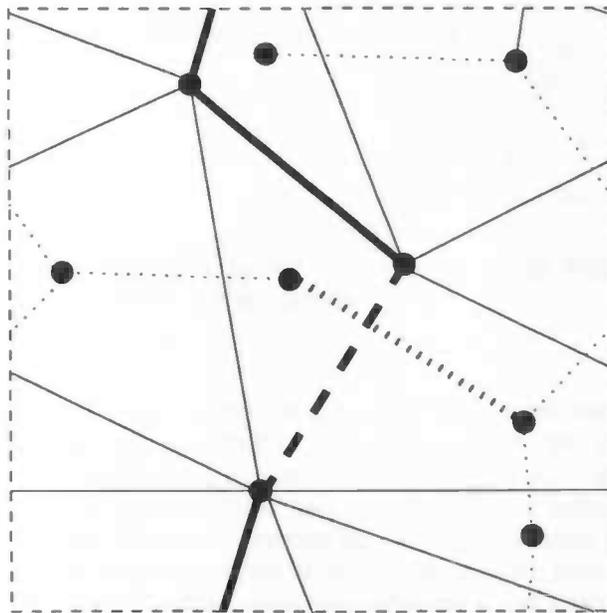


Figure 18: The bold edges form a cycle

In this figure, we see a 2D simplicial complex with its dual graph. The dashed black edge is added to the complex and will form a 1-cycle with the bold black edges, which belong to the complex when the simplices are processed in normal order. So the (blue) edges are the edges belonging to the complex, when the edges are processed in reversed order. The (red) unconnected nodes and dotted arcs form the dual graph. However the red dashed arc in the dual graph corresponding to the addition of the dashed black edge, will not merge two different components in the dual graph, since the components to the left and right of the cycle are already connected through the periodic boundary.

This problem can be solved by surrounding the triangulation with copies of itself in the positive direction along the dimensional axes. So in 2D this would mean four copies of the triangulation connected to itself.

Returning to the 3D case, we would need to connect eight copies of the triangulation. Now we can run Algorithm 1 and use almost the same detection methods. Since if

we now encounter a 2-cycle which cuts the periodic box in two, we will find an extra copy of such a 2-cycle and these two 2-cycles will cut the extended periodic box in three and it will separate two components in this extended periodic box. In the new 2-cycle detection algorithm, when a face is added to the complex, actually eight faces are added to the complex, each in its copy of the original triangulation. Note that we still need to reverse the processing order of the simplices. If the first of these additions, joins two different components in the union-find structure representing the extended triangulation, the face belongs to a 2-cycle.

7.2 Proof of correctness

To prove that the given algorithm for detecting 2-cycles is correct, we need to show that the following two implications are correct:

Consider a face f , which is added to the current complex

1. If f is an element of a cycle, then the corresponding addition of the arc in the dual graph merges two different components
2. If f causes two components in the dual graph to be merged, then f is an element of a cycle in the current complex

In other words the first implication states that all cycles are detected, the second implication states that faces which are not part of a cycle in the current complex are not detected as being part of a cycle.

Since the proof of correctness for the 3D algorithm is too complicated and time consuming to work out during this research project, we will prove the correctness of the corresponding algorithm in 2D and leave the 3D proof open.

In the 2D case we are trying to detect 1-cycles. The elements of the dual graph are the faces of the two-dimensional triangulation. As explained above we start with four connected copies of the triangulation. When we add an edge and its three copies to the current complex, we need to add all four corresponding arcs between the nodes corresponding to the adjacent faces of the edges. If the first added arc causes two components to be merged, we say that the edge belongs to a 1-cycle. Note that this method of detecting 1-cycles in 2D is only for illustration purposes, in practice the much easier method described in Section 6.2 should be used.

So for 2D, we need to prove: Consider an edge e , which is added to the current complex

1. If e is an element of a cycle, then the corresponding addition of the arc in the dual graph merges two different components
2. If e causes two components in the dual graph to be merged, then e is an element of a cycle in the current complex

To prove the first point, we will have to look at a few properties of the cycle. First if e is part of a cycle, it is also part of a cycle that does not intersect itself. Next, there are two types of cycles, ones that can be retracted to a point and ones that cannot, each type is shown in Figure 19.

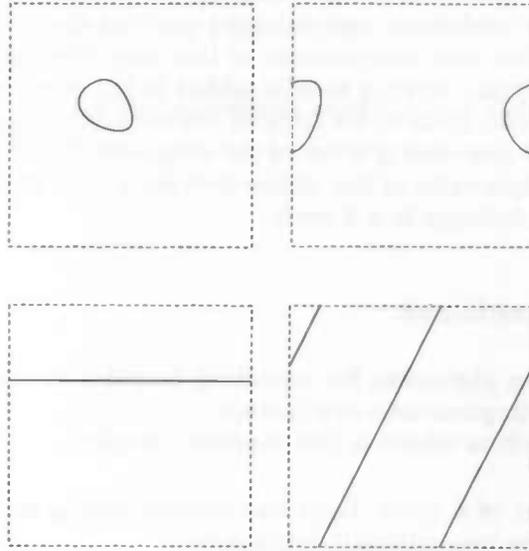


Figure 19: The two types of cycles. In the top row two cycles which retract to a point, in the bottom row two which do not retract to a point.

Suppose the cycle goes through some point $p_{0,0}$ in the periodic box, then the cycle also goes through the corresponding points $(p_{0,1}, p_{1,0}$ and $p_{1,1})$ in the copies of the box, see Figure 20. Suppose we walk along the cycle, after visiting $p_{0,0}$, we have to visit either $p_{0,0}, p_{0,1}, p_{1,0}$ or $p_{1,1}$. If the visited point has a higher x or y subscript coordinate, the next visited point will have the same change in x and y subscript coordinate. However, since we have a two by two grid of copies of the triangulation, this addition of subscript coordinates has to be done modulo two. So, whatever the point visited after $p_{0,0}$ is, the next point will be $p_{0,0}$ again. This means that in the extended periodic box we will find four (if the cycles goes from $p_{0,0}$ to $p_{0,0}$ directly) or two cycles, see Figure 20. From homology theory, we know that if we cut along a simple cycle on a 2-torus, the torus will be either cut in two or become a cylinder, see Figure 21. If the result of the cutting is a cylinder, we can cut this cylinder along the second cycle, this will split the cylinder in two. So we know that the cycles in the extended periodic box will cut the 2-torus in at least two components. The edges of these components are represented by the cycles.

If we now add the first edge e_1 of the four copies to the complex, the two components which the corresponding arc in the dual graph connects have to be different, since we process the simplices in reversed order, the arcs corresponding to the edges of the cycle are not present in the dual graph yet.

We will go on to prove the second implication, stating that if e causes two components in the dual graph to be merged, then e is an element of a cycle in the current complex. Take one of those components, suppose C is the sum of all the faces corresponding to the elements of this component. Each element of $\partial_2 C$ corresponds to an arc which has not been added to the graph yet. These in turn correspond to edges which are

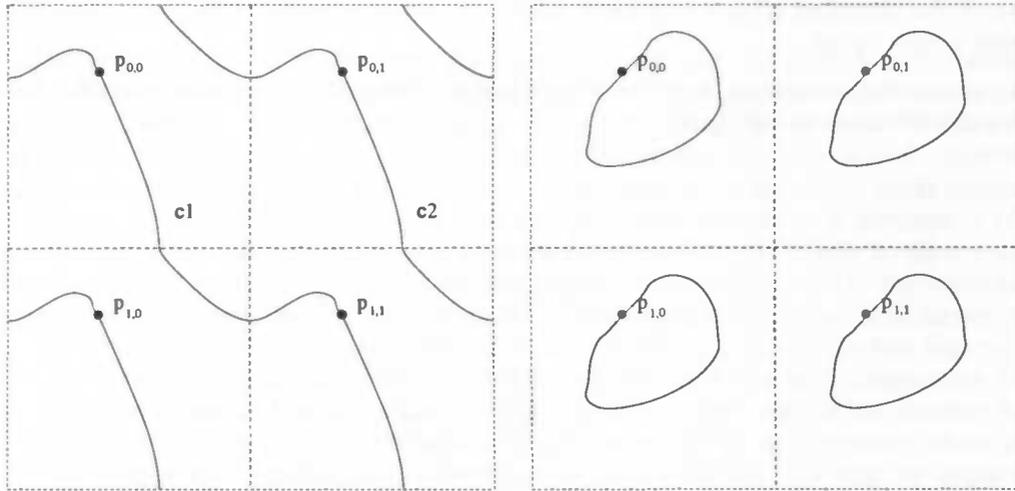


Figure 20: A cycle will form two or four cycles in the extended periodic box

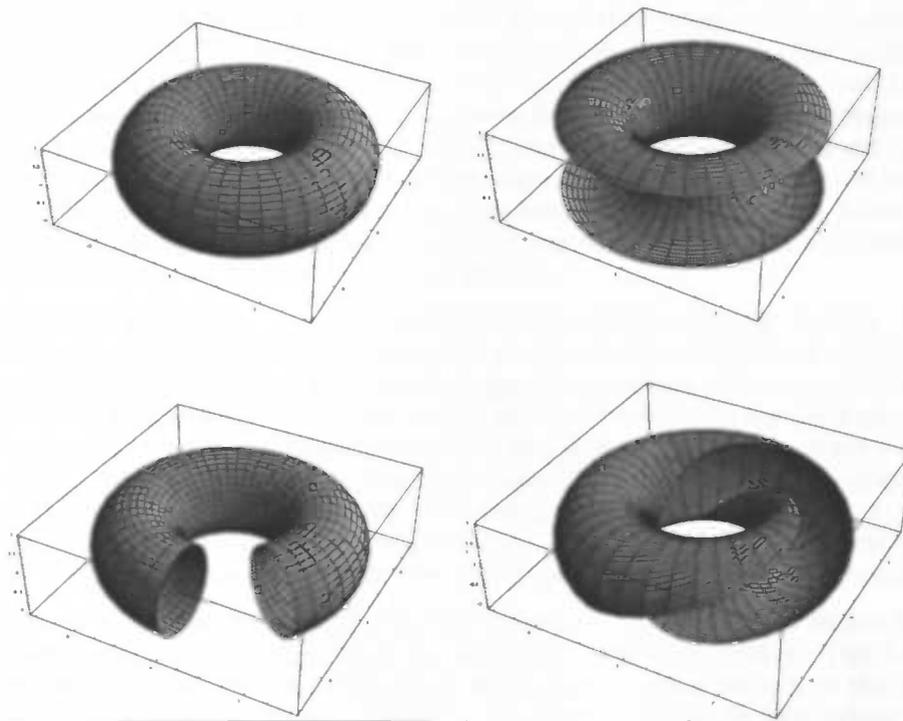


Figure 21: A 2-torus cut by 3 homologically different cycles

elements of the complex. Clearly e is an element of $\partial_2 C$, so all elements of $\partial_2 C$ will be edges of the complex after e has been added. $\partial_1 \partial_2 C = 0$ so $\partial_2 C$ is a cycle. So adding e creates a new cycle.

This proves the correctness of the algorithm working in a 2D environment. As said before the 3D case is left open.

The data we will use our algorithms on have been created by a computer model for the evolution of the distribution of galaxies, based on the Voronoi diagram (also called Voronoi tessellation) as described in [31]. To understand the rationale behind this model we need to focus on the evolution of the voids within the cosmic web, instead of the traditional focus on where the mass has accumulated. Inspired by early computer calculations, Icke (1984) pointed out that for the understanding of formation of the large coherent patterns pervading the Universe it may be worthwhile to direct attention to the complementary evolution of the underdense regions [31]. By contrast to the overdense features, the low-density regions start to take up larger and larger parts of the volume of the universe (see Figure 7). It will be as if matter in the intervening high-density domains will gradually be swept up in the wall-like and filamentary interstices, yielding a natural explanation for the resulting coherence of the cosmic foam. In realistic circumstances, expanding voids will sooner or later encounter their peers or run into dense surroundings. The volume of space available to a void for expansion is therefore restricted. For the purpose of the geometric viewpoint, the crucial question is whether it is possible to identify some characteristic and simplifying elements within such a complex. Indeed simulations of void evolution represent a suggestive illustration of a hierarchical process akin to the void hierarchy seen in realistic simulations, where the evolution of the galaxy distribution is purely governed by physical equations. The simulations of void evolution show the maturing of small-scale voids until their boundaries would intersect, after which they merge and dissolve into a larger embedding void. This process gets continuously repeated as the larger parent voids in turn dissolve into yet larger voids. A detailed assessment of the void hierarchy as it evolves from a primordial Gaussian density field (Sheth & Van de Weygaert 2000) suggests the gradual disappearance of small voids as they merge and get absorbed into the encompassing underdensities, while colossal and large voids would be rare by virtue of the fluctuation field statistics, the mainstay of voids would have sizes within a rather restricted range. Corresponding calculations then yield a void size distribution (broadly) peaked around a characteristic void size.

A bold leap then brings us to a geometrically interesting situation. Taking the voids as the dominant dynamical component of the universe we may think of the large scale structure as a close packing of spherically expanding regions. Then, approximating a void distribution, where the distribution of the scales of the voids are peaked around a characteristic value, by a void distribution of a single scale, we end up with a situation in which the matter distribution in the large scale universe is set up by matter being swept up in the in the sections of planes having equal distance to exactly two void centers and being further away from the rest of the void centers. This description of the cosmic clustering process leads to the model based on the Voronoi tessellation.

This model, creating the galaxy distributions used for this article, is called the kinematic Voronoi model, defined by Van de Weygaert and Icke (1989). The kinematic Voronoi model is based on the notion that when matter streams out of the voids towards the Voronoi skeleton, defined by the void centers, cell walls form when material from one void encounters that from an adjacent one [31]. The structure formation scenario of the kinematic model proceeds as follows (see Figure 22 for a schematic sketch of the various stages in the model). Within a void, the mean distance between galaxies increases uniformly in the course of time. When a galaxy tries to enter an ad-

adjacent cell, the gravity of the wall will slow its motion down. On average, this amounts to the disappearance of its velocity component perpendicular to the cell wall (which corresponds to the Voronoi face). Thereafter the galaxy continues to move within that wall, until it tries to enter the next cell; it loses its velocity component towards that cell, so that the galaxy continues along a filament (corresponding to a Voronoi edge). Finally, it comes to rest in a node, as soon as it tries to enter a third neighbouring void.

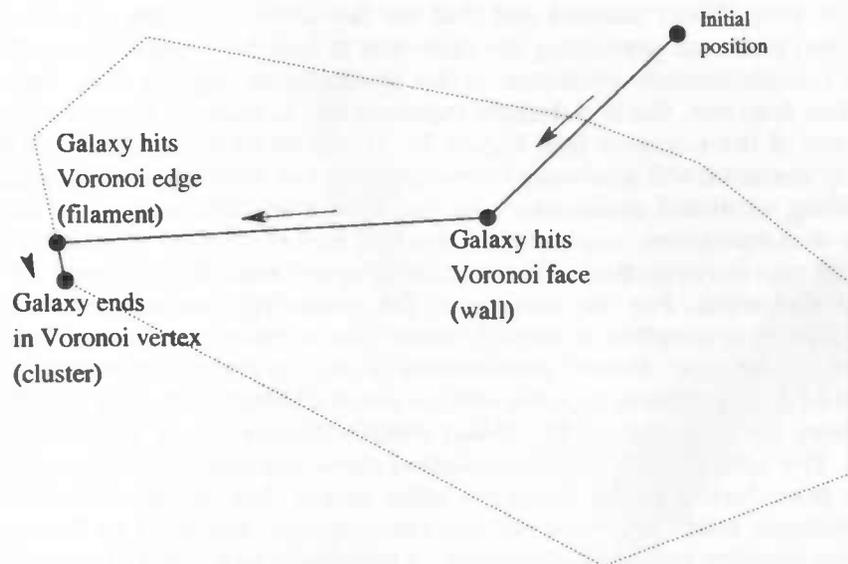


Figure 22: A galaxy moving within a Voronoi cell, face and edge

Using this model, twelve galaxy distribution data sets, six pairs, have been created. The model needs a number of input parameters:

- The number of void centers $\#V$
- The number of particles $\#\text{particles}$
- The filling factor ff , the fraction of the particles in the interior of the cell, so not in a wall, edge or vertex
- The expansion factor ef , the relative movement of the particles seen from the Voronoi nucleus, representing a void
- The initial distribution percentage of the particles over the the interior of cells, walls, edges and vertices.

The following sets of parameters have been used to create distributions (shown in Figure 23)

The first four distribution data sets (vorwall.1, vorwall.2, vorfil.1 and vorfil.2) are created by simply placing the particles in the walls and filaments respectively, for these

Name	#V	#particles	<i>ff</i>	<i>ef</i>	cell	wall	edge	vertex
vorwall.1	8	200000			0.00	100.00	0.00	0.00
vorwall.2	64	200000			0.00	100.00	0.00	0.00
vorfil.1	8	200000			0.00	0.00	100.00	0.00
vorfil.2	64	200000			0.00	0.00	100.00	0.00
vorkinm.1.1	8	200000	0.50	1.2599	50.92	36.46	11.32	1.30
vorkinm.1.2	64	200000	0.50	1.2599	49.93	38.52	10.46	1.08
vorkinm.2.1	8	200000	0.20	1.7100	21.83	36.27	32.38	9.53
vorkinm.2.2	64	200000	0.20	1.7100	20.04	40.52	30.27	9.17
vorkinm.3.1	8	200000	0.05	2.7144	5.06	23.98	43.35	27.62
vorkinm.3.2	64	200000	0.05	2.7144	5.03	23.50	41.26	30.22
vorkinm.4.1	8	200000	0.02	3.6840	2.17	14.09	40.79	42.94
vorkinm.4.2	64	200000	0.02	3.6840	2.00	14.72	39.81	43.47

Figure 23: Parameters used to create the galaxy distribution data sets

data sets the particles do not move like described above. This will create very “extreme” data sets in the sense that *all* particles are in the wall or filaments, which will not happen with the kinematic data sets

The kinematic sets (vorkinm.1.1, vorkinm.1.2, vorkinm.2.1, vorkinm.2.2, vorkinm.3.1, vorkinm.3.2, vorkinm.4.1 and vorkinm.4.2) on the other hand are created by starting with a box full of randomly placed particles. These particles then *do* move by the procedure described above. These data sets represent a more and more evolved structure. As we can see in Figure 23 each distribution has two versions, one with eight void centers and one with 64 void centers. Since these two versions represent the same underlying structure, we use these to see how our quantification methods respond to the same structure in different scales.

The next few pages shows slices of the data sets described in Figure 23. The particles of all distributions have coordinates between zero and one. To create these slices, the particles with *z*-coordinate between 0.45 and 0.55 are plotted.

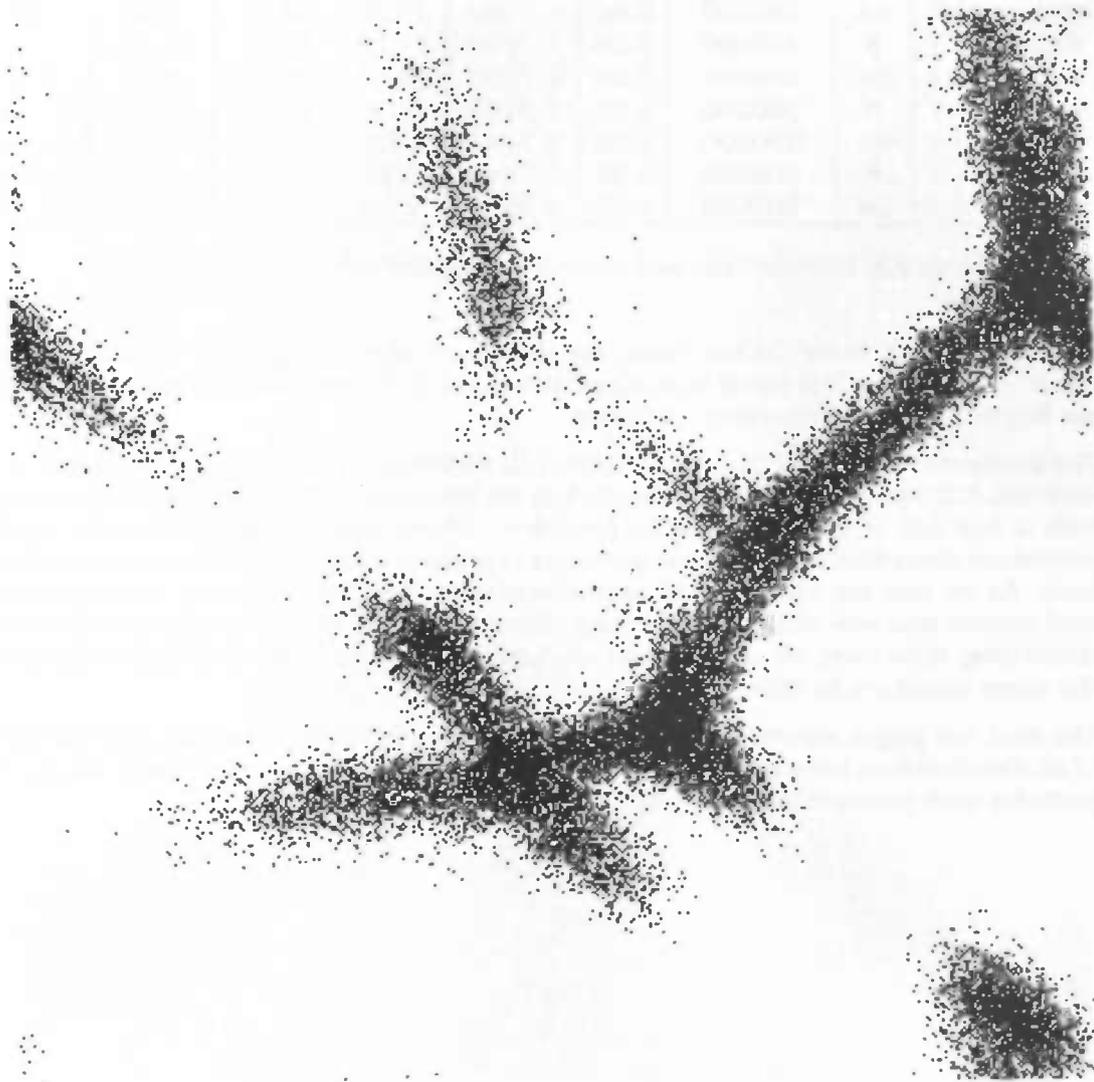


Figure 24: A slice of the distribution vorfil.1

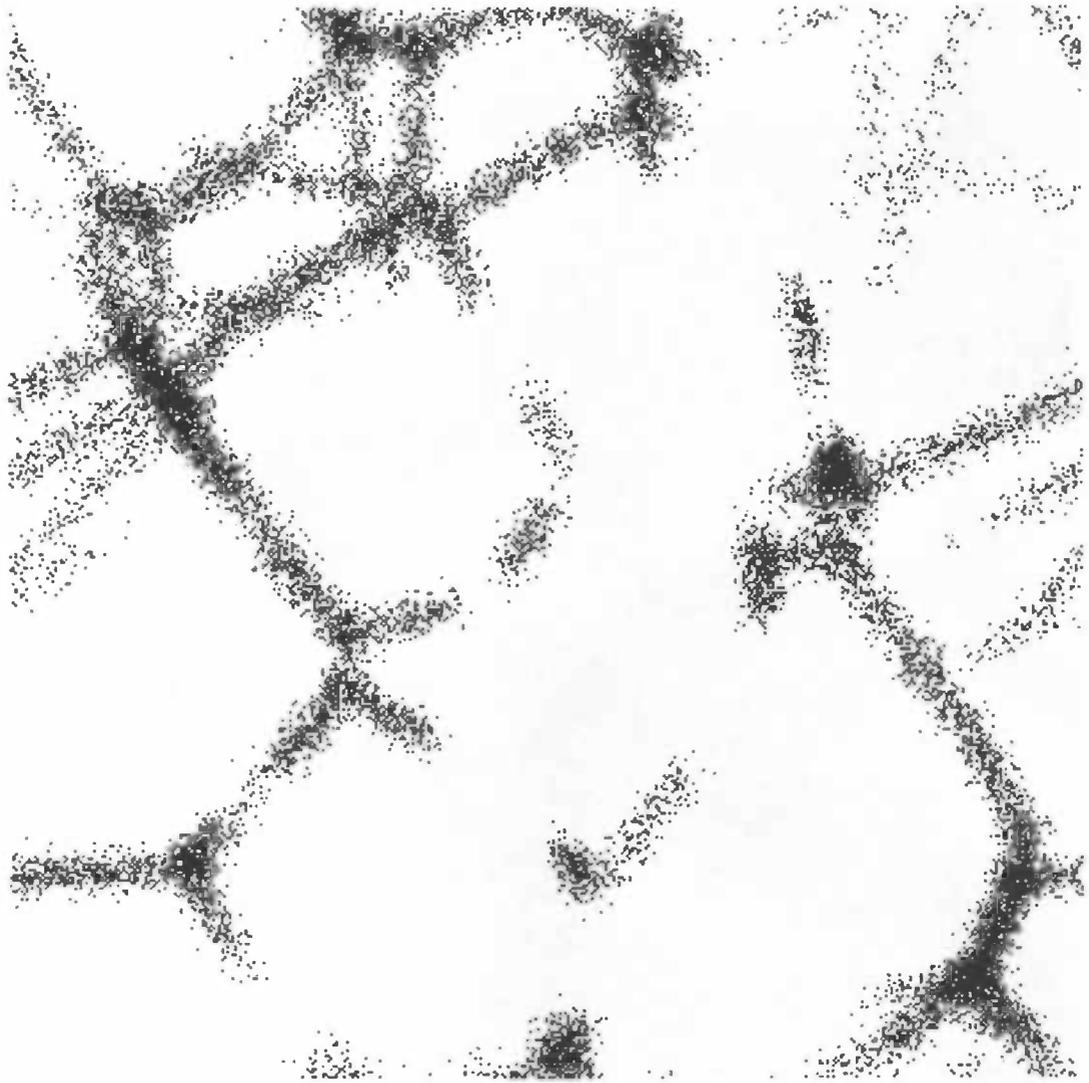


Figure 25: A slice of the distribution vorfl.2

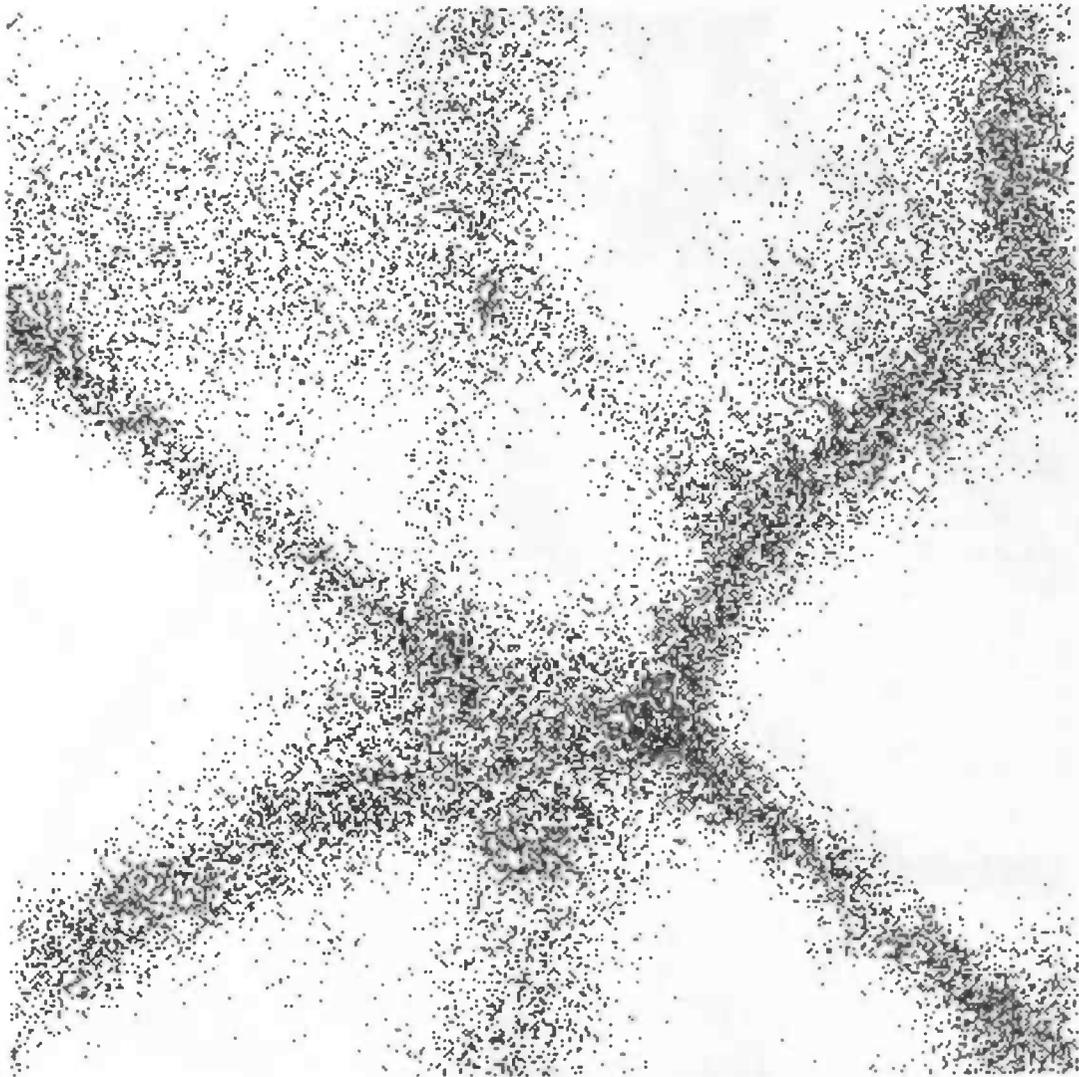


Figure 26: A slice of the distribution vorwall. 1

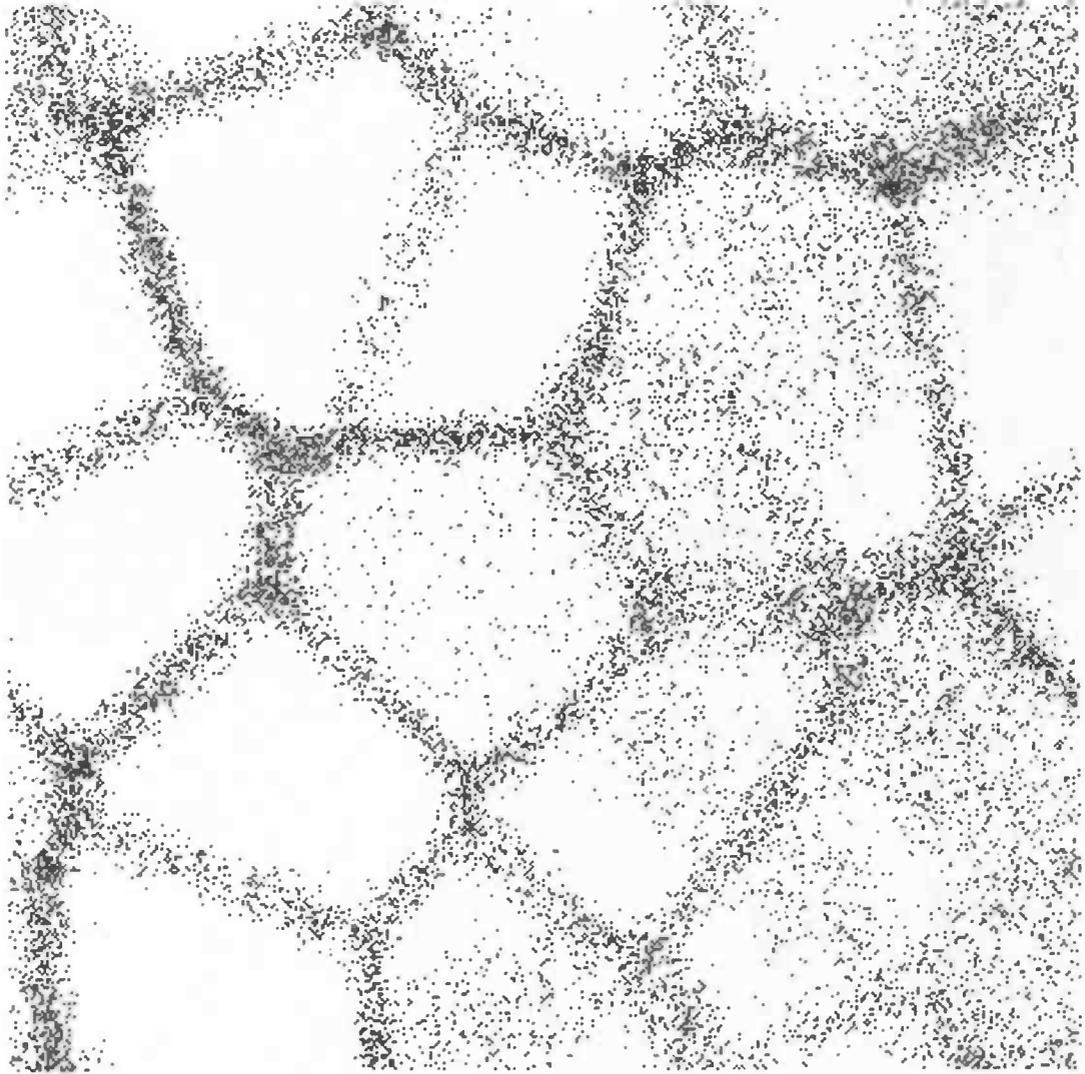


Figure 27: A slice of the distribution vorwall.2

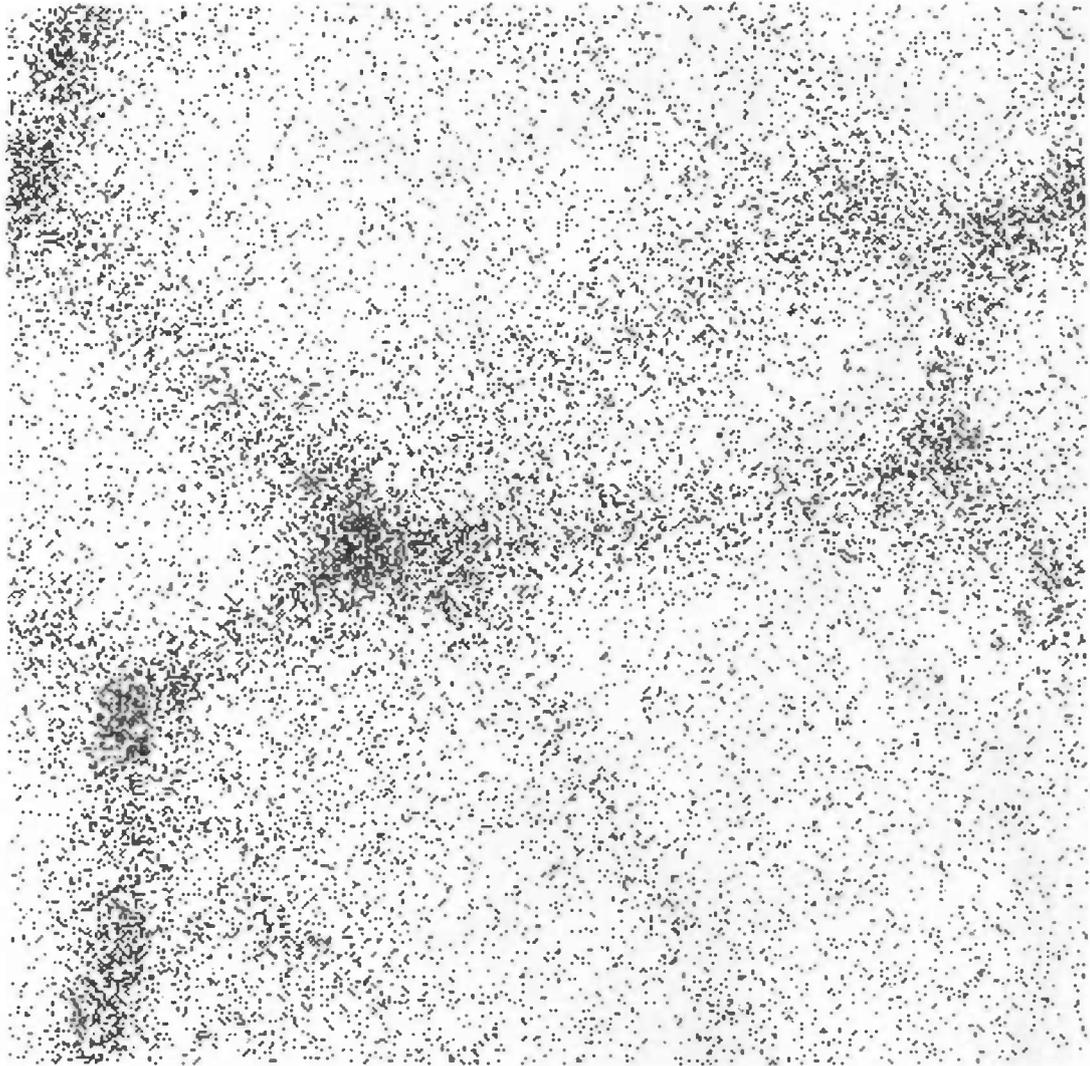


Figure 28: A slice of the distribution $vorkinm. 1. 1$

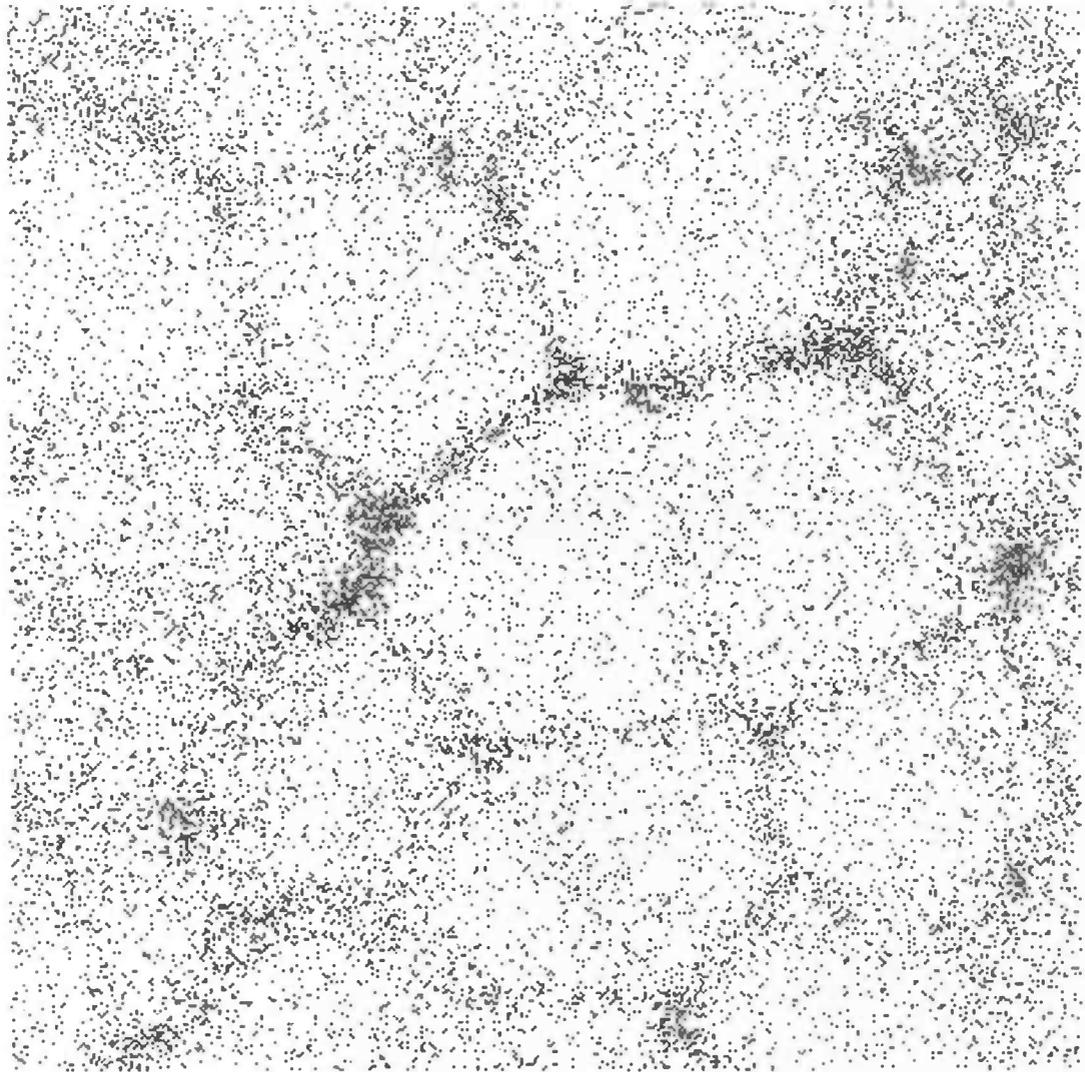


Figure 29: A slice of the distribution $v_{\text{okinm.1.2}}$

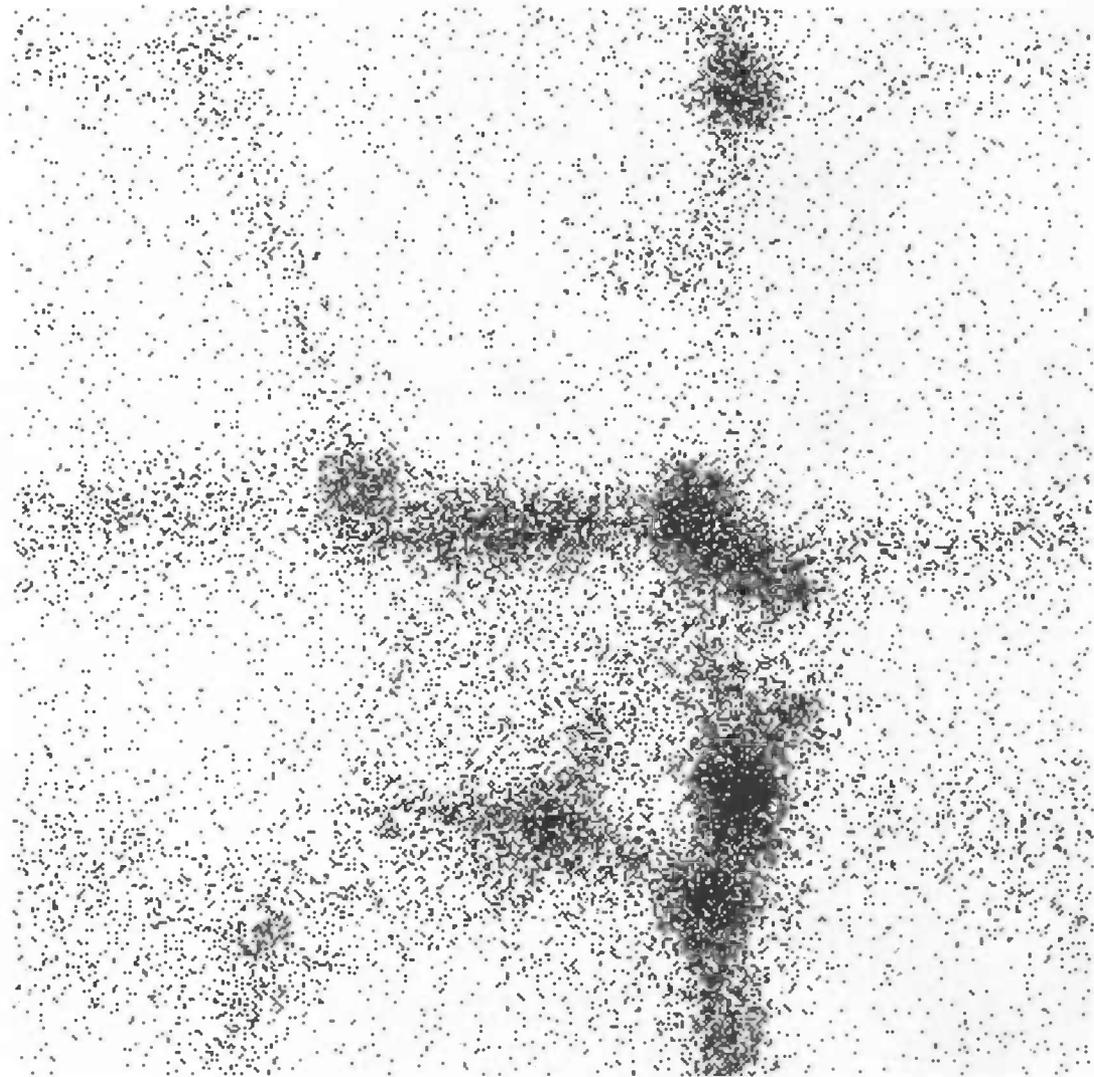


Figure 30: A slice of the distribution vorkinm.2.1

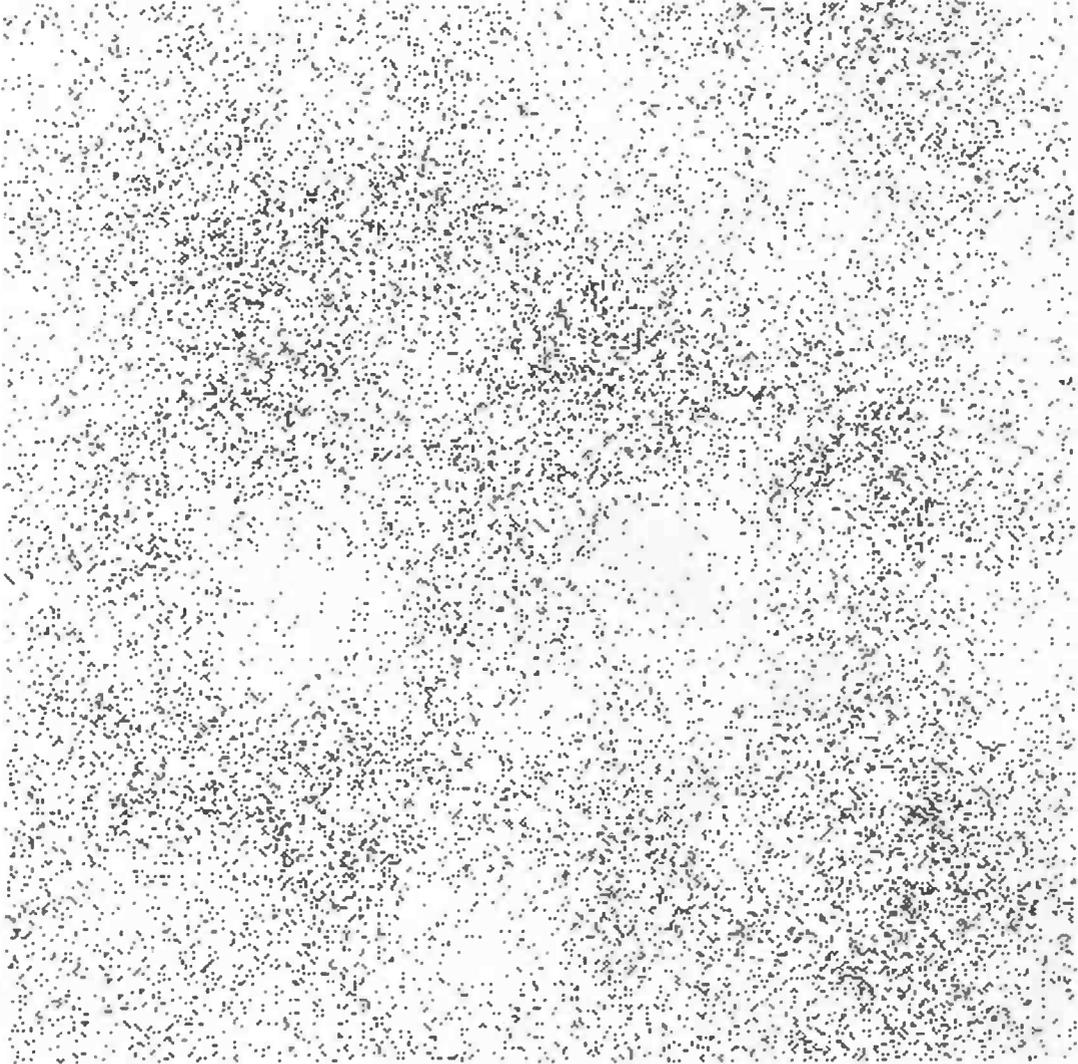


Figure 31: A slice of the distribution vorkinm.2.2

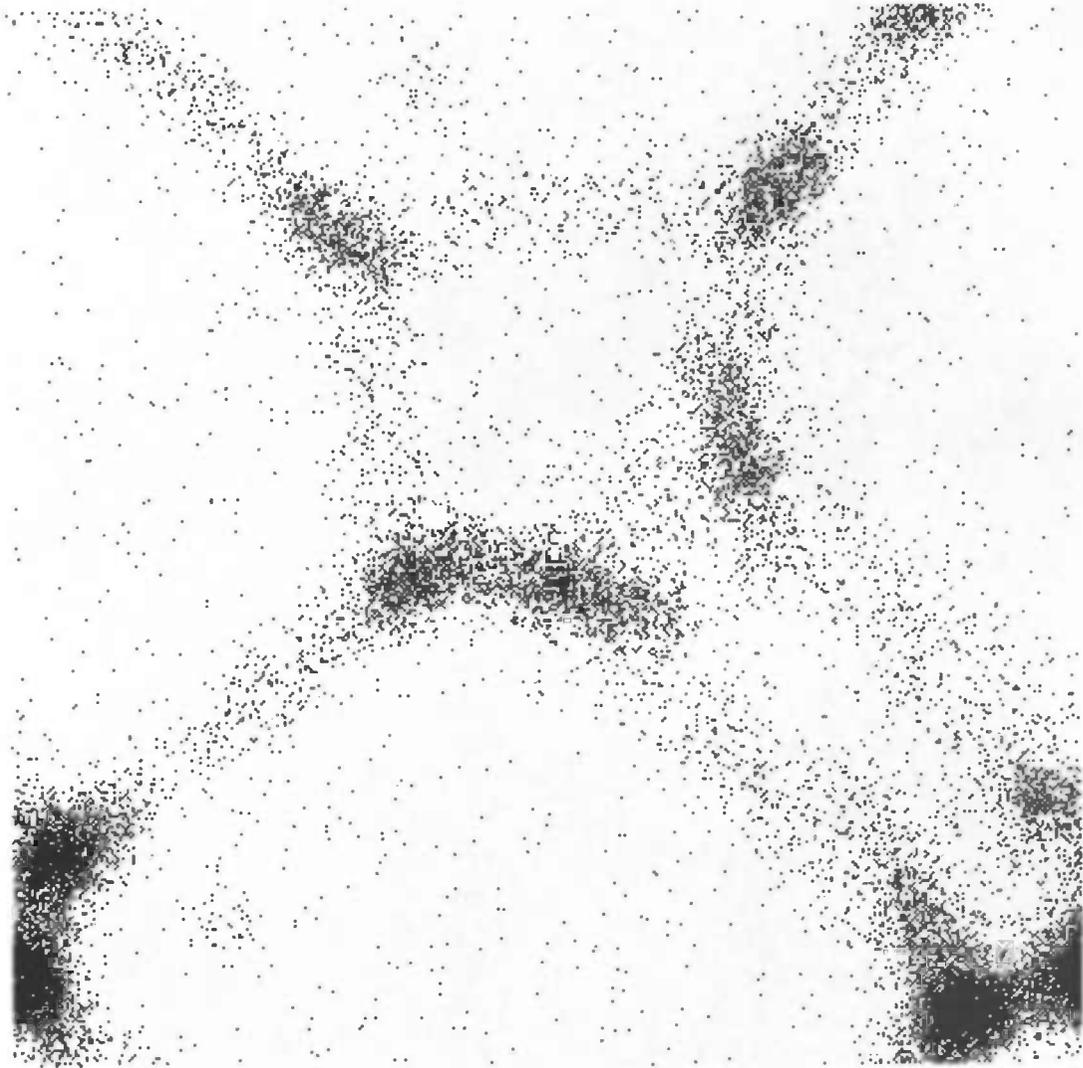


Figure 32: A slice of the distribution vorkinm.3.1

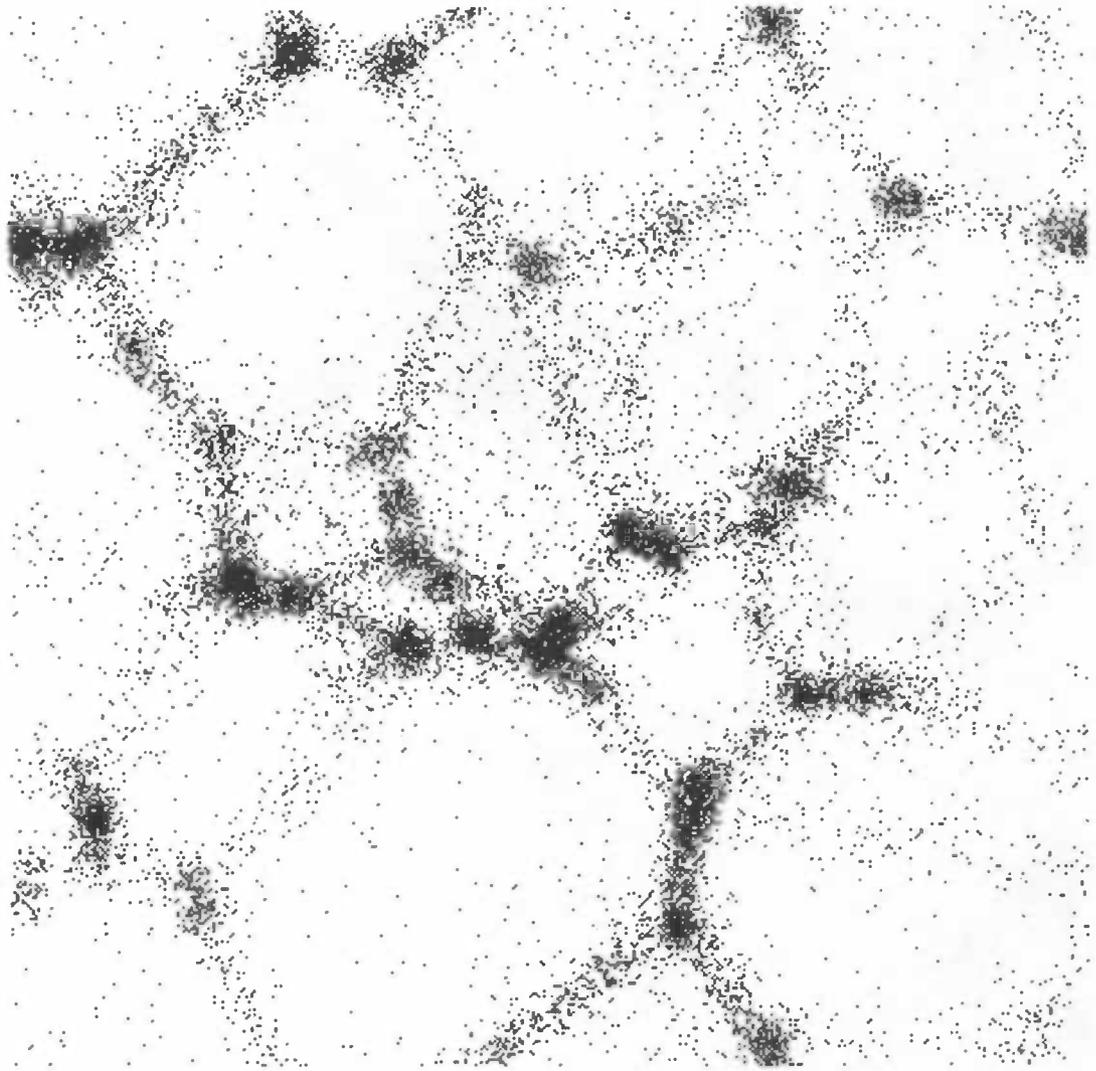


Figure 33: A slice of the distribution `vorkinm.3.2`

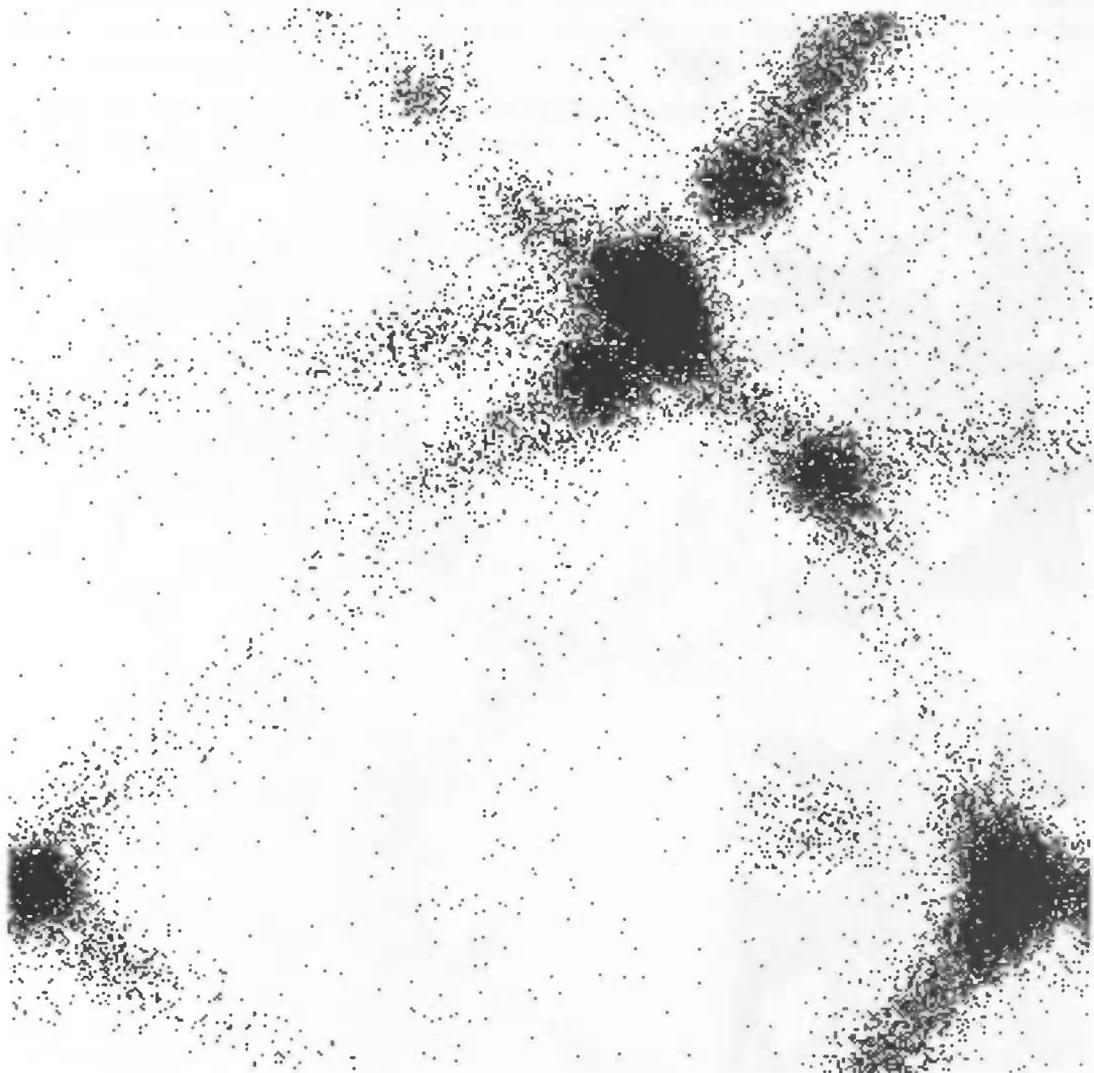


Figure 34: A slice of the distribution vorkinm.4.1

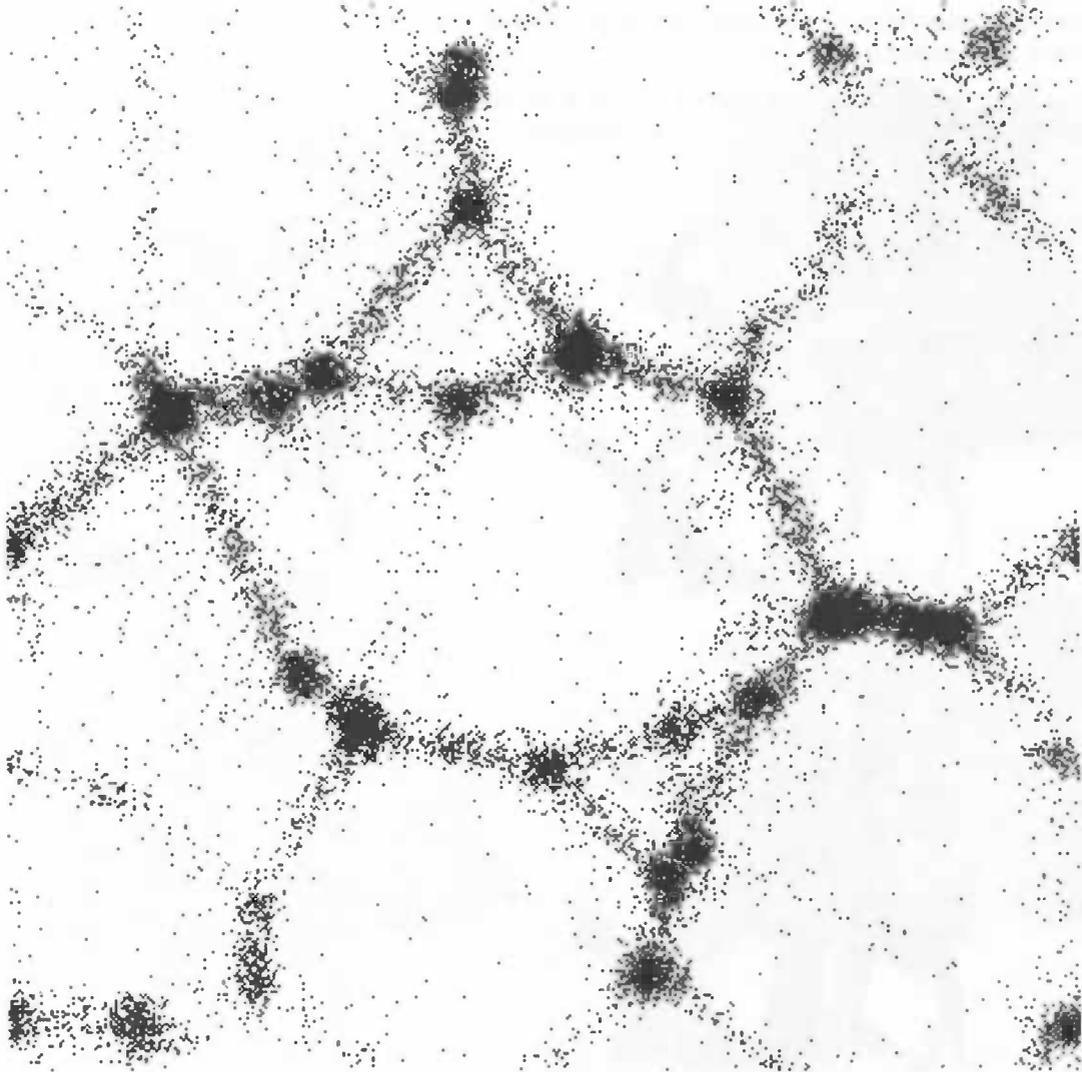


Figure 35: A slice of the distribution `vorkinm.4.2`

9 Results

In this section we will discuss the results of the calculations of the Betti numbers of the alpha shapes of the data sets discussed in the previous section. As mentioned in the introduction we will also present the volume and surface area of the alpha shapes. All graphs produced can be found in the appendix, Section A. Note that all galaxy distributions have been simulated in a periodic unit cube. Most graphs in this section show a selected interval of α values.

But first we will present figures of some alpha shapes of the galaxy distributions for selected value of α to give a visual reference.

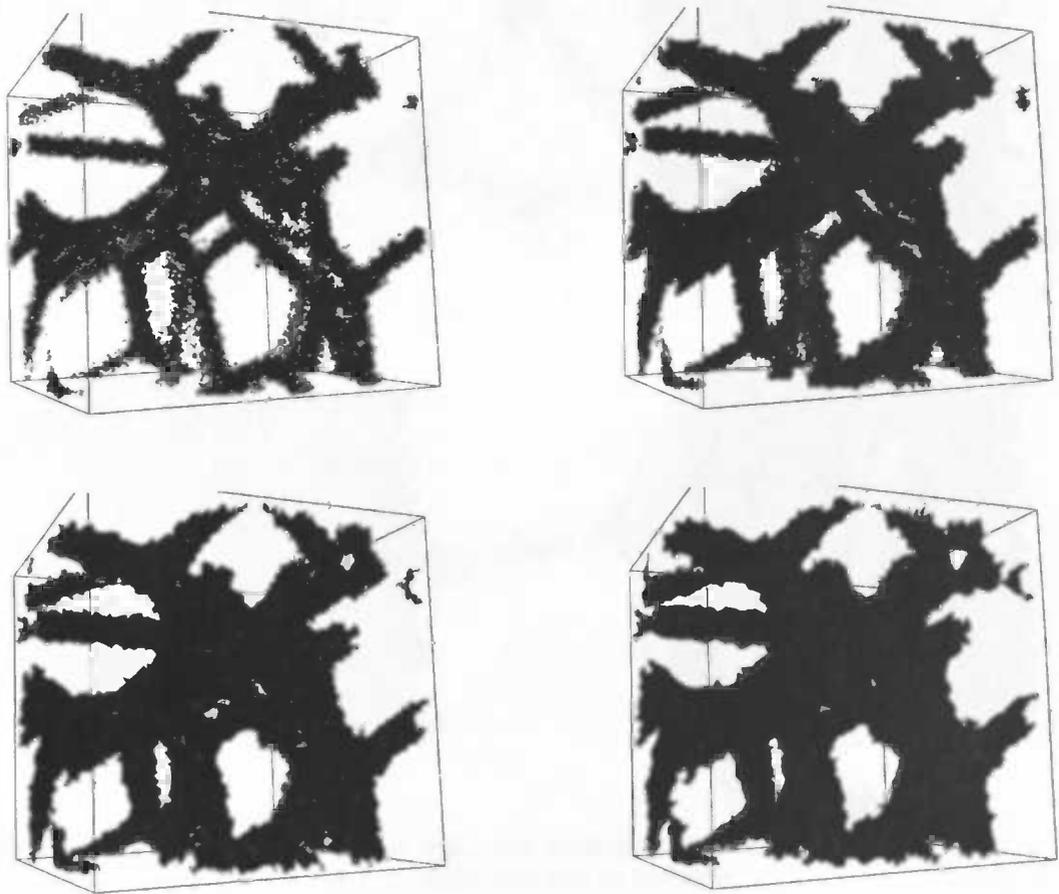


Figure 36: Four alpha shapes of the data set vorfil.1, for α values $0.5 \cdot 10^{-4}$, $1.0 \cdot 10^{-4}$, $2.0 \cdot 10^{-4}$ and $4.0 \cdot 10^{-4}$ (left to right and top to bottom)

In these figures only the alpha cells are shown, to avoid clutter. The different colors represent connected components of cells.

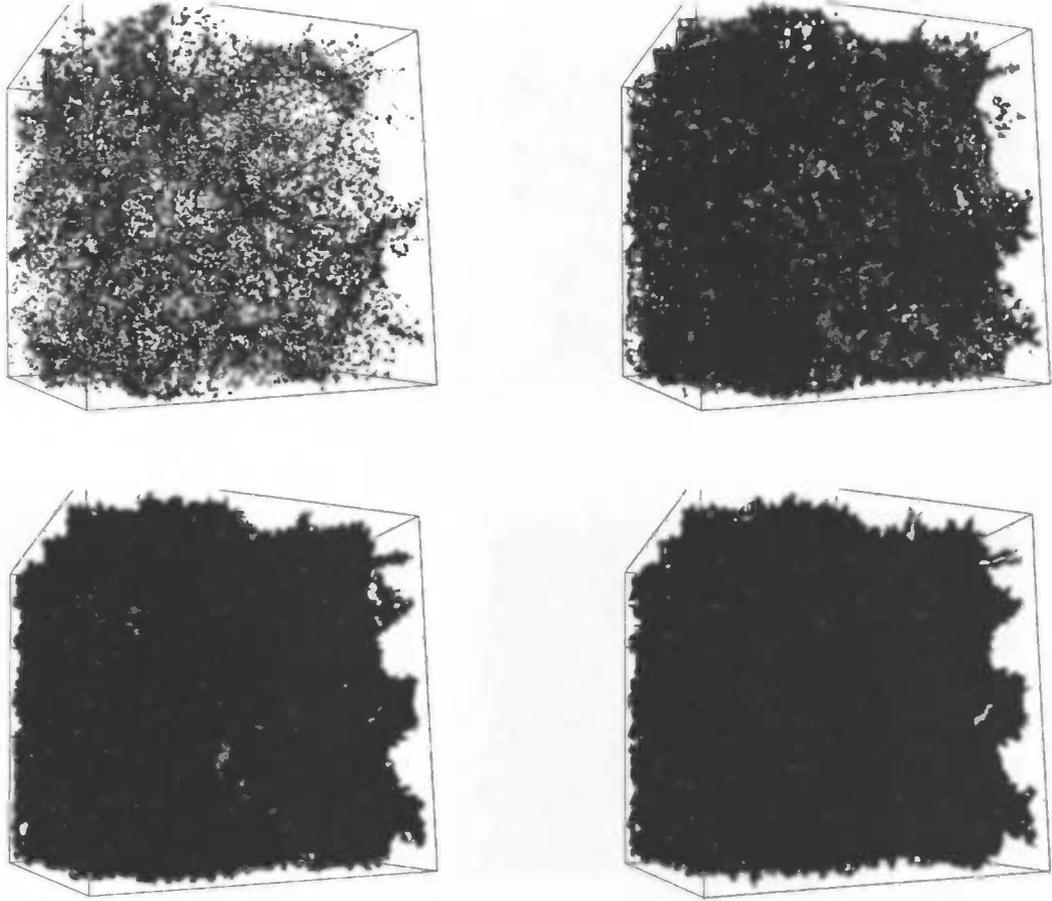


Figure 37: Four alpha shapes of the data set vorwall.2, for α values $0.5 \cdot 10^{-4}$, $1.0 \cdot 10^{-4}$, $2.0 \cdot 10^{-4}$ and $4.0 \cdot 10^{-4}$ (left to right and top to bottom)

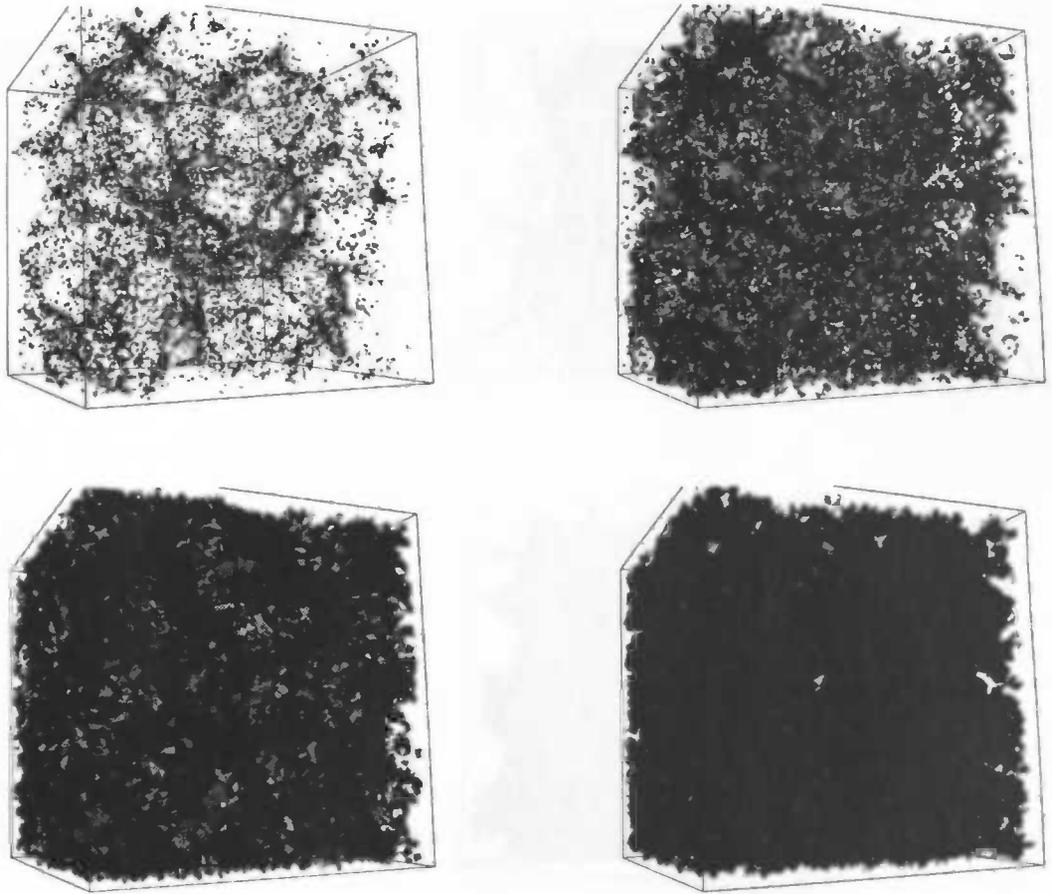


Figure 38: Four alpha shapes of the data set vorkinm.1.2, for α values $0.5 \cdot 10^{-4}$, $1.0 \cdot 10^{-4}$, $2.0 \cdot 10^{-4}$ and $4.0 \cdot 10^{-4}$ (left to right and top to bottom)

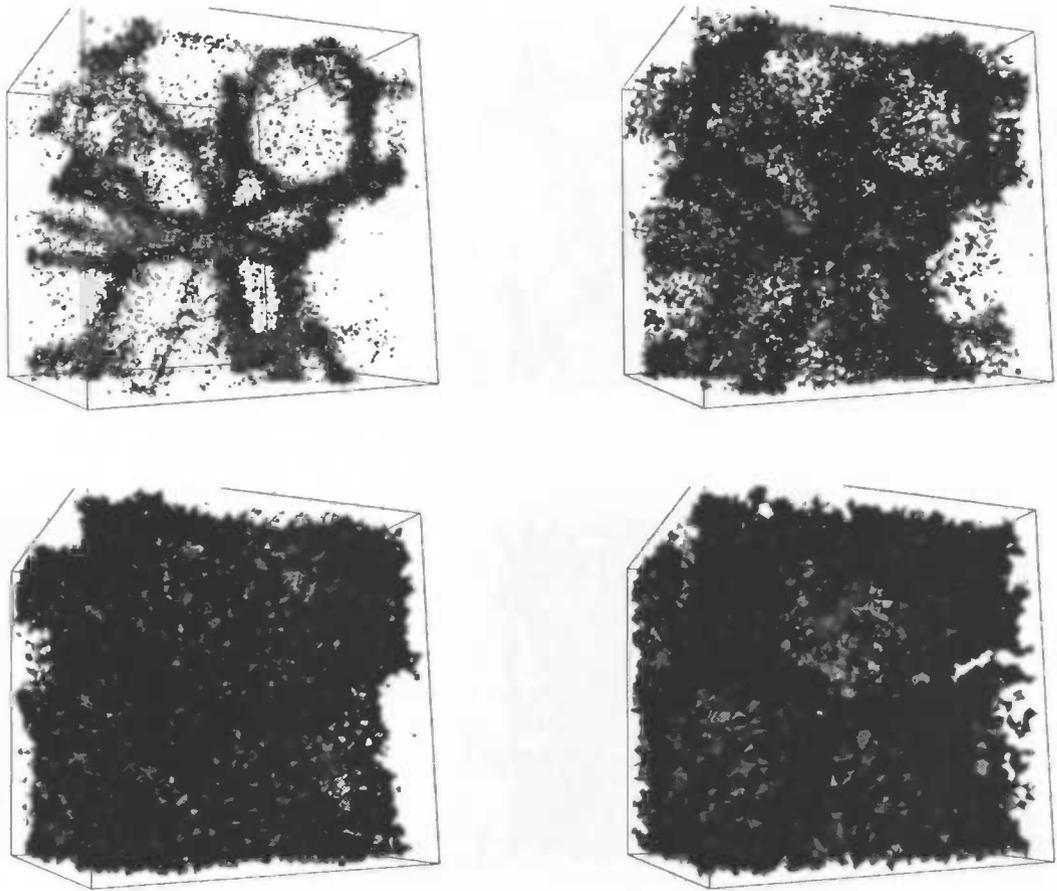


Figure 39: Four alpha shapes of the data set `vorkinm.2.1`, for α values $0.5 \cdot 10^{-4}$, $1.0 \cdot 10^{-4}$, $2.0 \cdot 10^{-4}$ and $4.0 \cdot 10^{-4}$ (left to right and top to bottom)

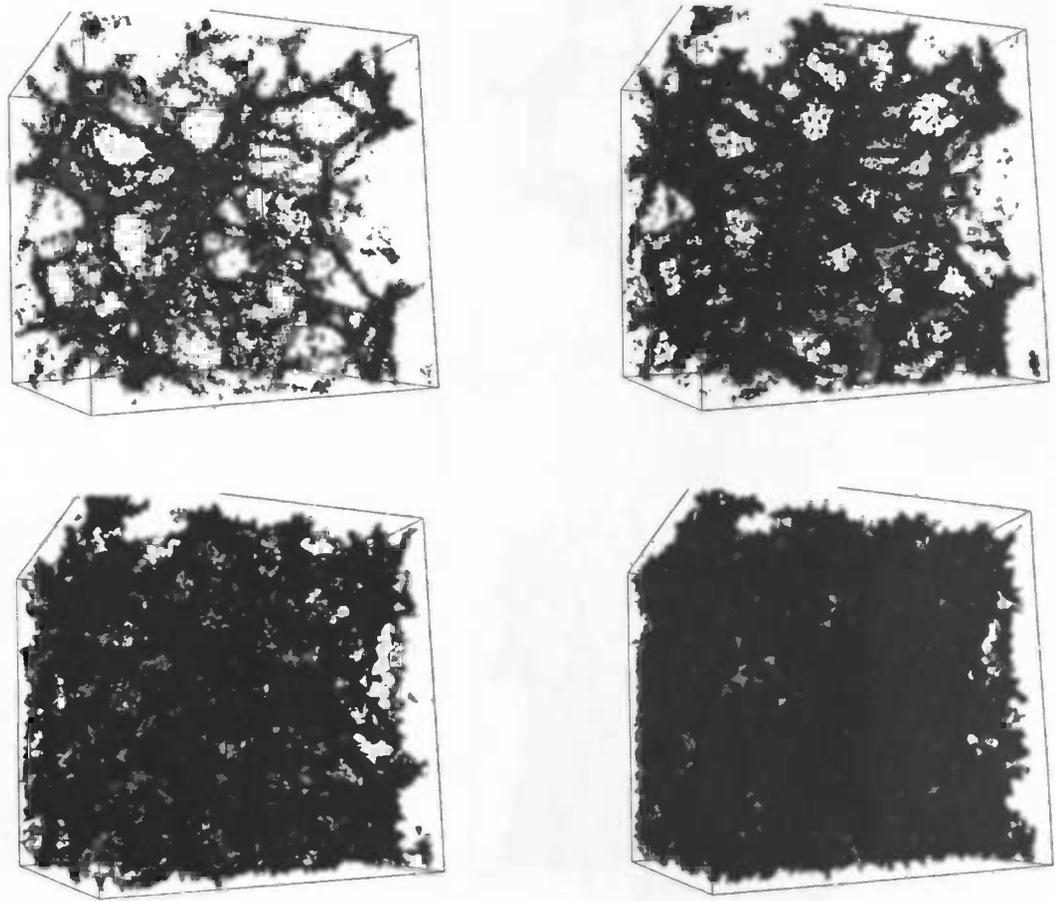


Figure 40: Four alpha shapes of the data set `vorkinm.3.2`, for α values $0.5 \cdot 10^{-4}$, $1.0 \cdot 10^{-4}$, $2.0 \cdot 10^{-4}$ and $4.0 \cdot 10^{-4}$ (left to right and top to bottom)

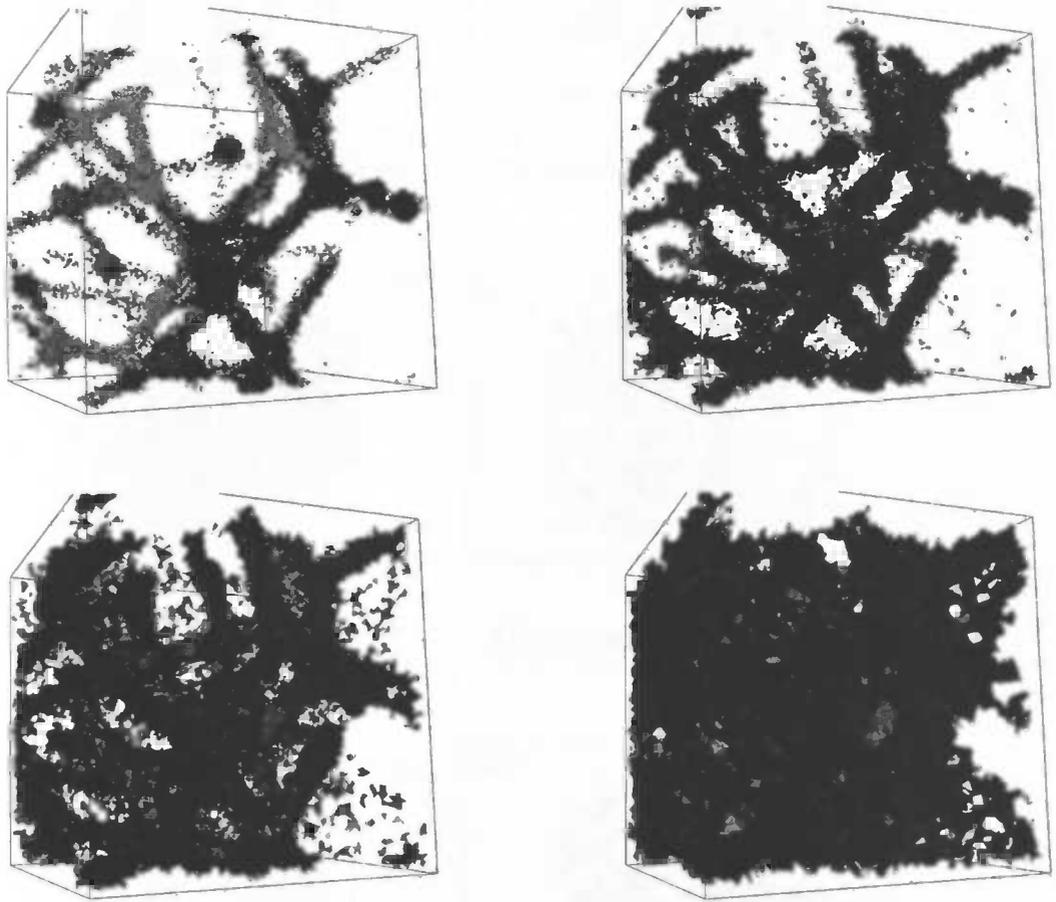
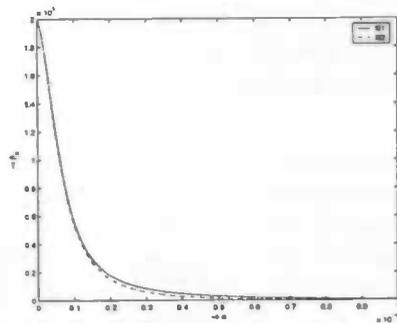
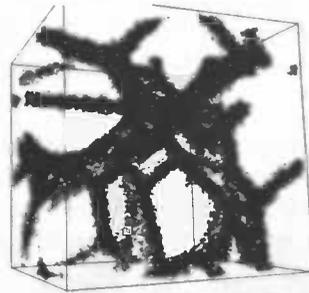


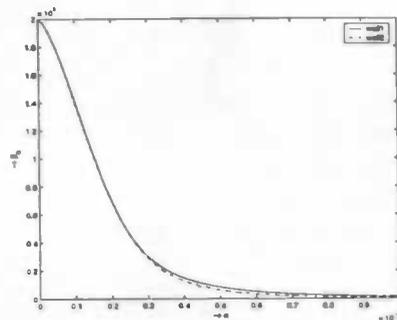
Figure 41: Four alpha shapes of the data set vorkinm.4.2, for α values $0.5 \cdot 10^{-4}$, $1.0 \cdot 10^{-4}$, $2.0 \cdot 10^{-4}$ and $4.0 \cdot 10^{-4}$ (left to right and top to bottom)



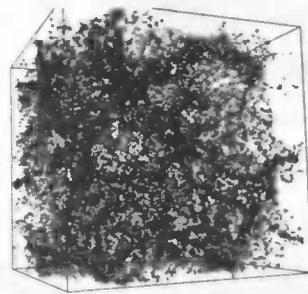
(a) The filament datasets (vorfil.1 and vorfil.2)



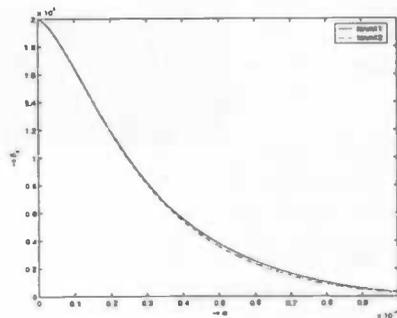
(b) And an example alpha shape (vorfil.1 at $\alpha = 0.5 \cdot 10^{-4}$)



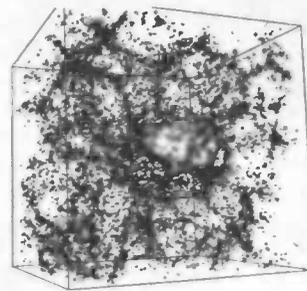
(c) The wall datasets (vorwall.1 and vorwall.2)



(d) And an example alpha shape (vorwall.2 at $\alpha = 0.5 \cdot 10^{-4}$)

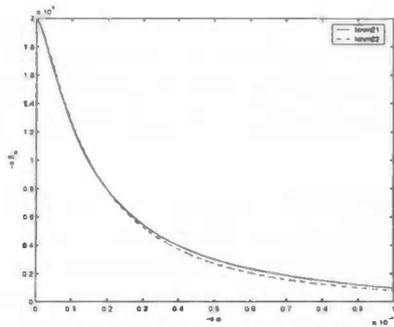


(e) The first pair of kinematic datasets (vorkinm.1.1 and vorkinm.1.2)

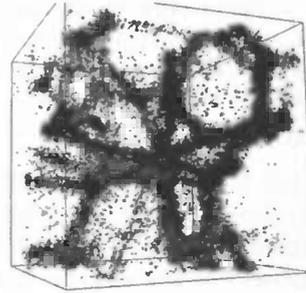


(f) And an example alpha shape (vorkinm.1.2 at $\alpha = 0.5 \cdot 10^{-4}$)

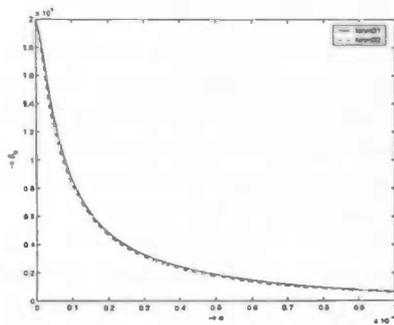
Figure 42: Figure continued at next page



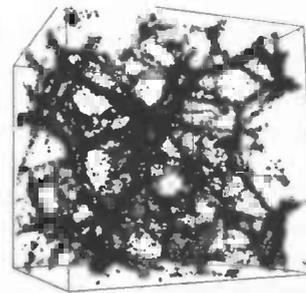
(a) The second pair of kinematic datasets (vorkinm.2.1 and vorkinm.2.2)



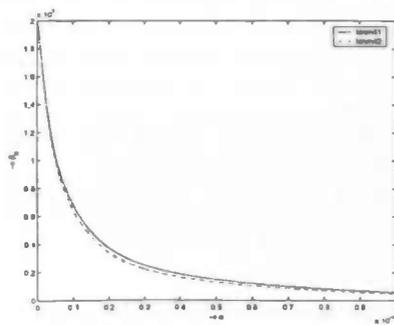
(b) And an example alpha shape (vorkinm.2.1 at $\alpha = 0.5 \cdot 10^{-4}$)



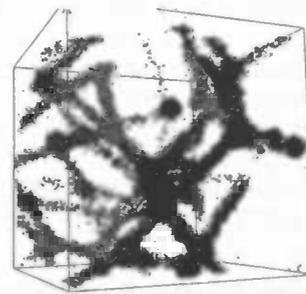
(c) The third pair of kinematic datasets (vorkinm.3.1 and vorkinm.3.2)



(d) And an example alpha shape (vorkinm.3.2 at $\alpha = 0.5 \cdot 10^{-4}$)



(e) The fourth pair of kinematic datasets (vorkinm.4.1 and vorkinm.4.2)



(f) And an example alpha shape (vorkinm.4.1 at $\alpha = 0.5 \cdot 10^{-4}$)

Figure 43: The first Betti number β_0 of the alpha shapes of the data sets discussed in the previous section. For the kinematic distributions, the rate of decrease of the Betti number for small values of α stands out.

The first sequence of graphs (Figure 43) shows the development of the first Betti number β_0 of the alpha shape for increasing α value. This Betti number is equal to the number of connected components of the shape. Since the alpha shape for $\alpha = 0$ consists of the vertices representing the input particles, the Betti number for this value is exactly 200,000. With an increasing α value, components can only be joined, so β_0 is a non-increasing number.

In Figure 43 the interval of α values where the Betti number decreases rapidly, is a clear discriminator between the distributions, because they differ clearly. To make this more precise we would have to look at the derivative of the function $\beta_0(\alpha)$, however this function is discrete, so the derivative of this function is not defined and an approximation has to be used as shown in Figure 44. The approximation used to calculate the derivative at a point uses the β_0 value at the two nearest (nearest with respect to α value) points (x and y) of topological change: $\frac{\partial \beta_0}{\partial \alpha} = \frac{\beta_0(y) - \beta_0(x)}{(y-x)}$. If we now take the interval to be where the smoothed version of the derivative is smaller than $-5 \cdot 10^9$, then the size of this interval and the minimum value of the derivative will be clear discriminators. The intended intervals are shown in Figure 44.

Comparing the filament and wall distribution, the filament has the interval at lower α values and a lower minimum derivative value. Since we can see the filament distribution as a more evolved structure, we expect to see the interval of rapid decrease of β_0 at lower α values and a lower minimum derivative value for the latter kinematic distributions. This is the case as we can see in the Figures 43 and 44. However the difference between the third and fourth kinematic distributions is too small to discriminate effectively between them using the interval only (the start of the interval is both close to zero and the difference of the end of the interval differ less than ten percent). The minimal derivative value is still clearly smaller for the more evolved structures of the fourth kinematic distribution. The two versions of the distributions, which have (as we recall from the previous section) eight and 64 generating void centers respectively, show the same development of β_0 for increasing α . This is a desired effect, since they represent the same structure.

Figure 46 shows the second Betti number, β_1 , plotted against α . β_1 represents the number of independent tunnels. Tunnels are formed when at a certain α value an edge is added between two vertices which are already connected, a tunnel can be filled, or destroyed, by adding faces.

Similar to β_0 , the maximum value of β_1 and the α value where this maximum value is reached can be used to discriminate between the distributions. This time we will see a lower maximum β_1 value for more evolved structure and this maximum will be reached at a lower α value. The difference between the two versions is once more very small.

Figure 48 shows the last Betti number (that is, the last having an interesting graph), β_2 , plotted against α . β_2 represents the number of holes in the shape. A hole is destroyed when a cell is added to the alpha shape.

The graphs of β_2 are strikingly different from those of the first two Betti numbers, in the sense that the two versions of the distributions show clearly different developments of the Betti number. Looking closely at the filament and wall distributions, it seems that the first versions (fil1 and wal1, the version with eight generating voids), have the Betti numbers about three times as high as the second versions (fil2 and wal2, the

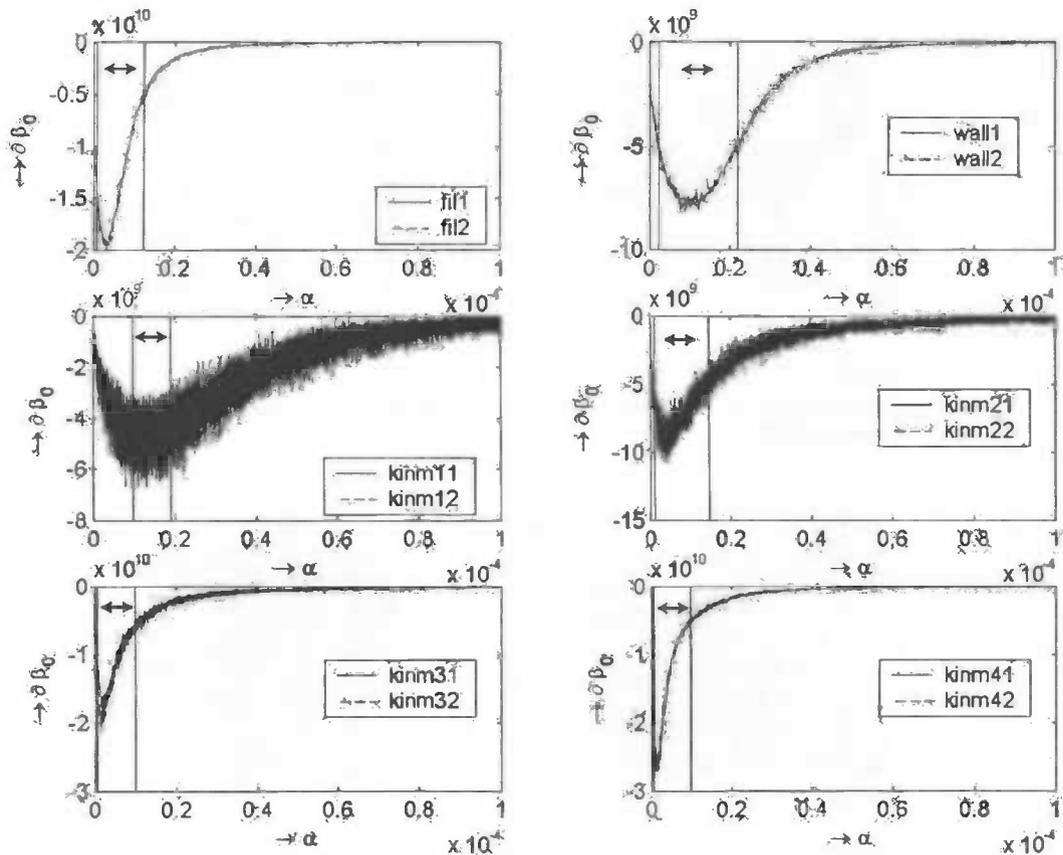
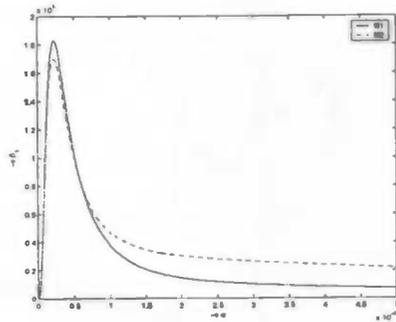


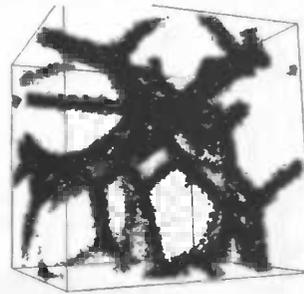
Figure 44: The "derivative" of β_0 : an approximation of $\frac{\partial \beta_0}{\partial \alpha}$ vs. α

version with 64 generating voids). β_2 for the kinematic distributions shows two local maxima, for the latter distributions (kinm3 and kinm4) the first maximum becomes more and more prominent compared to the second maximum. In this first maximum we can again see a factor of three difference between the two distributions just like in the filament and wall distributions. Actually, the filaments and wall distribution also have two local maxima, the second maximum however is too low to be visible in one graph with the first maximum (its highest β_2 value is around 40). This leads to the conclusion that the second maximum is probably due to the cells in the voids of the galaxy distribution. Since no particles are in these voids for the filament/wall distributions few holes can be formed there. However, how the factor three between the two versions comes to be, needs further investigation, which has not been done in this project because of time limitation.

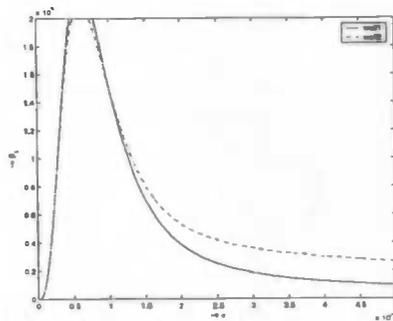
Figures 50 and 52 display the surface area and volume of the alpha shape. The surface area starts at zero and then grows quickly as many components appear in the alpha shape. When these components join and especially when holes within components are



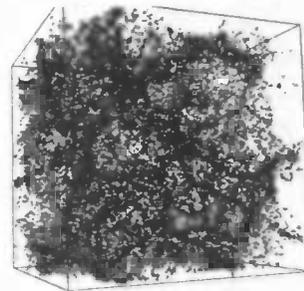
(a) The filament datasets (vorfil.1 and vorfil.2)



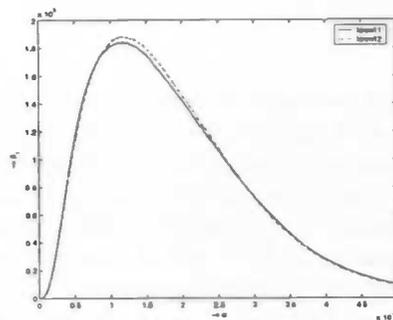
(b) And an example alpha shape (vorfil.1 at $\alpha = 0.5 \cdot 10^{-4}$)



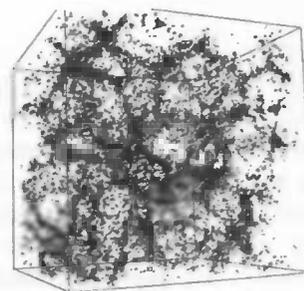
(c) The wall datasets (vorwall.1 and vorwall.2)



(d) And an example alpha shape (vorwal.2 at $0.5 \cdot 10^{-4}$)

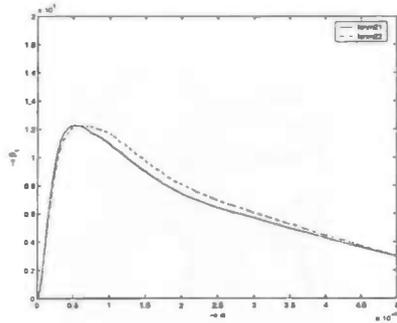


(e) The first pair of kinematic datasets (vorkinm.1.1 and vorkinm.1.2)

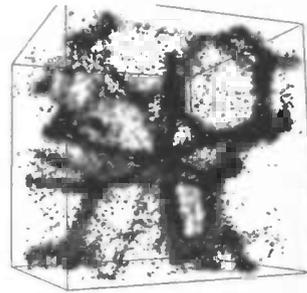


(f) And an example alpha shape (vorkinm.1.2 at $\alpha = 0.5 \cdot 10^{-4}$)

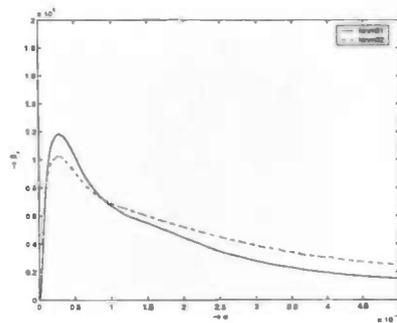
Figure 45: Figure continued at next page



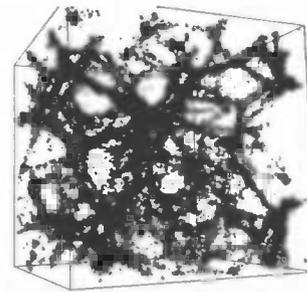
(a) The second pair of kinematic datasets (vorkinm.2.1 and vorkinm.2.2)



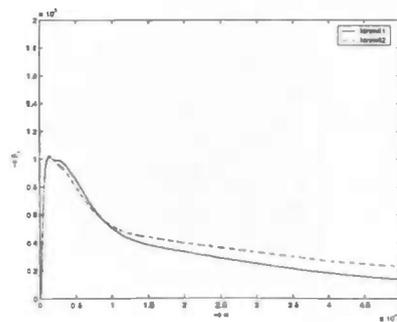
(b) And an example alpha shape (vorkinm.2.1 at $\alpha = 0.5 \cdot 10^{-4}$)



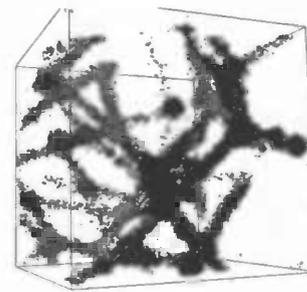
(c) The third pair of kinematic datasets (vorkinm.3.1 and vorkinm.3.2)



(d) And an example alpha shape (vorkinm.3.2 at $\alpha = 0.5 \cdot 10^{-4}$)

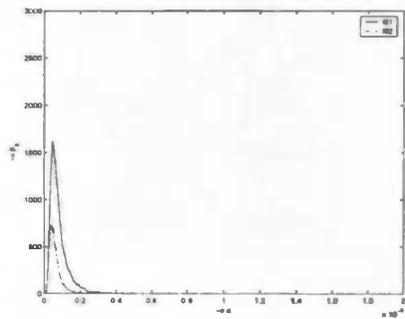


(e) The fourth pair of kinematic datasets (vorkinm.4.1 and vorkinm.4.2)

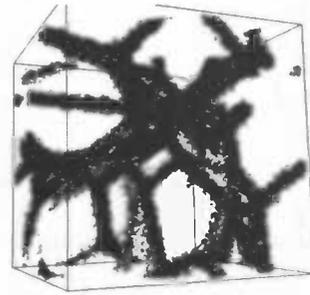


(f) And an example alpha shape (vorkinm.4.1 at $\alpha = 0.5 \cdot 10^{-4}$)

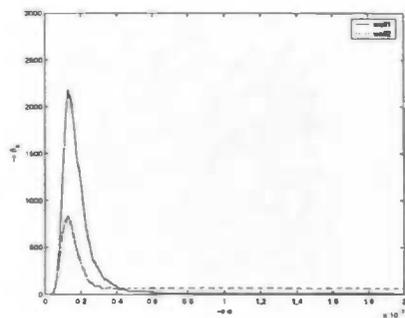
Figure 46: The second Betti number β_1 of the alpha shapes. The maximum value of β_1 and the α value where this maximum value is reached is clearly different for the distributions.



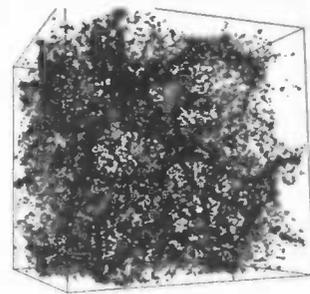
(a) The filament datasets (vorfil.1 and vorfil.2)



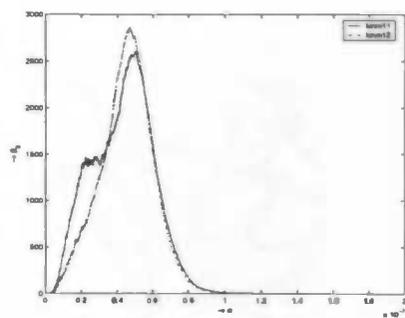
(b) And an example alpha shape (vorfil.1 at $\alpha = 0.5 \cdot 10^{-4}$)



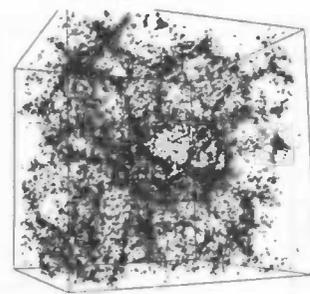
(c) The wall datasets (vorwall.1 and vorwall.2)



(d) And an example alpha shape (vorwal.2 at $0.5 \cdot 10^{-4}$)

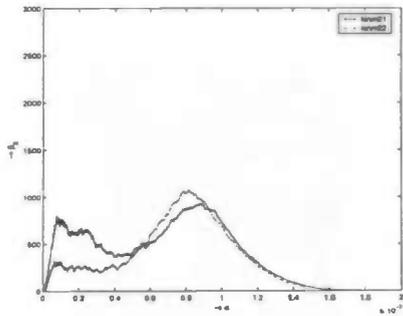


(e) The first pair of kinematic datasets (vorkinm.1.1 and vorkinm.1.2)

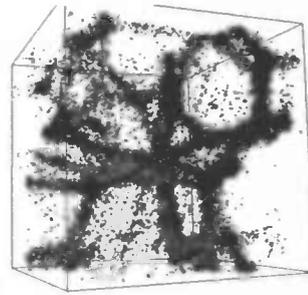


(f) And an example alpha shape (vorkinm.1.2 at $\alpha = 0.5 \cdot 10^{-4}$)

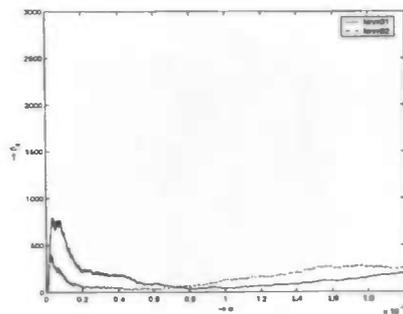
Figure 47: Figure continued at next page



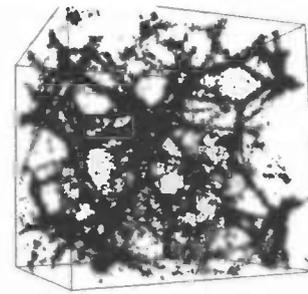
(a) The second pair of kinematic datasets (vorkinm.2.1 and vorkinm.2.2)



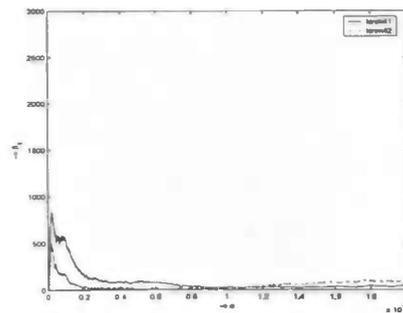
(b) And an example alpha shape (vorkinm.2.1 at $\alpha = 0.5 \cdot 10^{-4}$)



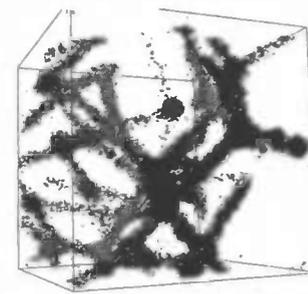
(c) The third pair of kinematic datasets (vorkinm.3.1 and vorkinm.3.2)



(d) And an example alpha shape (vorkinm.3.2 at $\alpha = 0.5 \cdot 10^{-4}$)

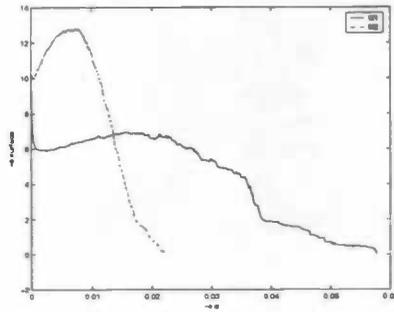


(e) The fourth pair of kinematic datasets (vorkinm.4.1 and vorkinm.4.2)

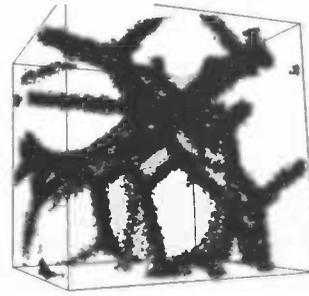


(f) And an example alpha shape (vorkinm.4.1 at $\alpha = 0.5 \cdot 10^{-4}$)

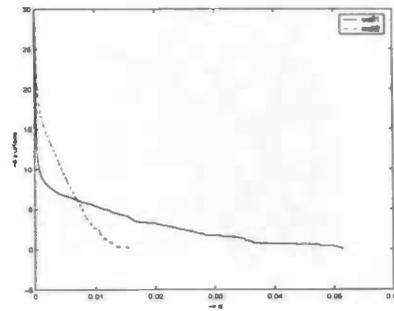
Figure 48: β_2 vs. α . The two local maxima of the kinematic distributions attract attention.



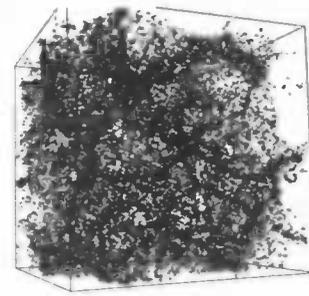
(a) The filament datasets (vorfil.1 and vorfil.2)



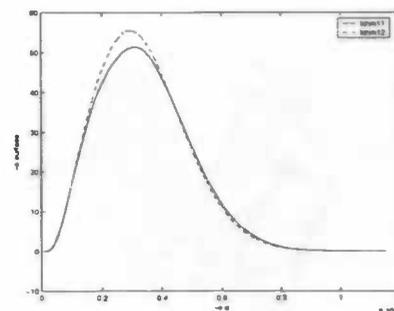
(b) And an example alpha shape (vorfil.1 at $\alpha = 0.5 \cdot 10^{-4}$)



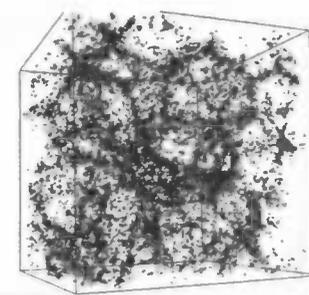
(c) The wall datasets (vorwall.1 and vorwall.2)



(d) And an example alpha shape (vorwal.2 at $0.5 \cdot 10^{-4}$)

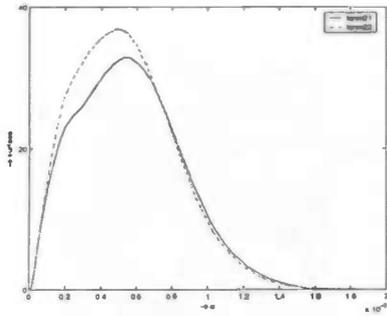


(e) The first pair of kinematic datasets (vorkinm.1.1 and vorkinm.1.2)

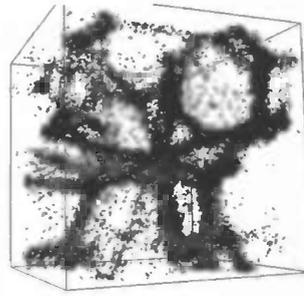


(f) And an example alpha shape (vorkinm.1.2 at $\alpha = 0.5 \cdot 10^{-4}$)

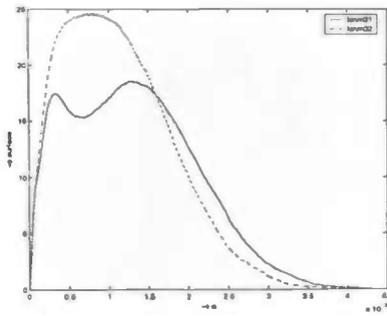
Figure 49: Figure continued at next page



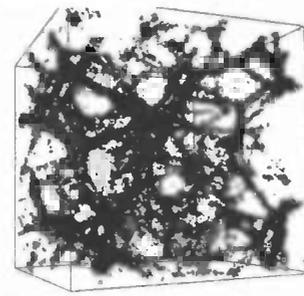
(a) The second pair of kinematic datasets (vorkinm.2.1 and vorkinm.2.2)



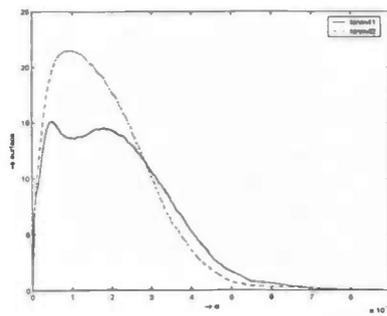
(b) And an example alpha shape (vorkinm.2.1 at $\alpha = 0.5 \cdot 10^{-4}$)



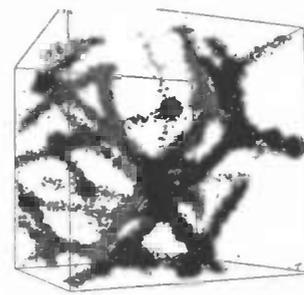
(c) The third pair of kinematic datasets (vorkinm.3.1 and vorkinm.3.2)



(d) And an example alpha shape (vorkinm.3.2 at $\alpha = 0.5 \cdot 10^{-4}$)

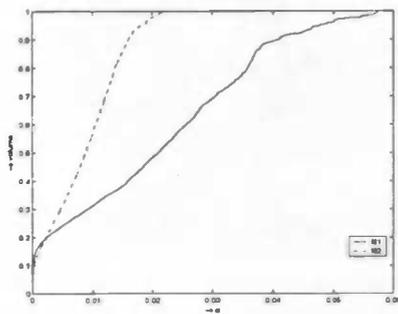


(e) The fourth pair of kinematic datasets (vorkinm.4.1 and vorkinm.4.2)

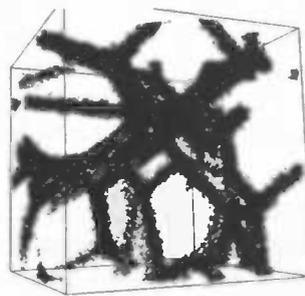


(f) And an example alpha shape (vorkinm.4.1 at $\alpha = 0.5 \cdot 10^{-4}$)

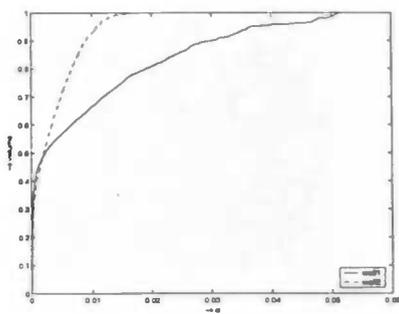
Figure 50: The surface area of the alpha shape as a function of α . Note the differences on the horizontal axis.



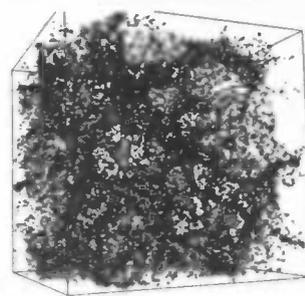
(a) The filament datasets (vorfil.1 and vorfil.2)



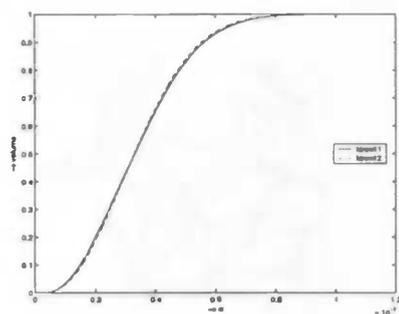
(b) And an example alpha shape (vorfil.1 at $\alpha = 0.5 \cdot 10^{-4}$)



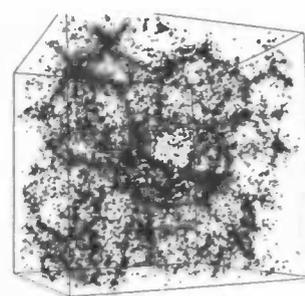
(c) The wall datasets (vorwall.1 and vorwall.2)



(d) And an example alpha shape (vorwall.2 at $0.5 \cdot 10^{-4}$)

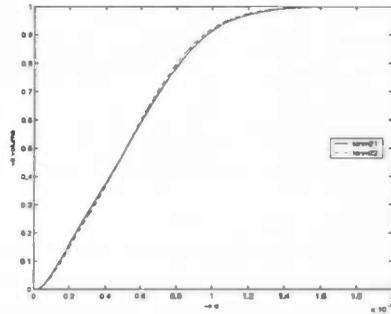


(e) The first pair of kinematic datasets (vorkinm.1.1 and vorkinm.1.2)

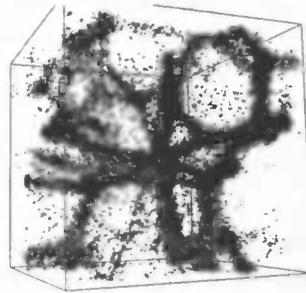


(f) And an example alpha shape (vorkinm.1.2 at $\alpha = 0.5 \cdot 10^{-4}$)

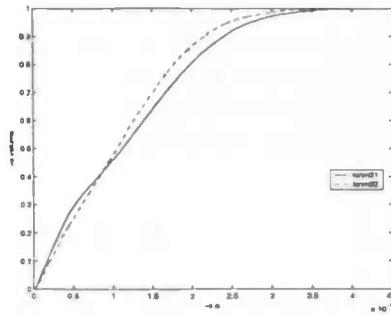
Figure 51: Figure continued at next page



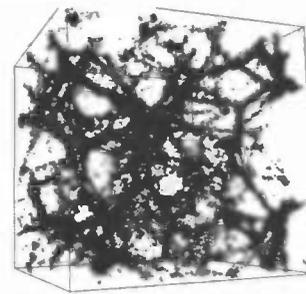
(a) The second pair of kinematic datasets (vorkinm.2.1 and vorkinm.2.2)



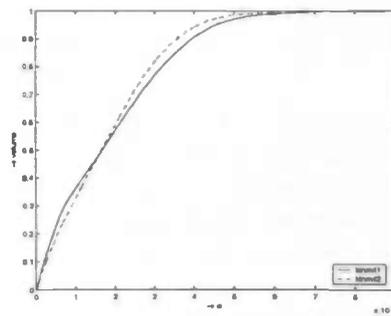
(b) And an example alpha shape (vorkinm.2.1 at $\alpha = 0.5 \cdot 10^{-4}$)



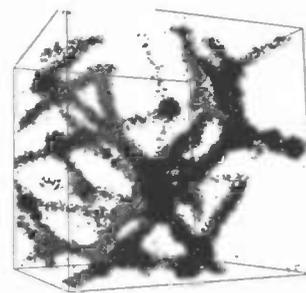
(c) The third pair of kinematic datasets (vorkinm.3.1 and vorkinm.3.2)



(d) And an example alpha shape (vorkinm.3.2 at $\alpha = 0.5 \cdot 10^{-4}$)



(e) The fourth pair of kinematic datasets (vorkinm.4.1 and vorkinm.4.2)



(f) And an example alpha shape (vorkinm.4.1 at $\alpha = 0.5 \cdot 10^{-4}$)

Figure 52: The volume of the alpha shape as a function of α . Note the differences on the horizontal axis.

filled the surface area shrinks again. When the whole unit cube is filled with the alpha shape the surface area is back at zero. The volume of the alpha shape also start at zero and grows when a cell is added to the alpha shape. Since the galaxy distribution model is simulated in a unit cube, its final volume is always one. The almost straight line during most of the growth of the volume is noticeable for the kinematic distributions. It means that the number of smaller cells added in the beginning of the development of the alpha shape compensates for the increased size during the latter stages of the development.

The most striking difference between filament/wall and kinematic distributions is that for the filament/wall distribution the difference between the two versions of the distributions is huge. This can be explained by noting that for the filament and wall distributions there are no particles in the voids at all. This will cause Delaunay cells with size proportional to the size of the voids. This size is obviously smaller for the second version, which implies that these cells are added to the alpha shape at a (much) lower α value.

10 Conclusions and future work

10.1 Astrophysical

Looking at the graphs in the previous section we have seen that Betti numbers can be used to discriminate between different distributions. A positive sign for usefulness of the Betti numbers of alpha shapes to discriminate between the galaxy distributions is that the graphs of the two versions of the kinematic distributions are nearly overlapping for β_0 , β_1 and the volume of the alpha shape. This is not the case for β_2 and the surface area of the alpha shape. To investigate this further, more distribution data sets are needed, it is hard to draw conclusions from twelve distributions. It would also be nice to have a direct interpretation from the Betti numbers about the actual structure of the galaxy distribution, comparable to the Shapefinders discussed in Section 2.

With more insight into the meaning of the Betti numbers for the structure of the distribution, it would be easier to understand the outcome of applying them to actual N -body simulations or maybe even to galaxy redshift surveys. Yet all the Betti numbers discussed can be used to discriminate between the galaxy distributions used for this research project, therefore they promise to be a useful tool to discriminate between models of the evolution of the structure in our universe.

10.2 Topological

When we try to manually count the number of holes in alpha shapes, like in Figure 3, we can maybe count a couple of dozens. However the graphs in the previous section show that the Betti numbers are in the hundreds. The reason for this much higher number is that many small holes, too small to see in figures like Figure 3, exist. Looking closely at the alpha shape, we will see a lot of irregularities at the surface and a lot of small connected components. These simplices seem to be noise in relation to the rest of the alpha shape. However all these simplices contribute to the Betti numbers (see Section 6), while they do not represent the holes we are most interested in, see Figure 53.

The holes we are interested in are the ones that remain for a longer period of alpha, the interval of alpha in which the hole exists. The authors of [13] propose a method to measure the lifetime of k -dimensional holes and to simplify the complex using this measurement. k -Dimensional holes that exist for a longer period than some parameter value p are used in the Betti numbers calculation, the others are discarded. Here the length of a period p can be measured by the length of the α interval or the number of topological changes that occur when α is gradually increased. This will allow the Betti number to be connected to the visual appearance of the alpha shape, however it will introduce another parameter p that needs to be chosen (ad hoc). To choose the value of this parameter a graph like the one shown in Figure 54 can be used, or one can choose to use this graph as a tool to discriminate instead of the two dimensional graphs used in this paper. The graph shows β_1 as a function of the persistence, that is the parameter p , and the number of topological changes.

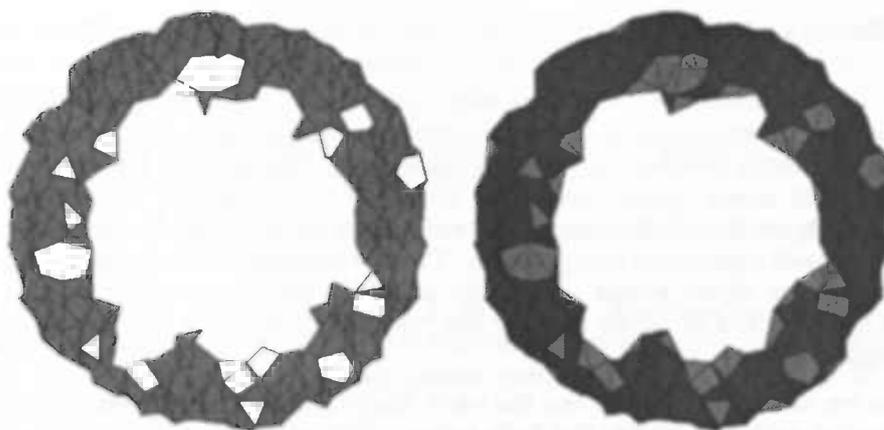
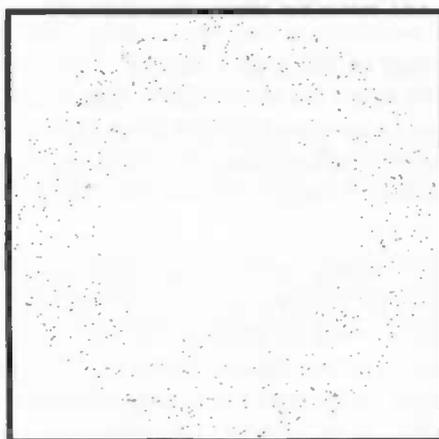


Figure 53: Two alpha shapes, for different values of α for the point set shown on the top, this point set is chosen randomly on a torus. The alpha shape on the left has 18 holes, only one of those is interesting to us. The alpha shape on the right still shows that one hole.

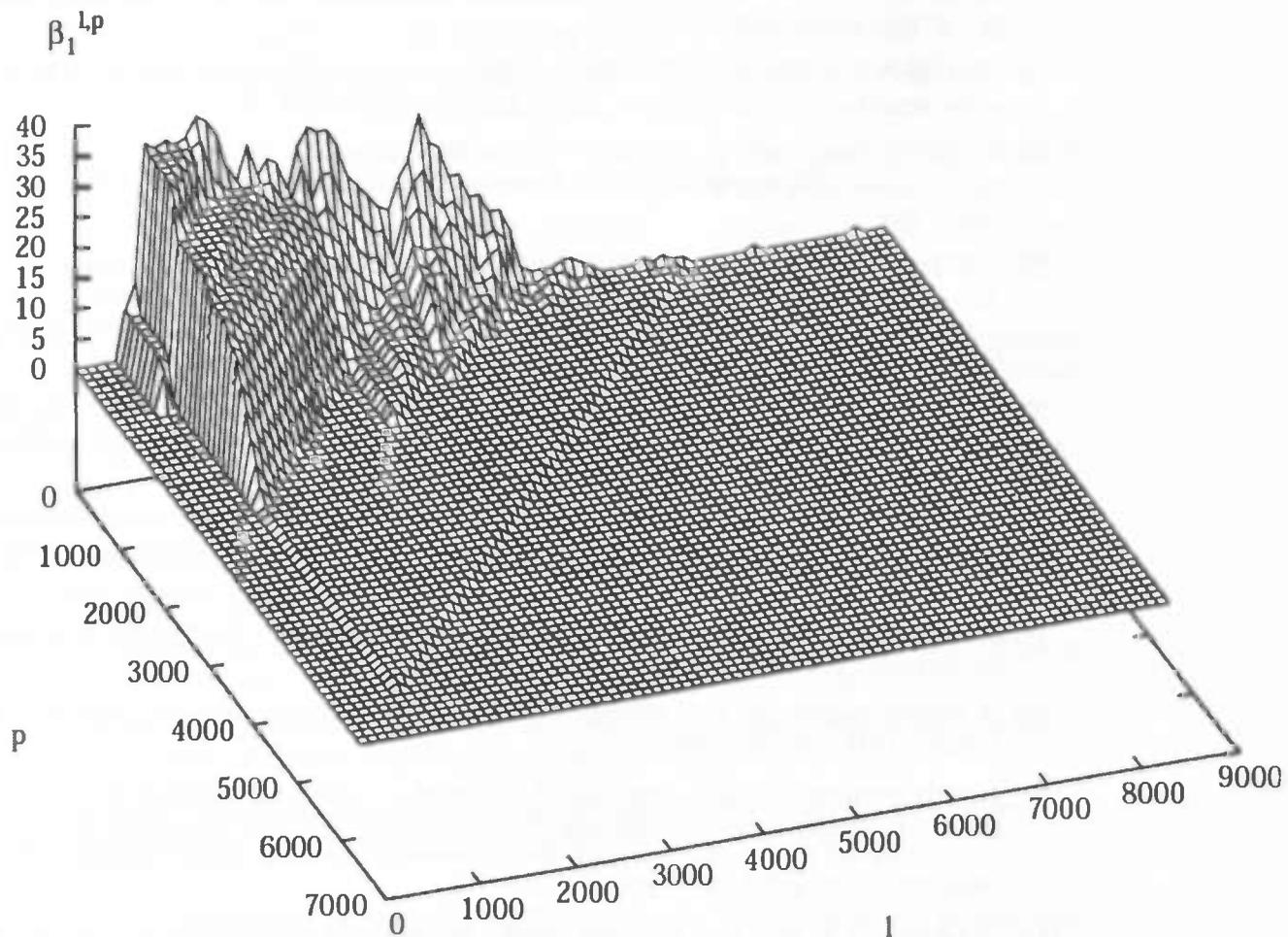


Figure 54: The persistence of β_1 of the alpha shape of a data set representing a Gramicidin A molecule (from [13]). Instead of the α value, the number of topological changes l is used to plot β_1 as a function of the persistence p .

References

- [1] F. Aurenhammer. Voronoi diagrams - a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23(3):345–405, 1991.
- [2] A.M. Ben-Amram, S. Alstrup and T. Rauhe. Worst-case and Amortised Optimality in Union-Find. *Proc. 31th Annual ACM Symposium on Theory of Computing (STOC'99)*, pages 499–506, 1999.
- [3] T Buchert K.R. Mecke and H. Wagner. Robust morphological measures for large-scale structure in the universe. *ArXiv Astrophysics*, 1993.
- [4] F. Da, T. Kai and M. Yvinec. 3D Alpha Shapes. *CGAL manual*, 2004. http://www.cgal.org/Manual/doc_html/cgal_manual/Alpha_shapes_3/Chapter_main.html
- [5] T.K. Dey, H. Edelsbrunner and S. Guba. Computational Topology Chapter in *Advances in Discrete and Computational Geometry edited by Chazelle, Goodman and Pollack*, vol. 223, *Contemporary Mathematics Series, American Mathematical Society*, pages 109–143, 1999.
- [6] G. De Frabitiis and P.V. Coveney. An empirical comparison of techniques for updating Delaunay triangulations *SCG '04: Proceedings of the twentieth annual symposium on Computational geometry*, 2004, 170–179.
- [7] C.J.A. Delfinado and H. Edelsbrunner. An incremental algorithm for Betti Numbers of Simplicial Complexes. *Symposium on Computational Geometry*, pages 232–239, 1993.
- [8] H. Edelsbrunner. Weighted alpha shapes. Technical report, University of Illinois at Urbana-Champaign, Champaign, IL, USA, 1992.
- [9] H. Edelsbrunner and E.P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, 1994.
- [10] H. Edelsbrunner, Y.E.A. Ban and J. Rudolph. Interface surfaces for protein-protein complexes. In *RECOMB '04: Proceedings of the eighth annual international conference on Resaerch in computational molecular biology*, pages 205–212, New York, NY, USA, 2004. ACM Press.
- [11] J. Giesen, T.K. Dey and M. John. Alpha-shapes and flow shapes are homotopy equivalent. *STOC '03: Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 493–502, 2003.
- [12] J. Giesen and M. John. The flow complex: a data structure for geometric modeling. *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 285–294, 2003.
- [13] D. Letscher, H. Edelsbrunner and A. Zomorodian. Topological persistence and simplification. *FOCS '00: Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, page 454, 2000.
- [14] D.A. Kraaijpoel. First order multivariate interpolation methods. *Seismic ray fields and ray field maps: theory and algorithms, Appendix D*, 2003.

- [15] K. Lemmens. An investigation, implementation and comparison of 3 important particle simulation techniques: PP : Particle-Particle, PM : Particle-Mesh and TC : Tree-code. Master's thesis, University of Technology, Delft, 1998. <http://ta.twi.tudelft.nl/DV/Staff/Lemmens/MThesis.TTH/report.html>
- [16] X.Y. Li and S.H. Teng. Generating Well-Shaped delaunay meshes in 3D. *Proc. 12th Annu. ACM-SIAM Sympos*, pages 28–37, 2001.
- [17] W.E. Lorensen and H.E. Cline. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *Computer Graphics (Proceedings of SIGGRAPH '87)*, Vol. 21, No. 4:pp. 163–169, 1987.
- [18] M. Overmars, M. de Berg, M. van Kreveld and O. Schwarzkopf. *Computational Geometry Algorithms and Applications*. Springer-Verlag, 2000.
- [19] E. Michael. The Cosmological Constant. *Department of Astrophysical and Planetary Sciences at the University of Colorado, Boulder*, 1999. <http://super.colorado.edu/~michaele/Lambda/lambda.html>
- [20] P. J. E. Peebles. *The large-scale structure of the universe*. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 435 p., 1980.
- [21] S. Pion and M. Teillaud. 3D Triangulations. *CGAL manual*, 2004. http://www.cgal.org/Manual/doc_html/cgal_manual/Triangulation_3/Chapter_main.html
- [22] D. Pogosyan, J.R. Bond and L. Kofman. How Filaments Are Woven into The Cosmic Web. *Nature*, page 603, 1996.
- [23] V.T. Rajan. Optimality of Delaunay Triangulation in \mathbb{R}^d . *SCG '91: Proceedings of the seventh annual symposium on Computational geometry*, pages 357–363, 1991.
- [24] V. Robins. Computational Topology for Point Data: Betti Numbers of α -Shapes. *LNP Vol. 600: Morphology of Condensed Matter*, 600:261–274, 2002.
- [25] J.V. Sheth and V. Sahni. Exploring the Geometry, Topology and Morphology of Large Scale Structure using Minkowski Functionals. *ArXiv Astrophysics e-prints*, 2005.
- [26] B.S. Sathyaprakash, V. Sahni, and S. Shandarin. Morphology of clusters and superclusters in N -body simulations of cosmological gravitational clustering. *The Astrophysical Journal*, 508:551-569, 1998.
- [27] S.F. Shandarin, J.V. Sheth, V. Sahni and B.S. Sathyaprakash. Measuring the Geometry and Topology of Large Scale Structure using SURFGEN: Methodology and Preliminary Results. *Mon. Not. Roy. Astron. Soc.*, 343:22, 2003.
- [28] R.A. Simcoe. The Cosmic Web: Observations and Simulations of the intergalactic medium reveal the largest structures in the universe. *American Scientist online*, 2004.
- [29] G. Vegter. Notes on Computational Topology <http://www.cs.rug.nl/~gert/>, 2003.
- [30] R. van de Weygaert. Het kosmische schuim. <http://www.astro.rug.nl/~weygaert/>.

- [31] R. van de Weygaert. The cosmic foam: spatial clustering and stochastic geometric patterns in the universe. In *Statistical Challenges in Astronomy*. Springer, 2003.
- [32] A.J. Zomorodian. *Topology for Computing*. Cambridge University Press, 2005.

A Graphs

This section presents all graphs made for this research project. Some of these graphs are not presented in Section 9, because the conclusion that can be drawn from these graphs are practically the same as the conclusions drawn from the graphs presented in Section 9. To be complete the graphs that are discussed in Section 9 are also included in this section.

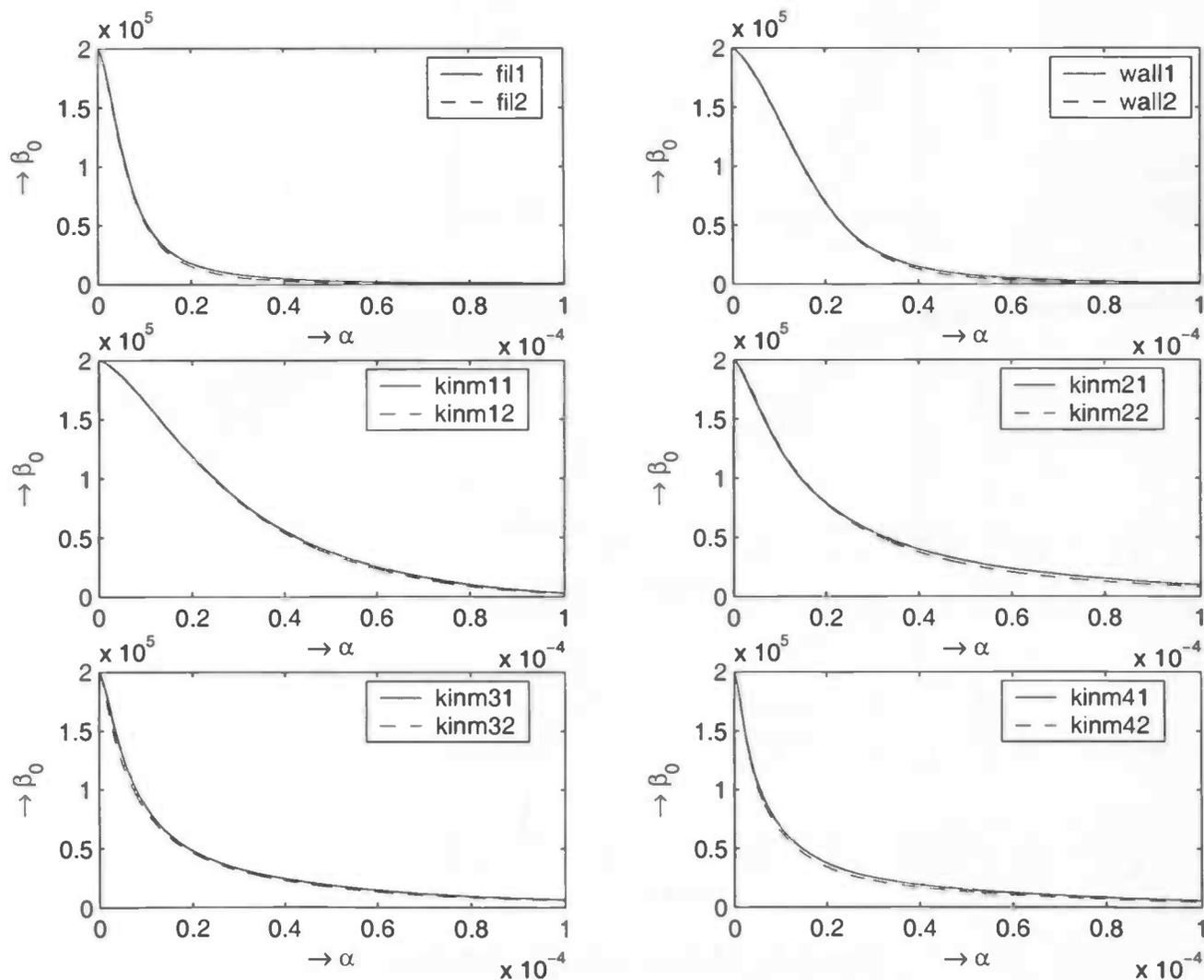


Figure 55: β_0 vs. α

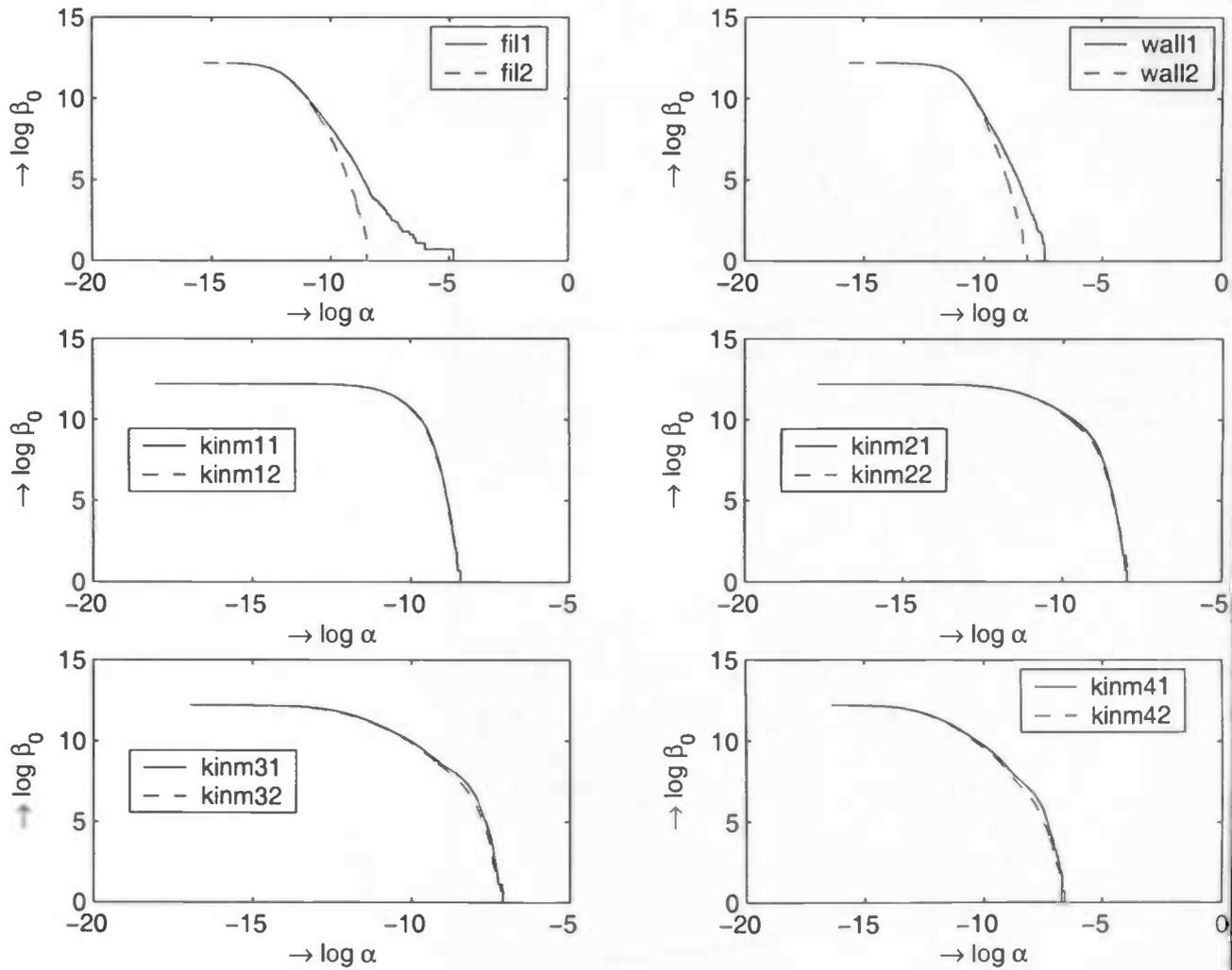


Figure 56: A log/log plot of the first Betti number: $\log \beta_0$ vs $\log \alpha$

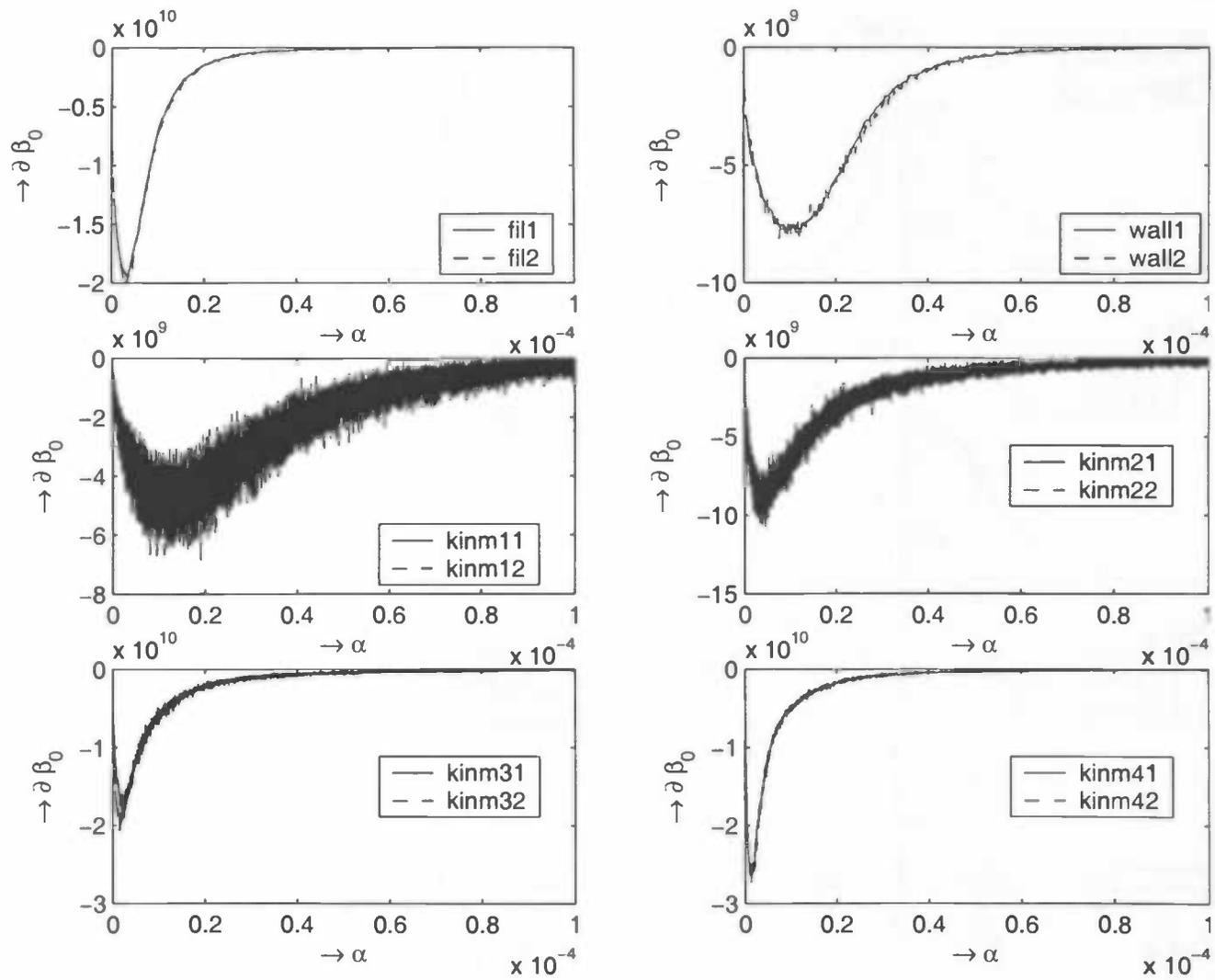


Figure 57: The "derivative" of β_0 : an approximation of $\frac{\partial \beta_0}{\partial \alpha}$ vs. α

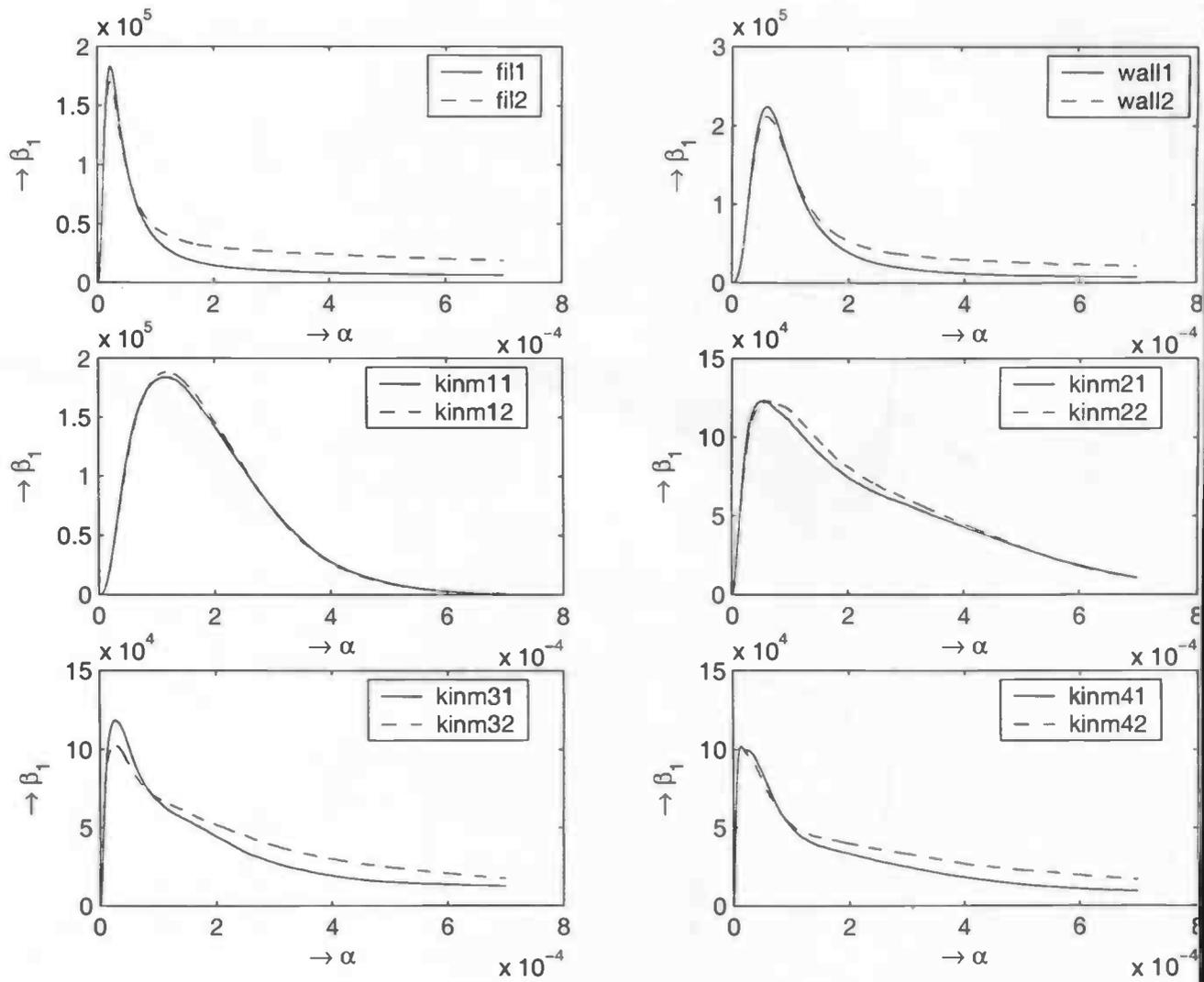


Figure 58: β_1 vs. α

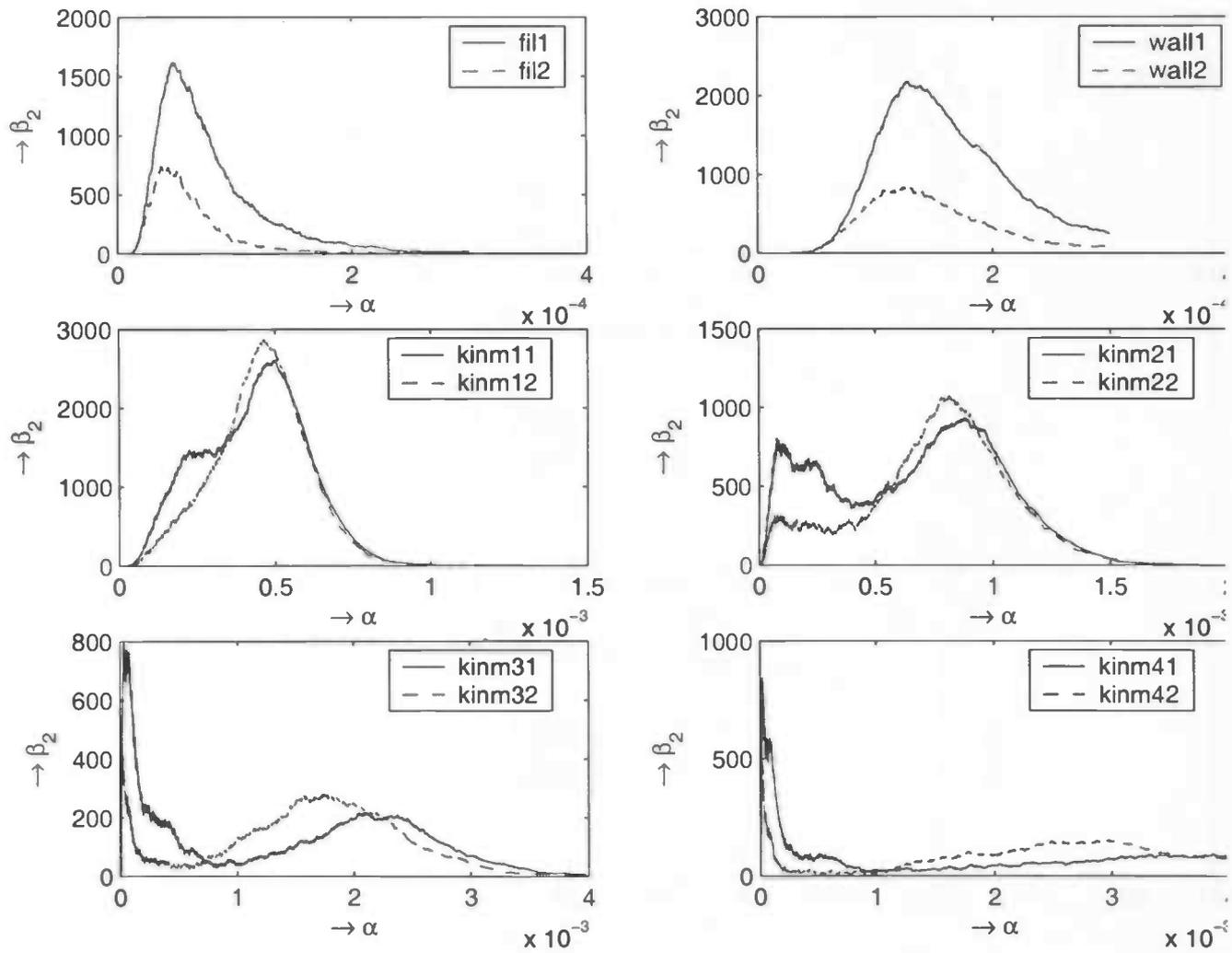


Figure 59: β_2 vs. α

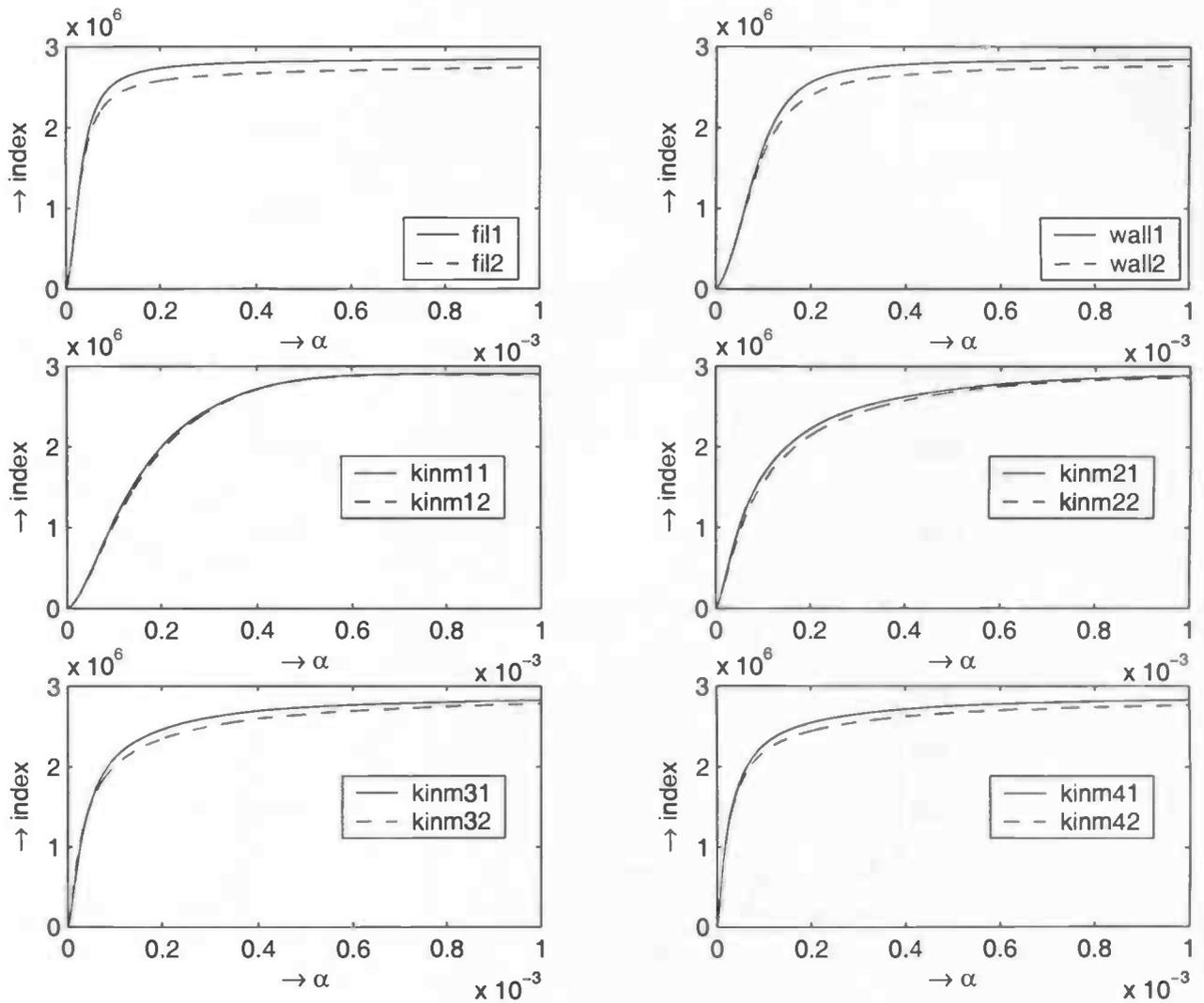


Figure 60: The number of topological changes vs. α . Here a topological change is when at a certain α value at least one vertex, edge, face or cell is added.

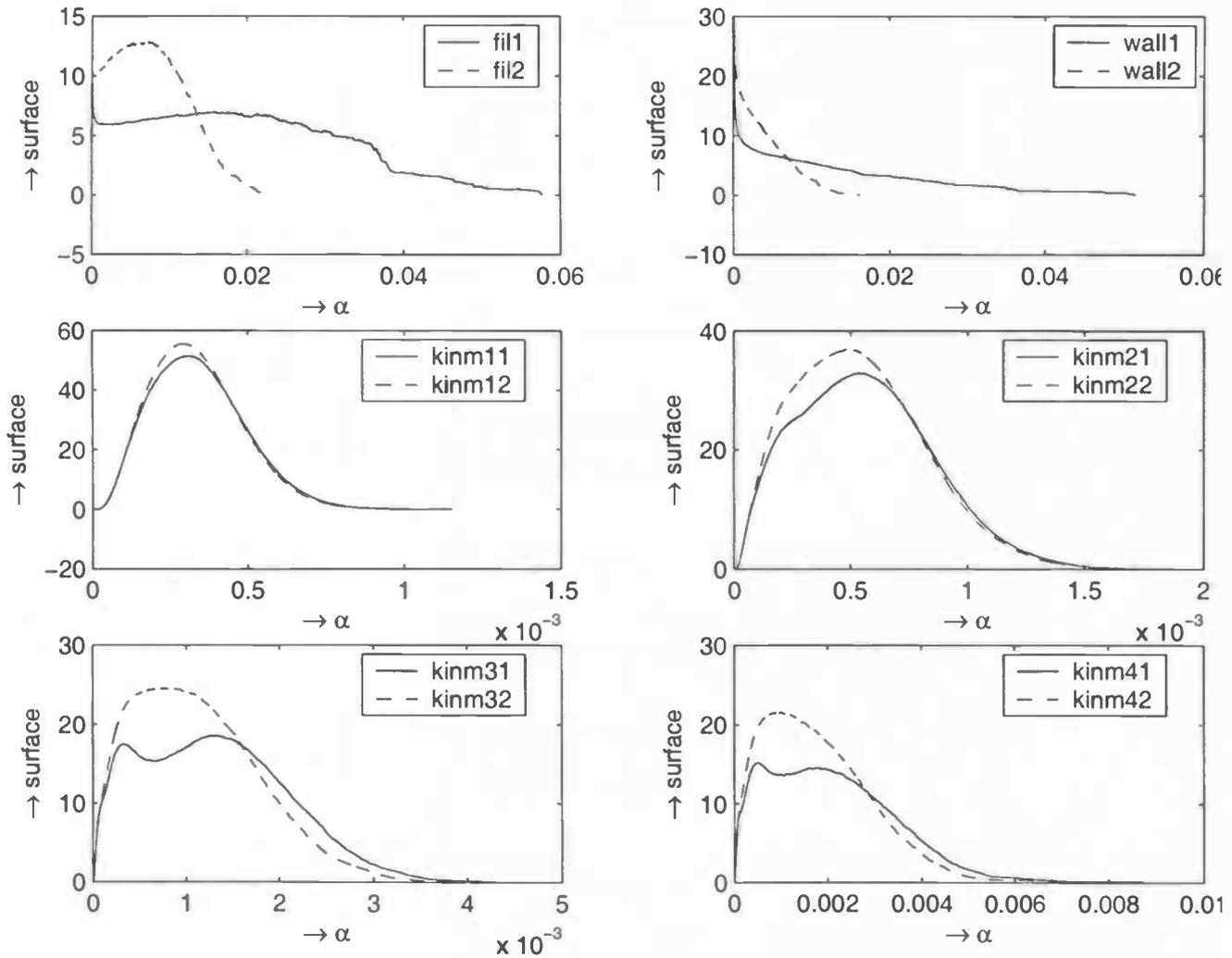


Figure 61: The surface area S vs. α

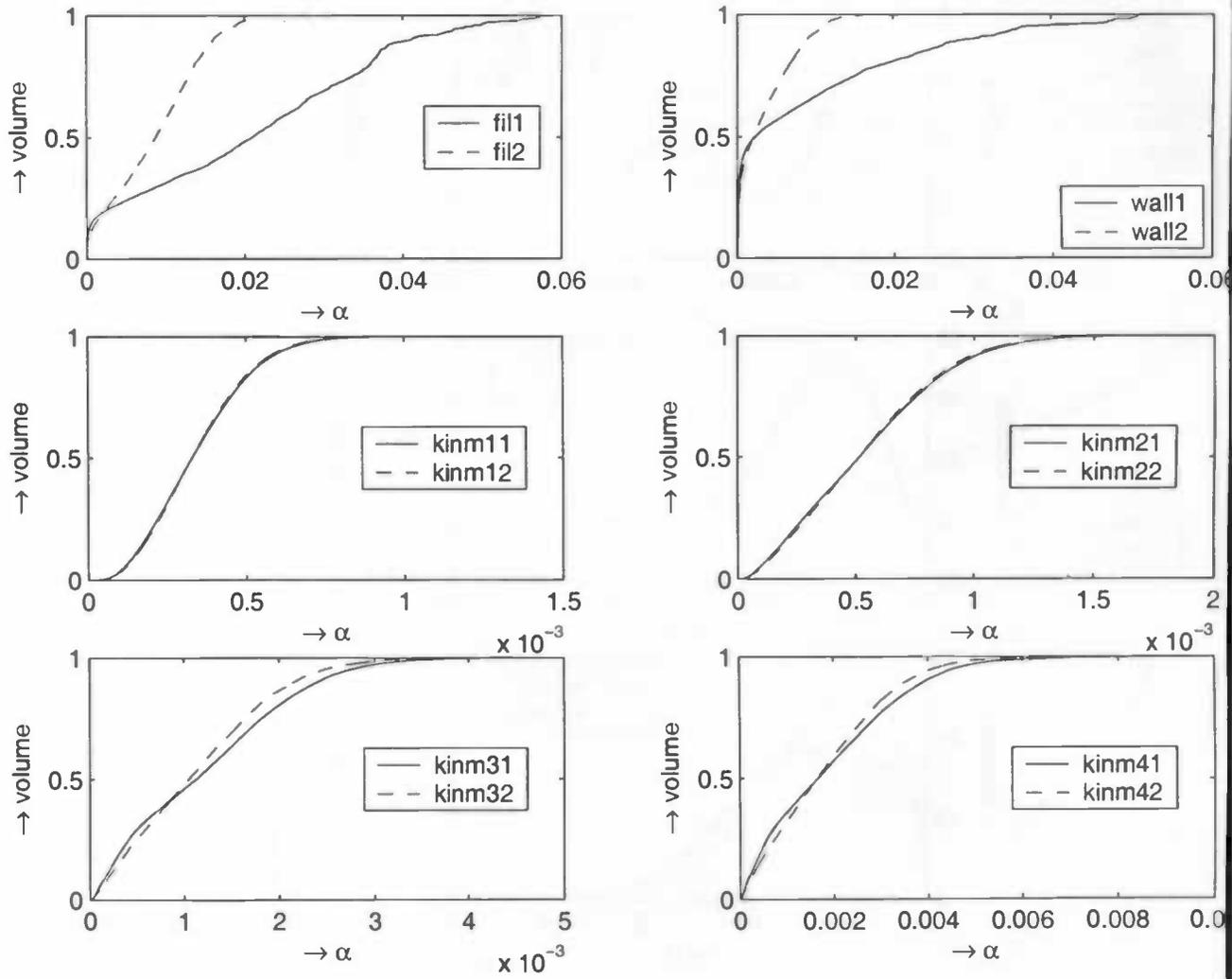


Figure 62: The volume V vs. α

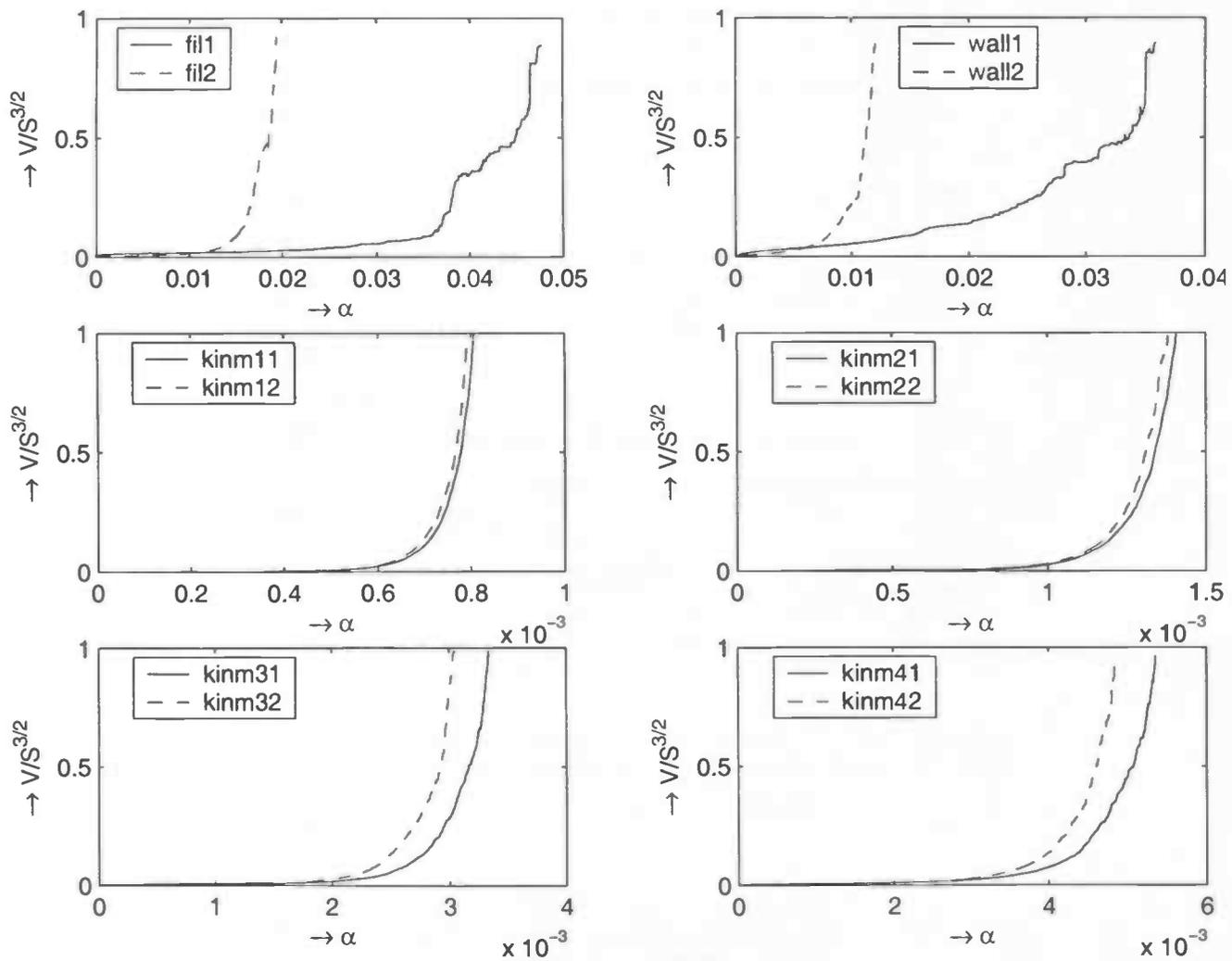


Figure 63: The dimensionless number $\frac{V}{S^{3/2}}$ vs. α , where V is the volume and S is the surface area of the alpha shape

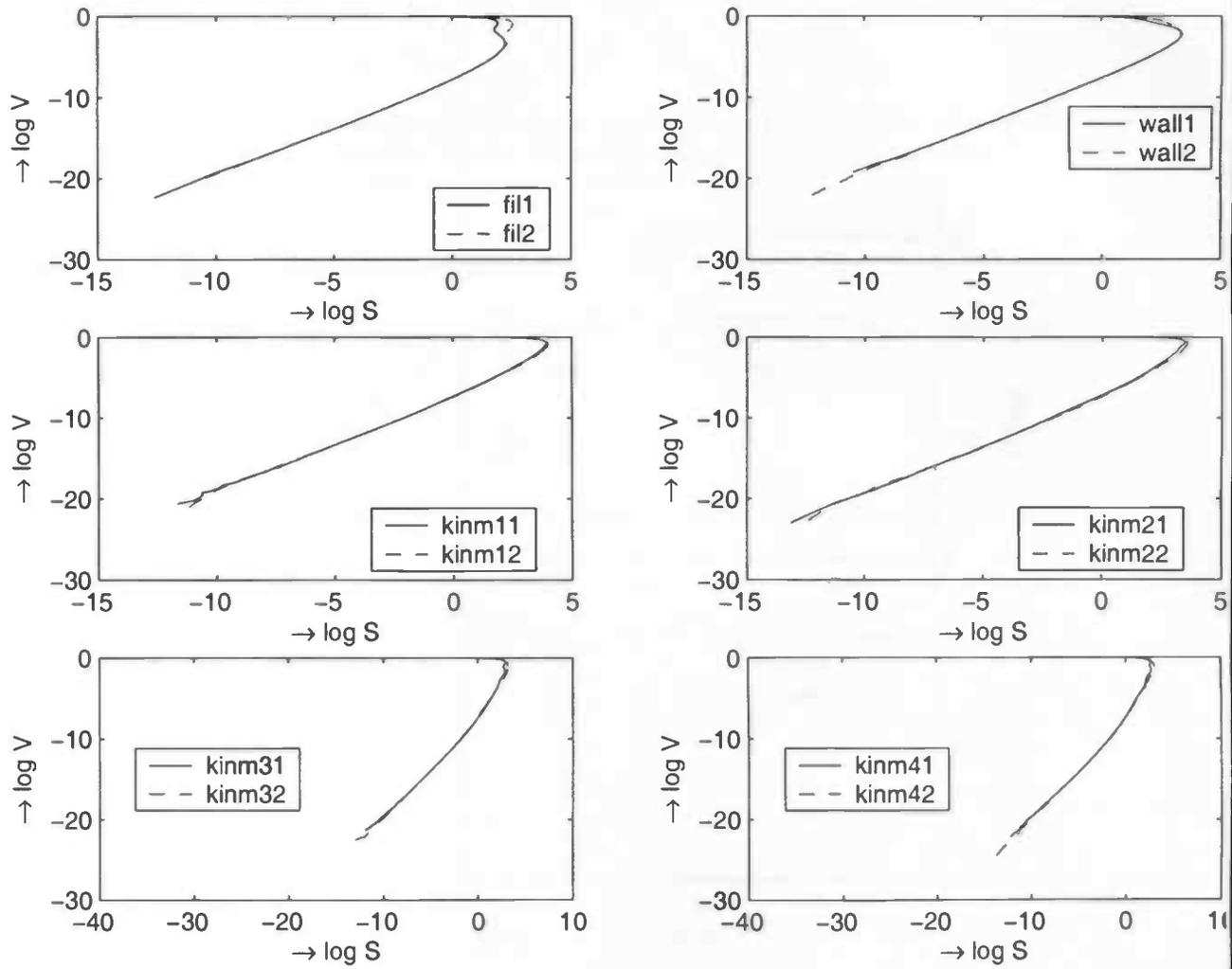


Figure 64: A log/log plot of the volume and surface area: $\log V$ vs. $\log S$

B Implementation

The program implemented has two functions: calculate the Betti numbers, surface area and volume of the serie of alpha shape of a galaxy distribution data set and output the alpha shape for visual inspection. For the underlying geometric structure, the CGAL library is used (see <http://www.cgal.org>). The CGAL library is a versatile collection of computational geometry algorithms, developed by several sites in Europe and Isreal.

The source code of the program consists of several files, which we will discuss in short in following subsections:

B.1 `alpha_shapes.C`

This is the main file where all main functions are called:

- reading the input galaxy distribution, or generating a completely random one
- making the data structure periodic
- computing the alpha shape
- printing the betti numbers, surface area and volume to the output
- writing parts of the components of the alpha shape to a `geomview` file. `Geomview` is an interactive 3D viewing program.

To run the program the executable has to be run with as arguments the input file and the output directory.

B.2 `defines.h`

This file is used to define some global parameters and functions. For example, which method for reading the data has to used and the functions `abs` and `sqr` are defined here.

B.3 `make_periodic.h`

Here are all the functions to approach the data structure as being periodic. The first that needs to be done, is to copy a number of particles in the periodi box to the neighboring boxes. It is not necessary to copy all particles to all neighboring boxes. The method to select which particles have to be copied is described in [6]. The main idea of this method is to place an extra particle at each vertex of the periodic cube, then the Delaunay triangulation is constructed. For each Delaunay sphere crossing a periodic boundary, the particles generating this sphere are copied to the box opposite to the crossed boundary. After all necessary particles are copied, a number of periodic vertices, but also edges, triangles and cell exist more than once in the data structure. Methods have been created to select one of them as the one used in all calculations, other methods can find this one unique simplex when one of the other copies is at hand.

B.4 output_betti.h

To calculate the Betti numbers, first all simplices (except vertices), have to be sorted on the α value of their first appearance in the alpha shape, as described in Section 5. After this sorting, the 2-cycle detection is run on the faces in reversed order and marked when required. Then the implementation of the abstract algorithm begins.

B.5 output_components.h

This file takes care of the output of the alpha shapes to a geomview file for visual inspection. It only outputs faces on the border of a component of connected cells, with a minimal number of cells (which can be one, to show all cells).

B.6 read_data.h and read_voronoi.h

These files read binary and ascii galaxy distribution data files respectively. The code will need to be adapted to the actual headers and format of the data files.

B.7 typedefs.h

Here the CGAL types are defined, like the definition of points, the Delaunay triangulation, alpha shape, etc.