Large-scale structure in the universe

D. G. Banhatti

School of Physics, Madurai Kamaraj University, Madurai 625 021, India

A survey of some cosmographic methods is presented here, with examples of results on the large-scale structure in the universe using these methods. Showing structure up to few 100 Mpc, the distribution of light tends to uniform random (i.e., Poisson) beyond that scale. The recently discovered substantial large-scale coherent flow of *all* visible matter relative to the rest frame of the 2.73 K cosmic background radiation is not discussed, nor are models and their predictions, including the dynamical importance of dark matter. The material presented forms part of the background needed to meaningfully discuss the 10 ppm ripples COBE observed in the 2.73 K radiation in 1992.

THE large-scale structure in the universe, as we observationally gauge it, consists of the distribution and motion of galaxies and galaxian objects and the distribution of diffuse radiation over the various windows through the electromagnetic spectrum. At present this means: (i) the distribution and motion of galaxies seen in the optical band over most of the sky, to a depth of $z \leq 0.1$ (Zwicky and Abell surveys of clusters, Center for Astrophysics (CfA) survey, etc.)[1]; (ii) the distribution of galaxies seen in the optical band over a portion of the celestial sphere deeper than $z \approx 0.1$, perhaps to $z > 1$ (Cambridge Automatic Plate-measuring Machine (APM) survey, Queen Mary-Durham-Oxford-Toronto (QDOT/QIGC) survey, etc.)[2]; (iii) the distribution of infrared sources over most of the sky (InfraRed Astronomical Satellite (IRAS) survey)[3]; (iv) the diffuse X-ray background which maps the overall distribution of clusters of galaxies and active galaxies, the nearer clusters identified with the broad maxima (mainly High Energy Astronomical Observatory 2 (HEAO 2 ≡ Einstein Observatory) satellite results)[4]; (v) the distribution of extragalactic radio sources (galaxies and quasars) over the celestial sphere much deeper than $z \approx 0.1$, perhaps to $z \geq 3$ (many low-frequency (≤ 1 GHz) radio surveys)[5] and (vi) the 2.73 K cosmic blackbody radiation peaking at microwaves (i.e., cmλ) and its (an)isotropies, observed on scales from few arcminutes to few degrees, including the recent very sensitive Cosmic Background Explorer (COBE) satellite results[6] showing an anisotropy of 10 ppm on scales > 7°. All these indicators

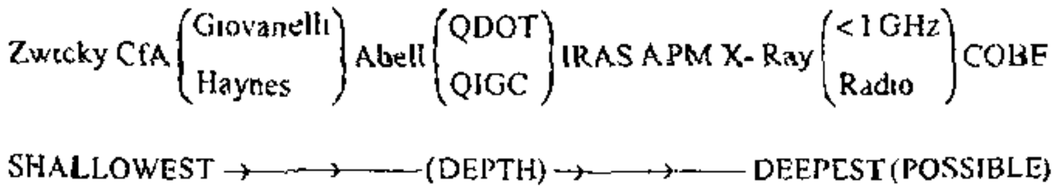

Figure 1. Surveys of large-scale structure

and determinants of large-scale structure in the universe have been listed above roughly in order of increasing depth: from the shallowest to the deepest possible (see Figure 1). The cosmic blackbody radiation is the deepest possible we can see. Matter was opaque to radiation before radiation decoupled from it and started expanding independently. No imprint of any structure before this decoupling epoch ($z \approx 1500$) is retained in the light we receive. Information about earlier epochs does come to us *indirectly*, e.g. from the abundance of light elements.

Dark matter[7]

On scales from galaxies to groups to clusters to superclusters and so on, dynamical mass as inferred from observed speeds of constituents is more than the visible mass inferred from their mass-to-light ratios. The discrepancy increases with increasing scale. For structures on the various scales to be sufficiently long-lived (or stable), dark matter must be present, and dominate the dynamics more and more on larger scales. So models must incorporate dark matter in some way. However, I only treat different ways of describing the observed large-scale structure, leaving models alone for the most part.

Methods

Analysis of observations of discrete objects (galaxies, groups or clusters of galaxies, radio sources, voids, etc.) is essentially the examination of the properties (like clustering) of (possibly weighted) point sets. If the distance (i.e. redshift) of each object in a survey is known, we have a three-dimensional point set. If not, the point set is two-dimensional, consisting of the projections of the objects' positions on the celestial sphere. Some methods use these point sets as such, while others use smoothed versions. I sketch below the methods that

Presented at a one-day meeting on COBE Results on 22 January 1993 at the Institute of Mathematical Sciences, Madras 600 113, India

Table 1. Source densities for 11 fields (Banhatti '90MN246 7P-10P)

Field	Area (deg^2)	No of srcs	Src density (x) (deg^{-2})	Texas counts σ	$(x-\bar{x})/\sigma$	6C-II counts src density (deg^{-2})
M 33 = 0130 + 30 = 133° – 32°	24 0	44	1 83 ± 0 28	0 27	+ 0 33	–
M 31 = 0041 + 41 = 121° – 22°	25 2	51	2 02 ± 0 28	0 26	+ 1 06	–
3C 236 = 1003 + 351 = 190° + 54°	18 4	39	2 12 ± 0 34	0 31	+ 1 24	2 28 ± 0 35
DA 240 = 0744 + 56 = 162° + 30°	17 4	25	1.44 ± 0 29	0 32	– 0 95	–
1331 – 099 = 320° + 51°	19 7	27	1 37 ± 0 26	0 30	– 1 25	–
12^h + 35° = 174° + 77°	25 0	41	1 64 ± 0 26	0 26	– 0 38	1 60 ± 0 25
00^h + 10° = 104° – 51°	25 0	39	1 56 ± 0 25	0.26	– 0 68	–
11^h + 25° = 212° + 65°	25 0	32	1 28 ± 0 23	0 26	– 1 74	–
02^h + 00° = 158° – 58°	25 0	53	2 12 ± 0 29	0 26	+ 1 44	–
1313 – 099 = 313° + 52°	24 6	41	1 66 ± 0 26	0 27	– 0 30	–
09^h15^m + 37° 5 = 186° + 44°	25 0	50	2 00 ± 0 28	0 26	+ 0 98	2.00 ± 0 29

where $\bar{x} = 1.73 \pm 0.08\ deg^{-2}$

In the first column, the source name or field designation is given, followed by the R A and Dec (epoch 1950) The approximate galactic coordinates are given in the second line

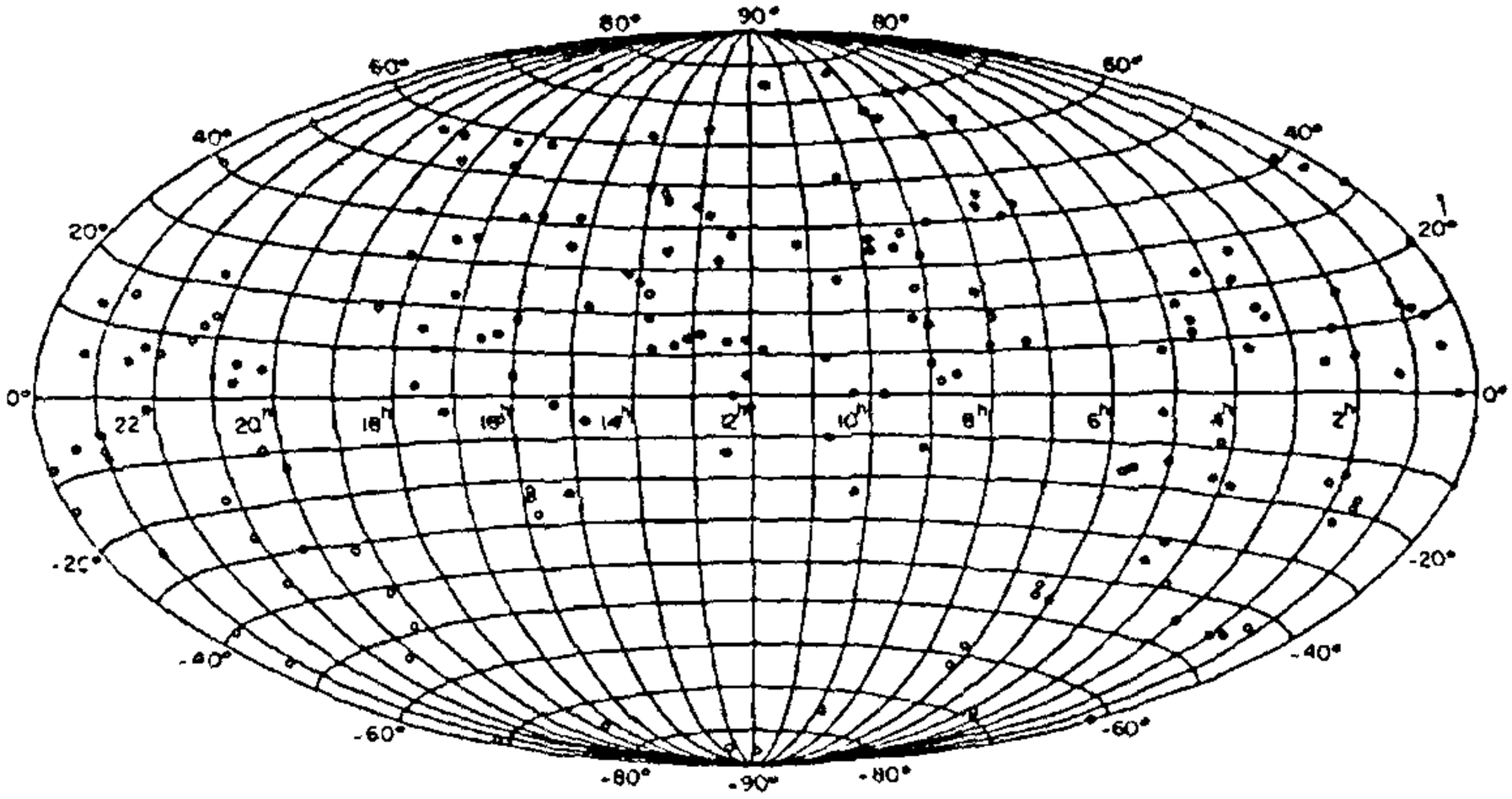

Figure 2. Distribution of over 180 voids in equatorial coordinates Limiting redshift 30,000 km s^{-1}.

have been used, and indicate broadly the kind of information that each method can give, with examples of results obtained. I have described the newer methods in more detail than the older ones.

1. Binning or counting the number of objects in different parts of a survey is the simplest method. This may be done in three dimensions if redshifts (indicators of depth) are available, or else in two dimensions (on the celestial sphere). Sometimes the (optical) apparent magnitude or (radio) flux density may serve as a crude indicator of distance (i.e. depth). For uniformly and randomly distributed points, Poisson statistics is expected – the population of a bin being proportional to its size, and the fluctuation in the number equal to its square root. Such an analysis of the 2nd Bologna (B2)

survey of (extragalactic) radio sources down to 0.25 Jy at 408 MHz shows that the number of radio sources within a cube of size 1 Gpc or more varies by < 3% as the cube is moved from place to place[8]. From the numbers of sources down to 0.25 Jy at 365 MHz (from Texas survey) in 11 randomly selected fields (avoiding any Galactic features) of sizes 18 to 25 deg^2, the distribution of (extragalactic) radio sources on the celestial sphere is fully consistent with Poisson statistics (Table 1)[9]. Extragalactic radio sources are generally associated with faint distant optical galaxies. In contrast, the distribution of the brighter (and nearer) galaxies is clustered (e.g. Zwicky and Abell clusters). Somewhat deeper, but still shallower than radio surveys, the optical CfA survey shows clusters of galaxies arranged in sheets enclosing voids. Data on these and other voids (Figure 2)[10] show more distant voids to be bigger (Figure 3 and Table 2)[10]. The sizes of voids range from 27 Mpc to 270 Mpc and the distances from $z \approx 0.003$ to $z \approx 0.1$. The enumeration of these voids is subjective, and needs confirmation by an appropriate objective procedure, similar to an algorithm which broke up the distribution of optical galaxies into groups and clusters[11].

2 Power spectrum analysis consists of resolving the sum of possibly weighted δ-functions at the source positions on the celestial sphere into orthogonal sets of functions (Fourier series or spherical harmonics), and using the sum of the squares of the amplitudes of the different components as relevant statistics. In practice, this method has given lower limits (generally ≈ degrees) above which any clustering occurs. Although in principle it is possible to use this method for three-dimensional data, a more convenient method of covariance function analysis (see below) is used for surveys with known redshifts. The results of applying power spectrum analysis to many radio surveys are summarized in Table 3 (ref. 5). The general conclusion is that radio sources are distributed independently, uniformly at random on the celestial sphere on scales of a degree and above, corresponding to more than a few 100 Mpc at estimated depths.

3 Covariance function analysis (or use of *two-point correlation functions*) gives the excess density of sources above that expected for a uniform random distribution at different linear or angular distances from each other, averaged over all the sources. Using two-dimensional surveys, this method has given upper limits (≤ degrees) below which any clustering occurs. This method and the previous one can be considered a Fourier pair. If the two-point spatial correlation function has a power-law form with index $-\gamma$, the corresponding angular function also has a power-law form with index $1 - \gamma$. Observations give $\gamma \approx 1.7$ from both spatial and angular surveys up to 10 Mpc and 1° respectively.

Table 2. Other possible relationships for the plot of Figure 3

Type	Equation	Correlation
Linear	$\log V = -5.798 + 2.558 \log d$	0.615
Power	$\log V = 0.156 \log d^{2.95}$	0.643

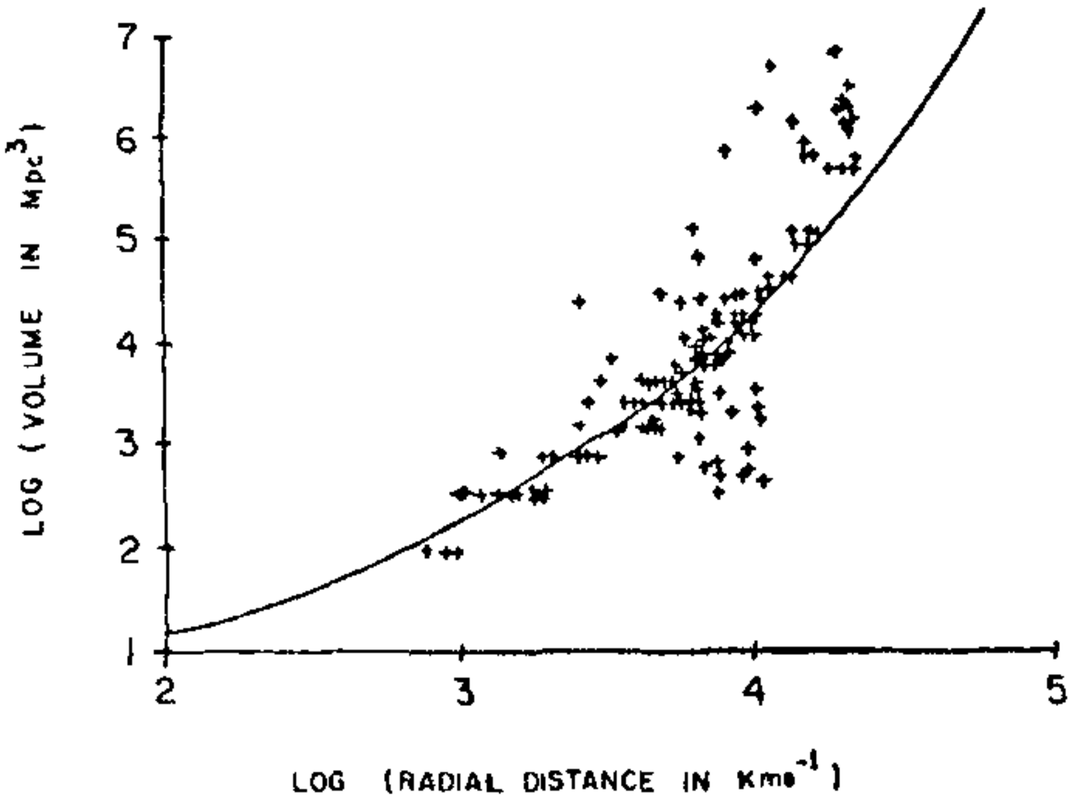

Figure 3. Correlation between volumes of voids and their distance. The curve is $\log V(\mathrm{Mpc}^3) = 0.308 \times \exp(0.66 \log d(\mathrm{km\ s}^{-1}))$ with correlation = 0.658.

Recent measurements of the angular correlation function on scales from 0.01° to 10° confirm the power-law of slope −0.7 to 1° and show a gradually steepening decline to zero at 6° (ref. 2, entry 1). Three-point and further correlation functions may also be used, but are not as well-determined as the two-point function. Two-point correlation functions are also used for two different types of surveys covering the same region of sky. Thus, the 3C radio sources (≈ 200) do not show clustering among themselves on scales of a few degrees, but 3C radio galaxies occur preferentially near the bright galaxy clusters (as seen in the optical band), while 3C quasars (which are an order of magnitude more distant) show no such preference[5].

4 Distribution functions are of two types. One, denoted by $f_R(N)$, measures the probability of finding different numbers N of objects in a given area of sky or a given volume of space specified by R. This was first applied systematically to galaxies by Hubble, who used samples with large values of N. The other complementary distribution function $f_N(R)$ measures the probability of a given number N of objects occupying different areas of sky or volumes of space given by R. Each may be computed for angular and spatial surveys with a specific procedure, and compared with the two-parameter prediction for relaxed gravitational clustering. In this article, I have avoided any discussion of models, but make an exception for this simple and powerful model.

Table 3. Summary of power spectrum analysis applied to many radio surveys[5]

Survey	Frequency (GHz)	Area (deg^2)	Flux density limit (Jy)	No. of beams per source	Results and remarks
4C	0.178	22800	2.0	30	No clustering on all scales ≥ 4° Slight anticlustering
GB	1.4	525	0.09	19	No clustering on all scales ≥ 1° Slight anticlustering
			0.25	90	Weak clustering on scales ≥ 10.7 with 1.7 sources/cluster
MC1	0.408	690	0.1	330	No clustering on all scales ≥ 1°.1
			0.40	690	No clustering on all scales ≥ 1°.8
PKS I, II	2.7	22700	0.35	1560	No clustering on all scales ≥ 5°
PKS III			0.26	1190	No clustering on all scales ≥ 8°
B2	0.408	4770	0.25	87	No clustering on all scales ≥ 0°.7 Slight anticlustering
5C	0.408	≈ 63	≈ 0.01	40	No clustering on all scales ≥ 0°.4

One model parameter is the average number of objects in the given area or volume, and the other, denoted by b, essentially equals half the ratio of the gravitational correlation energy to the kinetic energy of peculiar motion. For no clustering, i.e. uniform random (≡ Poisson) distribution, $b = 0$, and $b = 1$ if all objects are clustered. Application of this method to Zwicky and Abell surveys of clusters, and to a small deep 1.4 GHz radio survey (297 sources down to 5 mJy in a $4° \times 4°$ area) gives $b(\text{Zwicky}) = 0.70 \pm 0.05$ on scales 1 to 10 Mpc, $b(\text{Abell}) = 0.29 \pm 0.08$ (10 to 50 Mpc) and $b(\text{radio}) = 0.01 \pm 0.01$ (≥ 100 Mpc)[13]. Thus gravitational clustering decreases with scale, disappearing completely beyond 100 Mpc.

5 *Fractal geometry*[14] may be applicable on scales from a few Mpc to a few 100 Mpc. de Vaucouleurs found hierarchical structure in the distribution of galaxies singly or in pairs or multiplets (0.1 Mpc) forming groups and small clusters (few Mpc) as well as rich clusters (tens of Mpc) and all these forming superclusters (few 100 Mpc), to the depth of Abell survey[15]. Rich clusters have up to a thousand galaxies, although most galaxies are in smaller groups. The density of visible matter against the scale on which it is determined is a power-law of slope −1.7 (ref. 16) consistent with the hierarchical (or fractal) structure, and also with the slope of the two-point correlation function of the distribution of galaxies (see (3) above). The fractal dimension of the structure is thus $3 - 1.7 = 1.3$. Deeper, especially radio, surveys have shown that the hierarchy probably does not continue, radio sources being distributed uniformly at random (see (2) and (4) above). But since actual distances to radio sources (i.e. their redshifts) are largely not yet determined, it is well to keep an open mind.

In the method of analysis using the covariance function or the two-point (or pair) correlation function ((3) above) there is an implicit assumption of uniformity (or homogeneity) on a scale within the survey extent. An alternative equivalent method without this assumption is possible, and necessary if fractal structure persists to the largest scale accessible to a survey. This method uses the conditional number density instead of the covariance function.

To bring out clearly the difference and the relationship between these two methods, I give details. For a spatial survey over volume V having N objects of equal weights at positions $\mathbf{r}_i$, $i = 1, \ldots, N$, the number density $n(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i)$. Defining average by $\langle \ldots \rangle_{\mathbf{r}} \equiv \frac{1}{V}\int_V \ldots d\mathbf{r}'$, where $d\mathbf{r}'$ denotes an element of volume, the cross-correlation function of $n(\mathbf{r})$ is

$$G(\mathbf{r}) \equiv \langle n(\mathbf{r}')n(\mathbf{r}'+\mathbf{r})\rangle_{\mathbf{r}'} = \frac{1}{V}\sum_{i=1}^{N} n(\mathbf{r}_i + \mathbf{r})$$

$$= \frac{1}{V}\sum_{i=1}^{N} n(\mathbf{r}_i - \mathbf{r}) = \frac{1}{V}\sum_{i=1}^{N}\sum_{j=1}^{N} \delta(\mathbf{r} - (\mathbf{r}_i - \mathbf{r}_j)).$$

The two-point (or pair) correlation function $\xi(\mathbf{r})$ is defined by

$$1 + \xi(\mathbf{r}) \equiv \langle n(\mathbf{r}')n(\mathbf{r}'+\mathbf{r})\rangle_{\mathbf{r}'} / \langle n\rangle^2,$$

where, clearly, $\langle n\rangle = N/V$. So

$$1 + \xi(\mathbf{r}) = G(\mathbf{r}) / \langle n\rangle^2 = \frac{V}{N^2}\sum_{i=1}^{N} n(\mathbf{r}_i + \mathbf{r})$$

$$= \frac{V}{N^2} \sum_{i=1}^{N} n(\mathbf{r}_i - \mathbf{r}) = \frac{V}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(\mathbf{r} - (\mathbf{r}_i - \mathbf{r}_j)).$$

Thus $\xi(\mathbf{r})$ involves the survey size V. (Note that the cross-correlation function $G(\mathbf{r})$ also involves V.) Consider now the conditional density

$$\Gamma(\mathbf{r}) \equiv \langle n(\mathbf{r}')n(\mathbf{r}'+\mathbf{r})\rangle_{\mathbf{r}'} / \langle n \rangle = G(\mathbf{r}) / \langle n \rangle.$$

This function is calculated to be

$$\Gamma(\mathbf{r}) = \frac{1}{N} \sum_{i=1}^{N} n(\mathbf{r}_i + \mathbf{r}) = \frac{1}{N} \sum_{i=1}^{N} n(\mathbf{r}_i - \mathbf{r})$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(\mathbf{r} - (\mathbf{r}_i - \mathbf{r}_j))$$

and does not involve the survey volume V. Note that $\xi(\mathbf{r})$ and $\Gamma(\mathbf{r})$ are related by

$$\xi(\mathbf{r}) = \Gamma(\mathbf{r}) / \langle n \rangle - 1$$

and

$$\Gamma(\mathbf{r}) = \langle n \rangle (1 + \xi(\mathbf{r})).$$

Assuming that $\Gamma(\mathbf{r})$ depends only on the magnitude r of $\mathbf{r}$, and denoting it by $\Gamma(r)$, an average conditional density $\Gamma^*(r)$ is defined as

$$\Gamma^*(r) \equiv \frac{3}{4\pi r^3} \int_0^r dr' 4\pi r'^2 \Gamma(r')$$

The results of analysing a volume-limited sample of CfA survey using $\xi(r)$, $\Gamma(r)$ and $\Gamma^*(r)$ are shown in Figure 4 (ref. 17). Although $\xi(r)$ drops sharply, $\Gamma(r)$ *and* $\Gamma^*(r)$ *continue decreasing smoothly to the largest* scale which can be probed using the sample. As seen for appropriately simulated samples, $\Gamma(r)$ and $\Gamma^*(r)$ would level at the scale at which homogeneity (or uniformity) sets in. As the $\xi(r)$, there has been a tendency to derive 'characteristic scales' r_0 and r_0' from $\xi(r_0) = 1$ and $\xi(r_0') = 0$, although a power-law variation implies a scale-free behaviour. If the derived scales r_0 and r_0' are really *characteristic of the distribution, they should be* indepedent of the sample size. However, for volume-limited samples of CfA survey r_0 and r_0' depend on the sample size (Figure 5)[17] (ref. 17) precisely in the way expected if there is no characteristic scale up to the largest sample size. For $\xi(r)$ a power-law, say $Ar^{-\gamma}$, r_0 defined by $\xi(r_0) = 1$ equals $A^{1/\gamma}$ so that $\xi(r) = (r/r_0)^{-\gamma}$. This latter way of writing a power-law gives a false impression of there being a characteristic scale r_0.

An analysis of the cluster–cluster correlation in Abell survey using conditional densities shows that there is no uniformity to the largest accessible scale (Figure 6)[17]. A misguided analysis interpreting r_0 as a 'correlation *length' gives rise to different galaxy-galaxy and* cluster–cluster 'correlation lengths', while the distributions are actually scale-free (Figure 7)[17], although, as

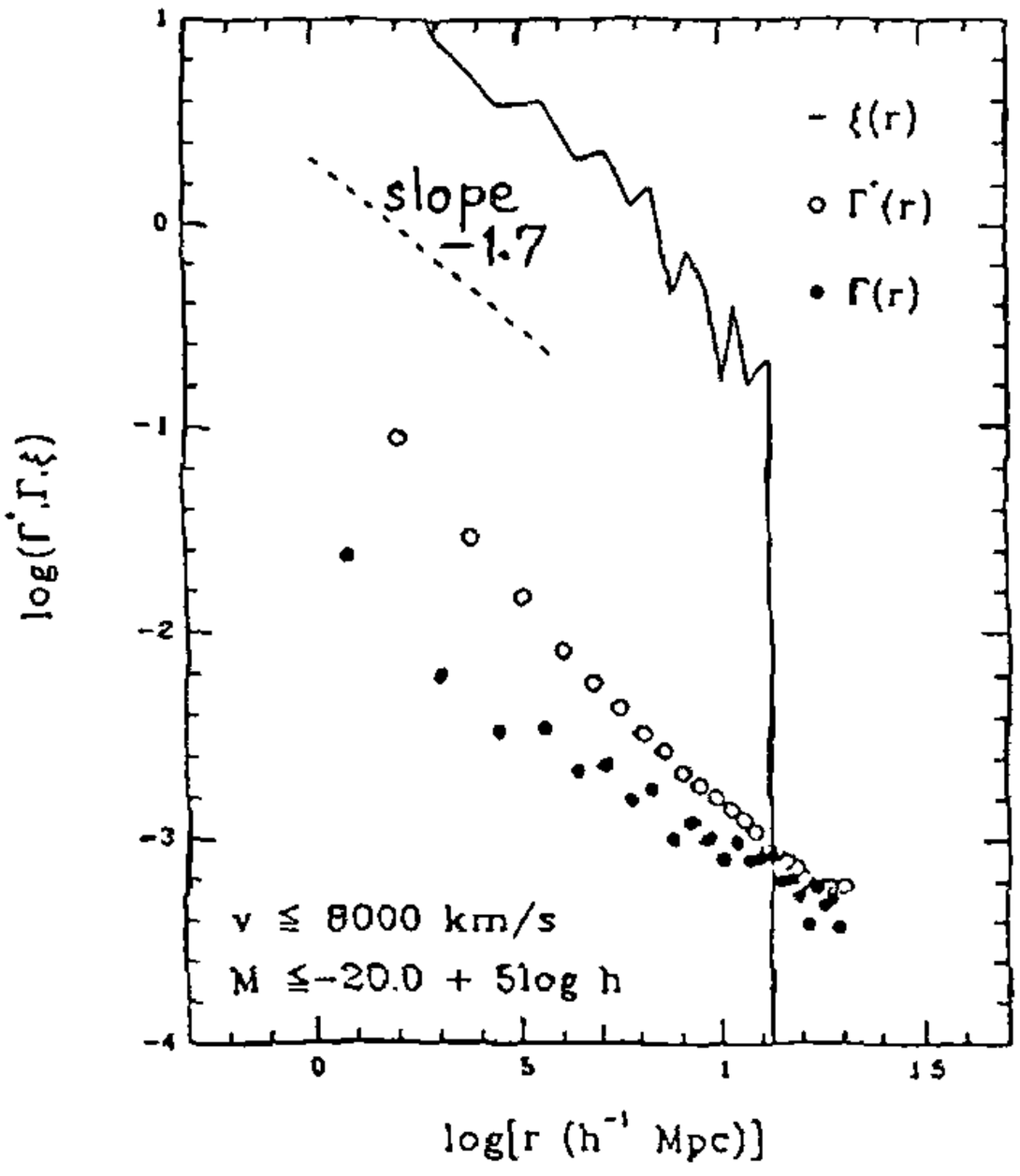

Figure 4. $\Gamma(r)$, $\Gamma^*(r)$ and $\xi(r)$ versus r, for the 226 galaxies of CfA survey with $cz \leq 8000$ km s^{-1} and $M \leq -20.0 + 5 \log h$

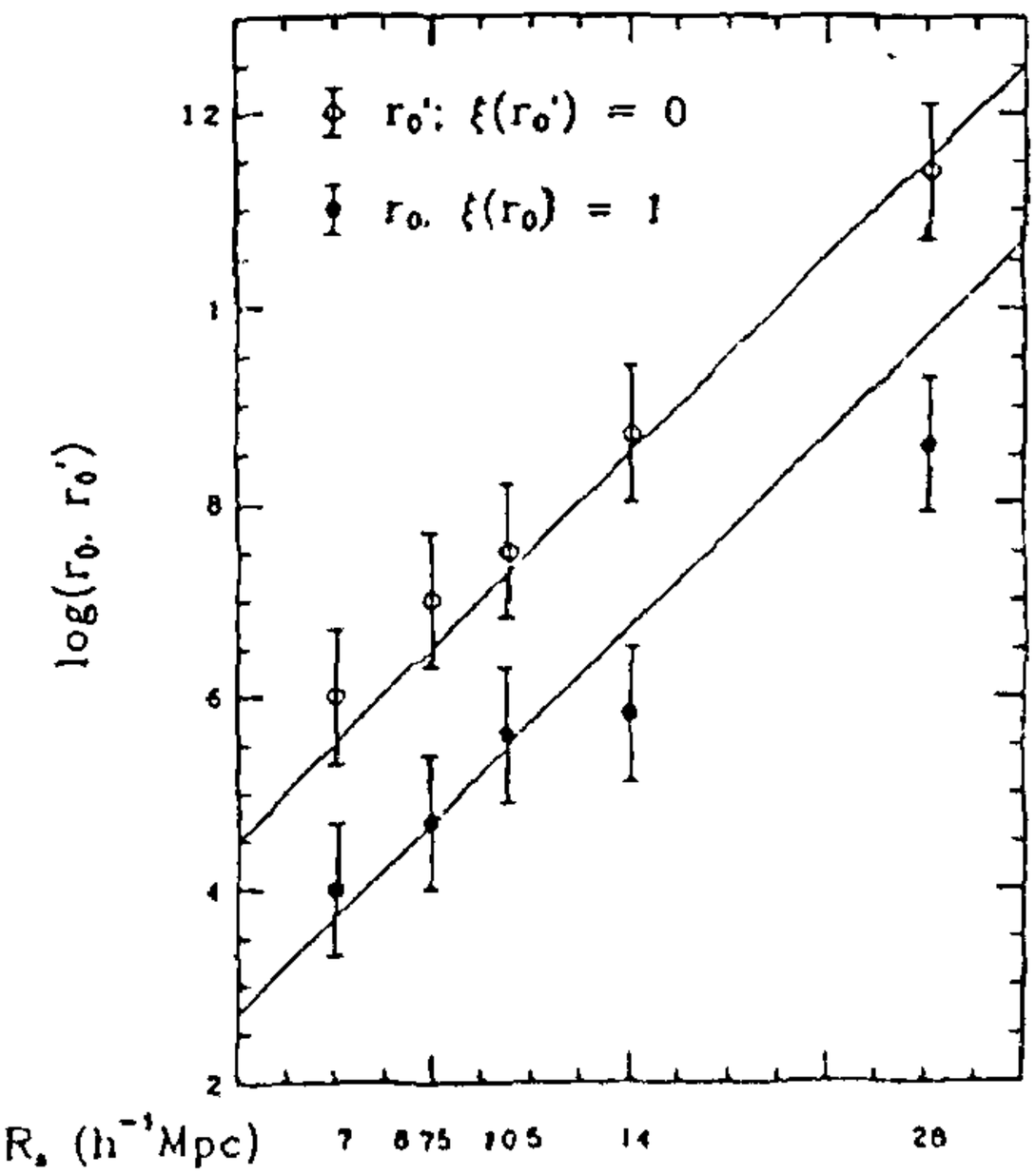

Figure 5. r_0 defined by $\xi(r_0) = 1$ and r_0' by $\xi(r_0') = 0$ plotted versus sample size R_s for which they were determined

stated toward the beginning of this section on fractal geometry, results from radio surveys, which are

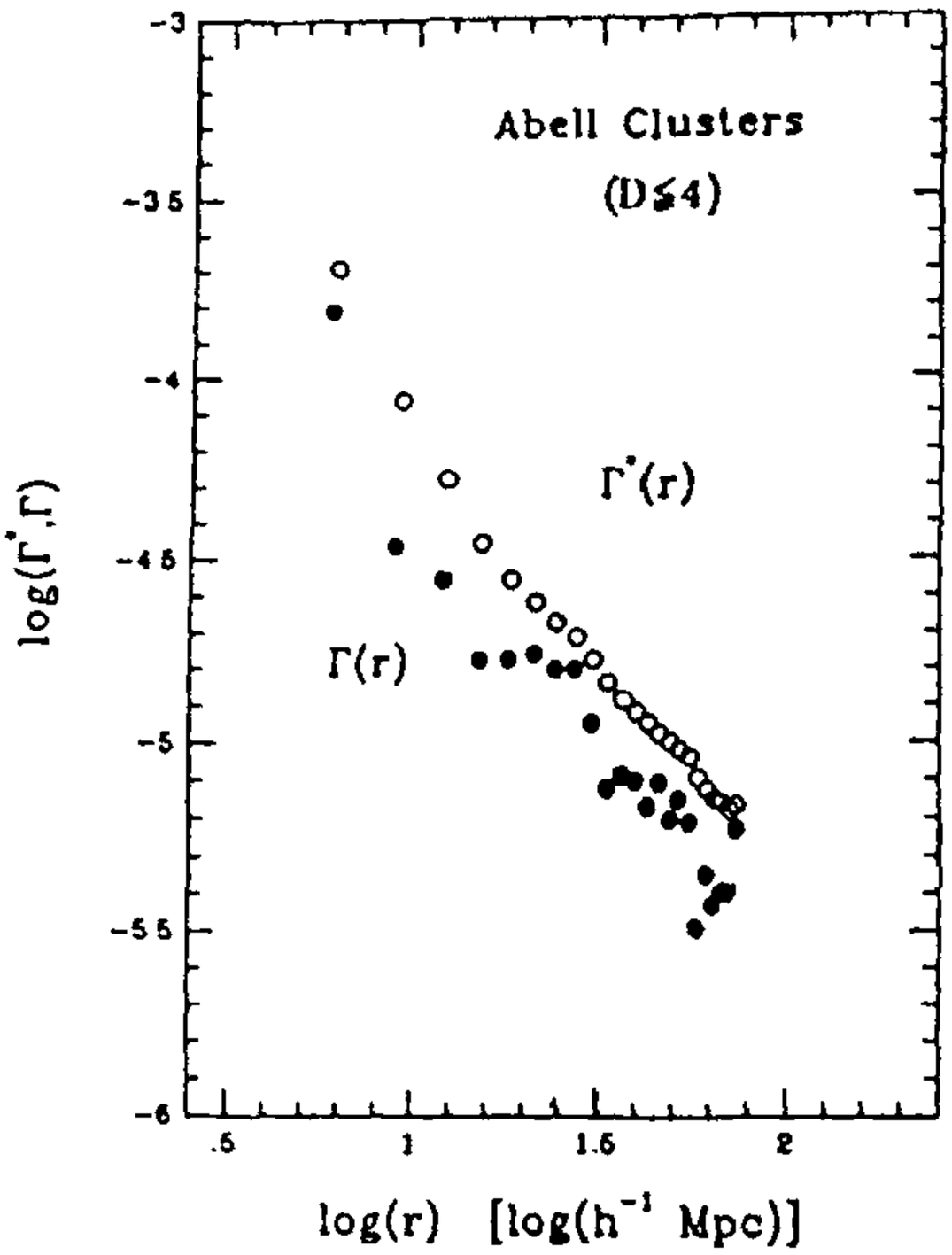

Figure 6. $\Gamma(r)$ and $\Gamma^*(r)$ for a sample of Abell survey

certainly deeper, but whose exact depth is not known due to lack of redshift measurements, indicate that the large-scale structure is uniform beyond a few 100 Mpc.

Visual impressions from maps have shown superclusters, voids, filaments and sheets in the distribution of galaxies on scales of a few tens to few 100 Mpc. Fractal geometry may provide a good description of these structures.

6. Graph theory deals with graphs, which are sets of points, some of them connected pairwise. Graphs without loops are called trees. All the points of a set may be interconnected in a single tree in many different ways. Each gives a certain length of the spanning tree. A spanning tree of the minimum length is a minimal spanning tree. The visual impression of filaments in a set of points can be quantified using minimal spanning trees of its subsets, pruning them, and rearranging in a random manner to see that filaments disappear. Thus, filaments have been shown to 'objectively' exist in galaxy surveys[18].

7. Topological studies are relevant since galaxies as members of (possibly weighted) point sets imply an underlying smoother mass density, certainly at the epoch of structure formation, perhaps before redshift 5 (i.e. $z \geq 5$). Isodensity surfaces of smoothed point sets with different smoothing scales and density levels (or threshold densities) may be profitably studied and compared with predictions of structure formation models, including the simplest possibility, viz. Gaussian (or random

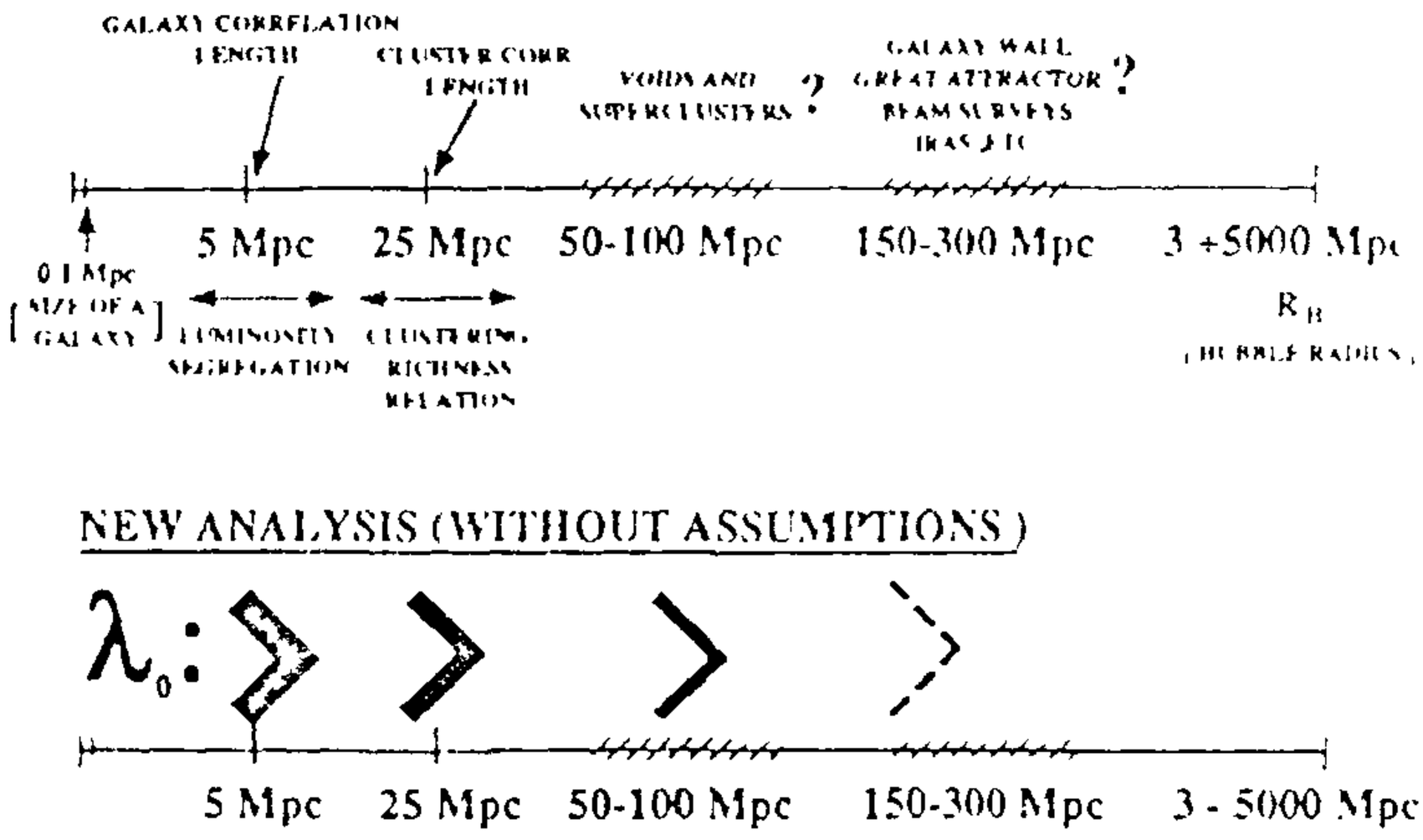

Figure 7.

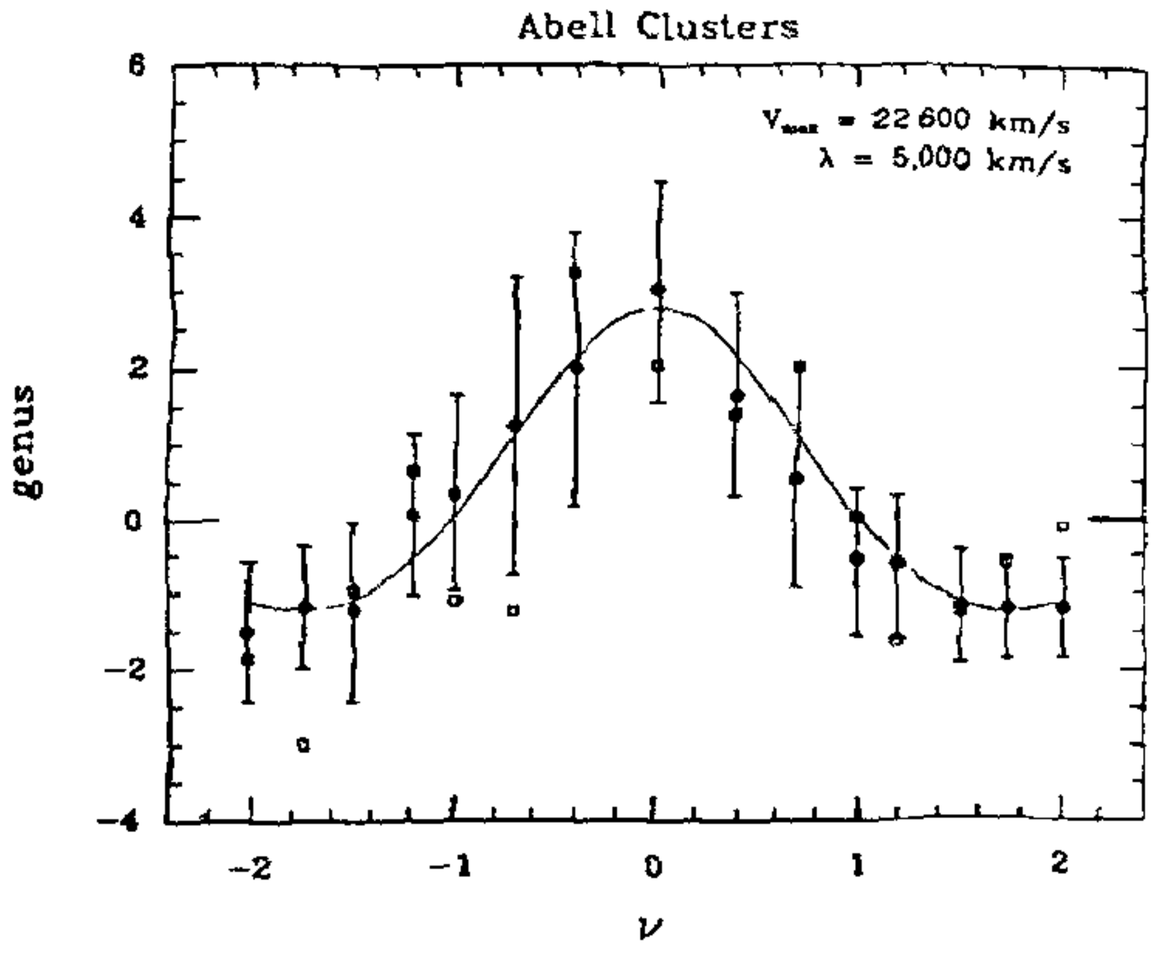

Figure 8. $g_s(\nu)$ for sample from Abell survey with $V_{max} = 22,600$ km s^{-1} and smoothing scale 5000 km s^{-1}

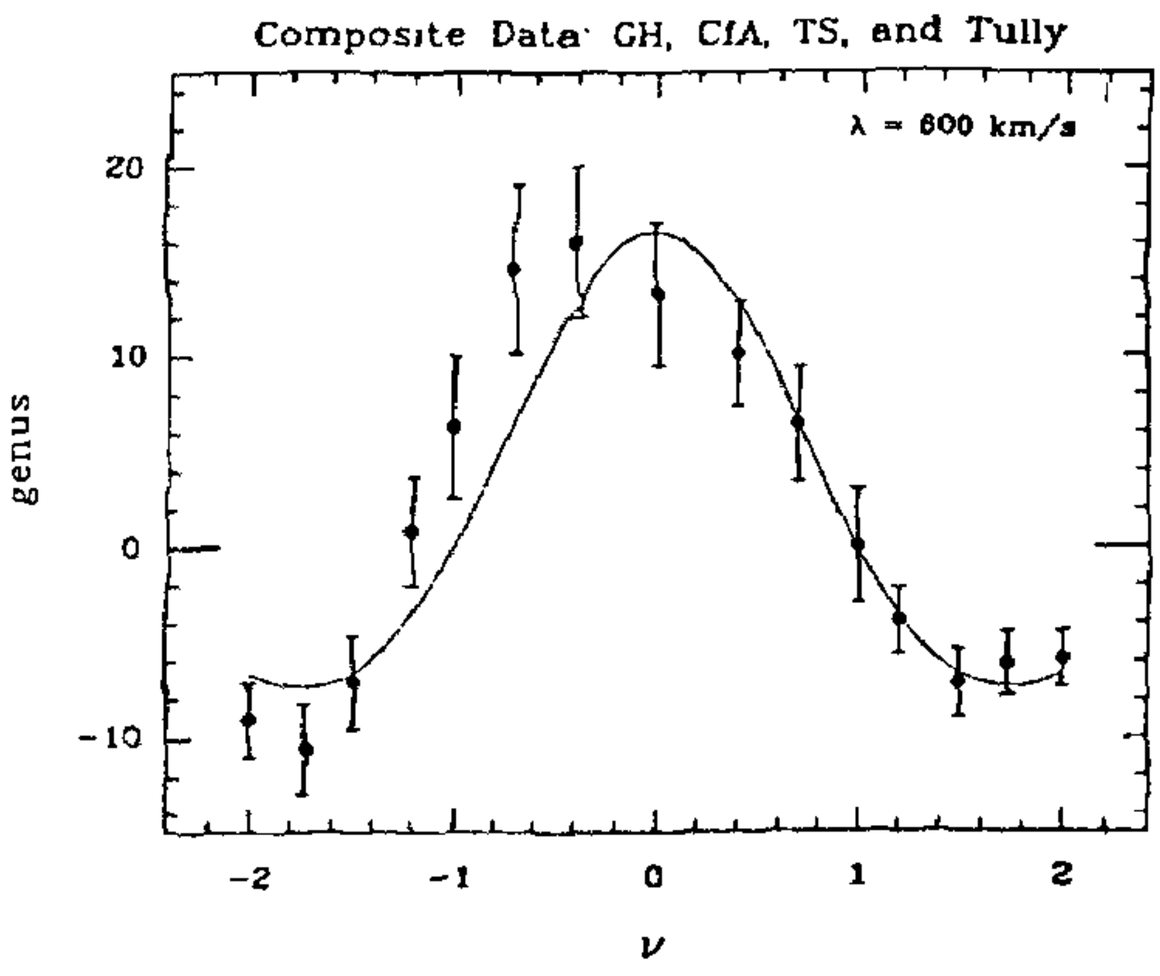

Figure 9. $g_s(\nu)$ for a composite sample

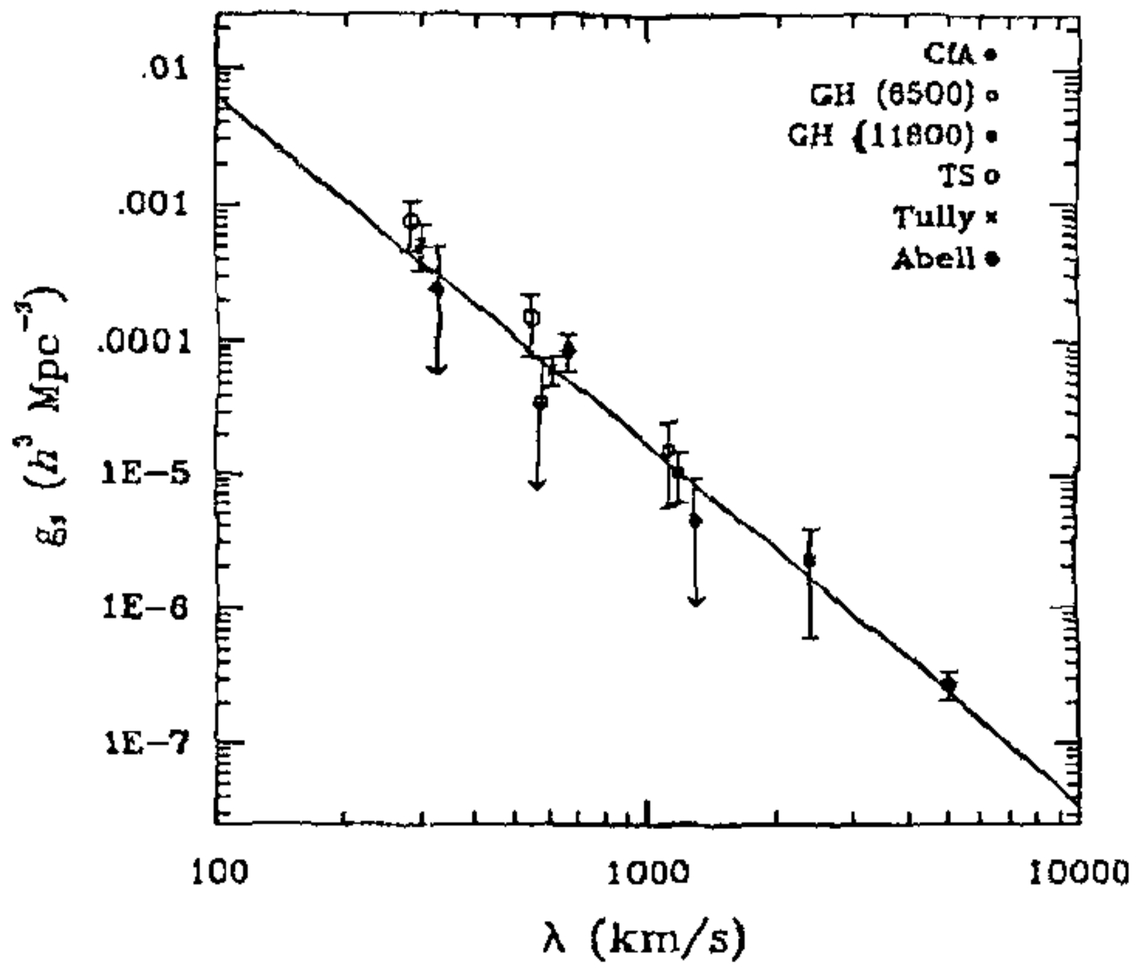

Figure 10. Amplitude of $g_s(\nu)$ for different smoothing lengths used in deriving $g_s(\nu)$

phase) distribution. The genus of the isodensity surface (*i* e. the number of holes it has) at a given density level (or threshold) may be calculated as an integral over the surface of its Gaussian curvature (by Gauss–Bonnet theorem). Superposing a grid of small cubes through the surface, one may, in practice, extract a topologically equivalent polyhedron for which this integral reduces to a sum of angle deficits (from 2π) over its many vertices. The diagnostic in this method is a plot of the genus per unit volume g_s against the density level expressed as the number of standard deviations ν from the mean density on both sides of the mean. Such a curve is shown for a sample from the Abell survey in Figure 8 (ref. 19) and for a composite of four samples in Figure 9 (ref. 19). The curves correspond to $g_s \propto (1 - \nu^2) \exp(-\nu^2/2)$ for the Gaussian (i.e. random phase) distribution. Within $|\nu| < 1$, the genus per unit volume is positive, indicating a sponge topology. For $\nu < -1$, there are voids or bubbles, while for $\nu > +1$, there are isolated clusters. Isolated voids surrounded by matter and isolated clumps of matter in empty space are the only two types of topology possible in two dimensions. In three dimensions, one can, in addition, have sponge topology, i.e. a connected matter region interlocking with a connected void region. A two-dimensional slice of a (three-dimensional) sponge shows isolated voids or isolated clumps. The amplitude of the $g_s(\nu)$ curve for galaxies decreases with increasing smoothing scales (Figure 10)[19], that is, the structure becomes less choppy (or smoother) as we average over larger volumes, in agreement with (examples of) results cited in sections (2) and (4) above.

8. Percolation study[20] of a point set is done by constructing a sphere of radius r around each point of the set, and increasing r, so that the sizes of connected groups of spheres increase, while, at the same time, the number of groups decreases. At the (critical) percolation radius r_c, all spheres connect in a single group. For galaxies it is found that the observed distribution has $r_c < r_p$, the percolation radius of a Poisson distribution having the same overall density of points. This means that there are denser clusters than expected of a uniform random distribution, and that these clusters are 'connected' across the volume sampled.

9. *Other methods* may be used to treat spatial, angular and even linear sky surveys. For example, a smoothed

spatial survey may be examined for the number of times a given line (i.e. direction) crosses the isodensity surface at a given level. This number (the contour crossing statistic) is proportional to the area of the isodensity contour, thus forming a connection between spatial and linear surveys[21]. Using topological methods to analyse the spatial fluctuations (i.e. anisotropies) in the 2.73 K cosmic microwave background, a model for the propagation of this light would give information on density fluctuations at the matter–radiation decoupling epoch without all the intermediate (currently poorly known/modelled) effects of structure formation.

A pancake detected?[22]

Since the visible universe contains 75% hydrogen, primordial structures are expected to show the λ21 cm neutral hydrogen signature line at the redshift when these structures formed. Two extreme possibilities are generally considered: merger of smaller structures to bigger and bigger ones (globular star clusters to galaxies to galaxy clusters and so on) or fragmentation of the biggest structures into smaller and smaller ones (superclusters to clusters to galaxies to stars). In the fragmentation picture, the biggest structures are expected to be highly flattened, called *pancakes*. In 1991, neutral hydrogen absorption and emission features within half a degree of each other were detected toward a giant radio galaxy of redshift 3.395 ± 0.005. The absorption feature at $z = 3.3968 \pm 0.0004$ has 270 ± 50 km s^{-1} width, corresponding to about $2.5 \times 10^{10}\, h^{-2}\, (T_s/10^4\ \mathrm{K})$ (size/arcsec^2) $M_\odot$ of absorbing material (where the Hubble constant H_0 is parametrized 100 h km s^{-1} Mpc^{-1} and T_s is the spin temperature). Small (proto)galaxies may be responsible for the absorption. The more interesting 5′ ($2h^{-1}$ Mpc) emission feature at $z = 3.3970 \pm 0.0003$ has 180 ± 40 km s^{-1} width, and is 33′ ($15h^{-1}$ Mpc) from the absorption feature. Its mass is estimated at $10^{14}\, h^{-2} M_\odot$, which tallies with the prediction for a Zel'dovich pancake[22]. The estimated surface mass density 48 $M_\odot$ pc^{-2} is 50% higher than the prediction. The narrow velocity width suggests that the system is yet to virialize. To be sure, this observation needs confirmation by other independent observers as well as by detection of more such structures.

Conclusion

In this article, I have presented a survey of the various methods of cosmography, with examples of results, using these methods, on the large-scale structure in the universe. The basic units are galaxies (0.1 Mpc), which are clustered relative to a Poisson distribution upto a few 100 Mpc, showing groups, clusters, superclusters, voids, filaments and sheets. Beyond that the distribution tends toward Poisson. Substantial large-scale motions (up to several 100 km s^{-1}) relative to the rest frame of the 2.73 K cosmic background radiation have been measured from almost all the surveys[23], but I have not touched this topic. Nor have I discussed in any detail models and their predictions, including the dynamical importance of dark matter. The material presented forms part of the background needed to discuss meaningfully the 10 ppm ripples COBE observed in the 2.73 K radiation in 1992.

1. Zwicky, F *et al*., *Catalogue of Galaxies and of Clusters of Galaxies*, 1961–1968, 6 Vols, California Institute of Technology, Nilson, P., *Uppsala General Catalogue of Galaxies*, Uppsala Astr Obs Ann., 1973, 6, Abell, G O., *Astrophys J Suppl*., 1958, **3**, 211–288, Huchra, J P., *et al*., *Astrophys. J*., 1983, **52**, 89–119, de Lapparent, V *et al*., *Astrophys J*., 1986, **302**, L1–L5, Hanes, M P and Giovanelli, R., *Large Scale Motions in the Unverse* (eds Rubin, V. C and Coyne, C V), Princeton University Press, 1988, pp 31–70, Haynes, M P. and Giovanelli, R., *Astrophys J*., 1986, **306**, L55–L59
2. Maddox, S J *et al*., *Mon Not R Astron Soc*., 1990, **242**, 43P–47P, Maddox, S J *et al*., *Mon Not R Astron Soc*., 1990, **243**, 692–712, Maddox, S J *et al*., *Mon Not R Astron Soc*., 1990, **246**, 433–457, Rowan-Robinson, M *et al*., *Mon Not R Astron Soc*., 1991, **253**, 485–495, Saunders, W *et al*., Oxford preprint OUTAS/92/4 for (*MNRAS*) 1992
3. Strauss, M A *et al*., *Astrophys J*., 1990, **361**, 49–62
4. Giacconi, R *et al*., *Astrophys J*., 1979, **230**, 540–550, Lahav, O *et al*., *Mon Not R Astron Soc*., 1989, **238**, 881–895, Micela, G *et al*., *Astrophys J*., 1991, **380**, 495–510
5. Webster, A., *Mon Not R Astron Soc*., 1976, **175**, 71–83, Webster, A., *Mon Not. R Astron Soc*., 1977, **179**, 511–515, Webster, A and Pearson, T J., *Mon Not R Astron Soc*., 1976, **179**, 517–519, Peebles, P J E., *The Large Scale Structure of the Universe*, Princeton, 1980 and references therein
6. Smoot, G F *et al*., *Astrophys J*., 1992, **396**, L1–L5, Wright, E L *et al*., *Astrophys J*., 1992, **396**, L13–L18
7. Carney, B W and Latham, D. W., *IAU Symposium 117*, 1987, 39, Tremaine, S and Lee, H M., in *Dark Matter in the Universe*, *World Scientific*, Singapore, 1987, p 103, Rubin, V C., in *After the First Three Minutes* (eds Holt, S S *et al*.) AIP, 1991
8. Webster, A., *IAU Symposium 74*, 1977, 75–84
9. Banhatti, D G., *Mon Not R Astron Soc*., 1990, **246**, 7P–10P
10. Haque-Copilah, S and Basu, D., *Int J. Mod Phys*., 1992, **D1**, 211–222
11. Huchra, J P and Geller, M J., *Astrophys J*., 1982, 257, 423–437, Geller, M J and Huchra, J P., *Astrophys J Suppl*., 1983, **52**, 61–87
12. Saslaw, W C., *Gravitational Physics of Stellar and Galactic Systems*, Cambridge University Press, 1985
13. Coleman, P H and Saslaw, W C., *Astrophys J*., 1990, **353**, 354–371
14. Falconer, K J., *Fractal Geometry*, Wiley, New York, 1990, Mandelbrot, B B., *The Fractal Geometry of Nature*, Freeman, 1983
15. de Vaucouleurs, G., *Publ. Astron Soc Pac*., 1971, **83**, 113–143
16. de Vaucouleurs, G., *Science*, 1970, **167**, 1203–1213
17. Coleman, P. H and Pietronero, L., *Phys Rep*., 1992, **213** 311–389
18. Bhavsar, S P. and Ling, E N., *Astrophys J*., 1988, **331**, L63–L68
19. Gott, J R *et al*., *Astrophys J*., 1989, **340**, 625–646
20. Shandarin, S F and Zel'dovich, Ya B., *Rev Mod Phys*., 1989, **61**, 185
21. Ryden, B S *et al*., *Astrophys J*., 1989, **340**, 647–660
22. Uson, J M *et al*., *Phys Rev Lett*., 1991, **67**, 3328

23 Ferris, T, *Sky Telesc*, 1987, **73**, 486, *Nature*, 1992, **356**, 657, *Science*, 1992, **257**, 1208, Lynden-Bell, D *et al*, *Astrophys J*, 1988, **326**, 19, Willick, J, *Astrophys J*, 1990, **351**, L5, Dressler, A and Faber, S, *Astrophys. J*, 1990, **354**, L45, Mathewson, D S *et al*, *Astrophys J*, 1992, **389**, L5

ACKNOWLEDGMENTS I thank N D Hari Dass and S Uma Sankar for arranging the One-Day Meeting and Institute of Mathematical Sciences for support and use of their facilities K R Anantharamaiah's comments helped improve the presentation

Received 25 May 1993; accepted 14 July 1993

X-ray crystal structure and computer modelling studies of HIV protease and its inhibitor complexes

Bangalore K. Sathyanarayana and Alexander Wlodawer

Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, BRP-ABL, P O Box B, Frederick, MD 21702, USA

HIV-1 protease is essential for the replication of HIV or the acquired immunodeficiency (AIDS) virus and is considered as an attractive target for the design of specific inhibitors. In order to design drugs which inhibit the action of HIV protease, it is essential to obtain the 3-D structures of these proteases. The native HIV protease and the very first inhibitor complex of the protease were studied at our laboratory and in this article we summarize the X-ray structure analysis and molecular modelling studies of the HIV-1 protease both in its native as well as with different inhibitor complexes studied both at our laboratory as well as laboratories elsewhere.

THIS article summarizes the results of X-ray crystallographic and computer modelling studies of the human immunodeficiency virus protease (HIV-PR) and its inhibitor complexes. These studies are quite promising and may lead to the design of drugs that could be therapeutically beneficial in cases of acquired immunodeficiency syndrome (AIDS). HIV-PR is an aspartic protease (PR), that is encoded by the human immunodeficiency virus (HIV). This enzyme is essential for proper maturation of the HIV virons, and if it is inactivated by either mutation or chemical inhibition, the assembled viral particles are not infectious. The techniques of rational drug design are being employed in many industrial and academic laboratories all over the world, in search of compounds that could accomplish such inactivation in the human body. However, for a rational approach to drug design to succeed, it is important to characterize the structures of the target enzyme and of enzyme–inhibitor complexes. Observations based on structural data, when used with other sources of information such as the results of investigations of binding constants and kinetics parameters, can lead to a better understanding of the enzyme–inhibitor interactions. Currently, extensive efforts have been devoted to the structure determination of HIV protease at various laboratories, and in this article, we summarize the work originating from our laboratory, as well as the results from other laboratories throughout the world.

HIV has been classified as a member of the lentivirus family[1] and from the array of processes required to sustain the viral life cycle, it may be possible that there are numerous points that could be exploited in development of drugs for AIDS therapy[2]. Figure 1 shows the HIV life cycle and the places of possible attack by the drugs at several stages. The cycle begins when HIV binds to the outside of the host cell and injects its core, which includes two identical strands of RNA and structural proteins and enzymes, all of which are needed at the later stages of the virus life cycle. Since HIV is a retrovirus with RNA as the source of its genetic information, instead of DNA as in most other viruses, it needs the help of two specific enzymes, namely polymerase and ribonuclease. These two enzymes form reverses transcriptase whose role is to make a second DNA copy using the first one as a template and to destroy the original RNA of the virus. RT actually causes the genes in the cell to make the proteins that the virus needs to reproduce. This is the first stage wherein the function of the RT can be halted by the intervention of any suitable drug. The genetic information of the virus in the regular DNA form is now carried on to the cell nucleus, where another viral enzyme called integrase (IN) will then splice the HIV genome into the host cell's DNA. This may well be considered as the second potential major stage of drug intervention against the virus. At the final stage, the