Halsey *et al.* 1986; Jones *et al.* 1988; Domínguez-Tenreiro and Martínez 1989). The "multifractal" or "multidimensional" formalism (Hentschel and Procaccia 1983) was first introduced to characterize some sets which appear in nonlinear physics and turbulence, such as strange attractors and aggregates. Those methods describe statistical time series or point sets in metric spaces, where the number of points N can tend to infinity and the time or space interval between two points, r, can be made to tend to zero. Jones *et al.* (1988) used this formalism as a clustering descriptor. Pietronero (1987) has suggested a different multifractal analysis, but his approach is based on a previous knowledge of the galaxy masses, and therefore it is not applicable to the present catalogs.

Galaxy sets are characterized by a finite value of both N and r, so the standard multifractal formalism has to be reformulated to circumvent the requirement $N \to \infty$ or $r \to 0$. In this paper we present such a reformulation. We test the techniques against specific clustering prescriptions. This gives us an appreciation of the scope and applicability of multidimensional analysis to clustered point sets.

The clustering paradigms we use are the continuum clustering hierarchy of Soneira and Peebles (1978), which is a homogeneous fractal, and the sample of vertices of a three-dimensional Voronoi tessellation (van de Weygaert and Icke 1989). We illustrate the formalism by means of an analytically calculable multifractal set. Then the formalism is applied to real world data.

This paper is organized in nine subsections. After presenting a subsample from the CfA catalog and both clustering models in §§ II and III, § IV contains the determination of the two-point correlation function for these sets. In § V we introduce a new, optimal, technique for determining the Hausdorff dimensions of the sets, based on their minimal spanning trees.

Although correlation functions of the data sets appear not to differ much, there are differences in their Hausdorff dimension which reflect the differences in the distributions. The multifractal formalism is described in § VI and illustrated in § VII. The differences revealed by the multifractal analysis are described in § VIII. General conclusions will be drawn in § IX.

II. THE CfA SUBSAMPLE

The CfA catalog is complete down to apparent magnitude $m = 14.5$ on the northern hemisphere for Galactic latitude $b^{II} \geq 40°$ and declination $\delta \geq 0°$. An equal area projection of the survey is shown in Figure 1. From the CfA catalog, we have selected a complete volume-limited subsample of 359 objects. The depth in terms of the radial velocity V_{max} is 6500 km s^{-1}. As the CfA catalog is complete down to apparent magnitude $m = 14.5$, the absolute limiting magnitude is $M_B^{max} = -19.56$. We have removed galaxies with radial velocity less than 1700 km s^{-1}. This is a representative volume-limited subsample drawn from the CfA catalog. We call it S65.

Radial velocities have been corrected for solar motion, Virgocentric flow and peculiar velocities due to galaxy clusters (Einasto *et al.* 1984). Apparent magnitudes have been corrected for galactic absorption. Calculations have been performed with a Hubble parameter $H_0 = 100$ km s^{-1} Mpc^{-1}.

III. THE CLUSTERING PARADIGMS

a) Model 1: Continuous Hierarchical Clustering

This model was proposed by Soneira and Peebles in 1978, and provides a three-dimensional point process having the same two-point correlation function as the observed one. The construction is as follows: we place randomly in a sphere of radius R the centers of η spheres of radius R/λ where $\lambda > 1$. In
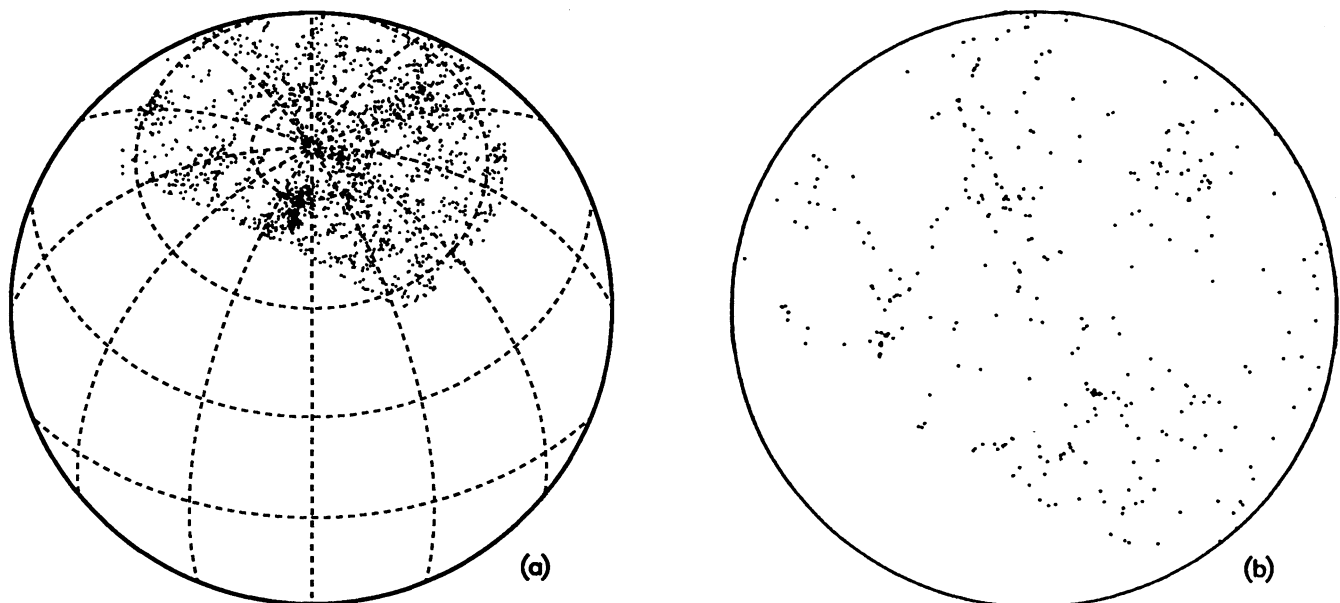
FIG. 1.—Equal-area Lambert's projection of the distribution of 2061 CfA catalog galaxies in the northern sky ($b^{II} \geq 40°$ and $\delta \geq 0°$). The north pole of the plot is the north Galactic pole. (*b*) Projection of the complete volume-limited subsample S65; 359 galaxies lie in the range between the velocity limits: $1700 \leq v \leq 6500$ km s^{-1} and with absolute magnitude $M_B \leq -19.56$. The Galactic pole is now at the center of the plot.
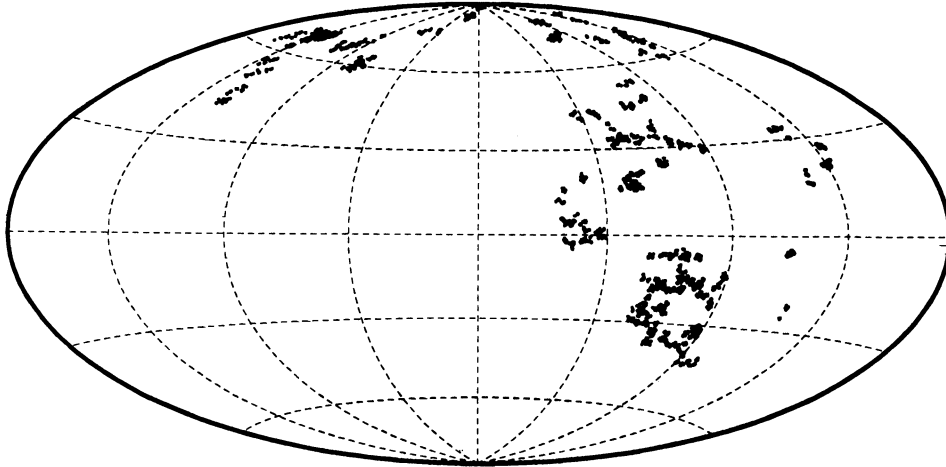
FIG. 2.—Equal-area Hammer's projection of one realization of a simple Soneira-Peebles model. This model is a homogeneous hierarchical fractal. 8192 points have been plotted.

each of these spheres we place now η new spheres with radius R/λ^2 and so on. Repeating this an infinite number of times yields a homogeneous fractal with Hausdorff dimension log $\eta/\log \lambda$. However, when we stop the construction at level L we have η^L positions where galaxies are placed. The two-point correlation function of the galaxies in this model is then given by

$$\xi(r) \sim r^{-\gamma} \tag{3}$$

with

$$\gamma = 3 - \left(\frac{\log \eta}{\log \lambda}\right) \text{ for } \frac{R}{\lambda^{L-1}} < r < R . \tag{4}$$

To illustrate what this model looks like we have selected the following parameters: $R = 50h^{-1}$ Mpc, $\eta = 2$, $L = 13$, and $\gamma = 1.77$, the observed slope of the correlation function. The value $\lambda = 1.76$ is determined by equation (4). The whole model contains 8192 galaxies. Figure 2 shows the equal-area Hammer's projection of this model. It should be noted that Soneira and Peebles used a superposition of the purely fractal models shown here, thereby creating a distribution that looks more realistic. In doing that they break the simple fractal nature of the clustering.

b) Model 2: Voronoi Tesselation Vertices

The Voronoi tessellation is a partitioning of space uniquely defined by a discrete point set. Each point of the set (*nucleus*) is surrounded by a Voronoi cell which encloses that part of space which is closer to its nucleus than to any other. In three-dimensional space, any realization of this process, called a "Voronoi foam" (Icke and van de Weygaert 1987) is built of three topologically distinct elements: *walls*, being the bisecting planes between two neighboring nuclei; *lines*, where three walls intersect; and *vertices*, where four lines come together. The vertices are the centers of the circumscribing spheres of the Delaunay tetrahedrons, whose packing is the Delaunay tessellation, the dual of the Voronoi tessellation. Each Delaunay tetrahedron consists of four nuclei whose circumscribing sphere does not contain any other nucleus. A stereoplot of a packing of three Voronoi cells is shown in Figure 3.

Icke and van de Weygaert (1987) and van de Weygaert and Icke (1989) used the Voronoi tessellation as a description of the geometrical skeleton of the distribution of galaxies in space, based on the notion by Icke (1984) that underdense regions will expand with respect to the background in such a way that they become more and more spherical. The Voronoi walls were identified with pancakes, the lines with filaments of galaxies,
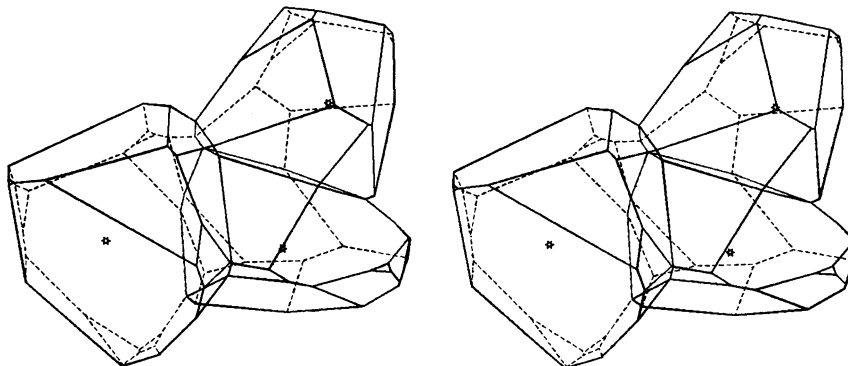
FIG. 3.—Stereoplot of a packing of three Voronoi cells

the vertices with clusters of galaxies and the interior of the cells with the voids in the galaxy distribution. One of the most outstanding results was that the two-point correlation function of the Voronoi vertices with Poissonian nuclei has a power-law form on scales smaller than the average cell size with a slope ≈ -2 and an amplitude completely in accordance with the observational cluster-cluster correlation function determined by Bahcall and Soneira (1983), a result which nearly all proposed models of large-scale structure formation have failed to reproduce.

Present knowledge about Voronoi tessellations stems from the pioneering works of Meyering (1953), Gilbert (1962), and Miles (1970). The Voronoi tessellation has acquired several alternative names: Dirichlet regions, Voronoi polygons/polyhedrons, Wigner-Seitz cells, Thiessen figures, and domains.

In this paper we will use the Voronoi vertices as a paradigm for point clustering. Note that here we use the Voronoi vertex distribution purely as an example of a stochastic point process, without the physical background that underlies the use of Voronoi tesselations as description of the large-scale structure in Icke and van de Weygaert (1987) and van de Weygaert and Icke (1989). The motivation for using the Voronoi vertices in this way is the convenient power-law clustering properties as expressed in their two-point correlation function. Moreover, to get a point process resembling the galaxy distribution as close as possible, the generating nuclei have a correlated rather than a Poisson distribution. The Voronoi vertex catalog used here consists of the 6001 vertices of a Voronoi tessellation of 996 cells defined by 996 clustered nuclei. As nuclei we have used the Voronoi vertices of a Voronoi tessellation resulting from 150 randomly distributed nuclei. The 6001 vertices are situated in a box of 100.0 Mpc × 100.0 Mpc × 100.0 Mpc. The Hammer projection of all vertices within a radius of 50.0 Mpc from the center of the box as seen from that center is shown in Figure 4; 3185 points have been projected.

IV. CORRELATION ANALYSIS

In Figure 5a we show the correlation function of the subsample of the CfA catalog. The dashed line corresponds to a subsample containing 50% of the galaxies randomly selected from the whole sample.

Log-log plots of the two-point correlation functions of the two stochastic models (Soneira-Peebles and Voronoi) are shown in Figures 5b and 5c. Both models have a power-law correlation function with a slope close to -1.77. Correlation functions were also computed for randomly selected subsamples having 10% of the original sample galaxies (*dashed lines*). The agreement between the subsamples shows the reliability of the correlation function determinations.

The two models as well as the CfA sample have correlation functions which are rather similar. Nevertheless, the two stochastic models have completely different point distributions as perceived by eye (Figs 2 and 4). We shall establish the mathematical reason for this discrepancy: the Soneira-Peebles hierarchical distribution is a homogeneous fractal, whereas the Voronoi model is not. Demonstrating this fact requires that we be able to determine the Hausdorff dimension of the point set.

V. THE HAUSDORFF DIMENSION

The Hausdorff dimension is defined for any set $\mathcal{A}$ by considering for each given value of $\epsilon > 0$ all possible coverings of $\mathcal{A}$ formed by domains with diameters $\epsilon_i \leq \epsilon$. Let us call this family of coverings $\Upsilon^\epsilon_{\mathcal{A}}$. We need first a well-defined measure of the set $\mathcal{A}$ (Falconer 1985). It depends on a real nonnegative parameter $\beta \geq 0$, and it is the so-called Hausdorff β-dimensional outer measure of $\mathcal{A}$:

$$H^\beta(\mathcal{A}) = \lim_{\epsilon \to 0} \inf_{\Upsilon^\epsilon_{\mathcal{A}}} \sum_i \epsilon_i^\beta . \tag{5}$$

Although it is always well defined, only if the set has uncountable support the β-dimensional outer measure can be different of zero. Now, the Hausdorff dimension of $\mathcal{A}$, $D_H(\mathcal{A})$ is defined as the exponent which produces the following behavior of the measure:

$$H^\beta(\mathcal{A}) = \infty \text{ if } \beta < D_H(\mathcal{A}) , \tag{6}$$

$$H^\beta(\mathcal{A}) = 0 \text{ if } \beta > D_H(\mathcal{A}) . \tag{7}$$

This definition (Hausdorff 1919) is not straightforward in its application to finite subsamples of fractal sets.
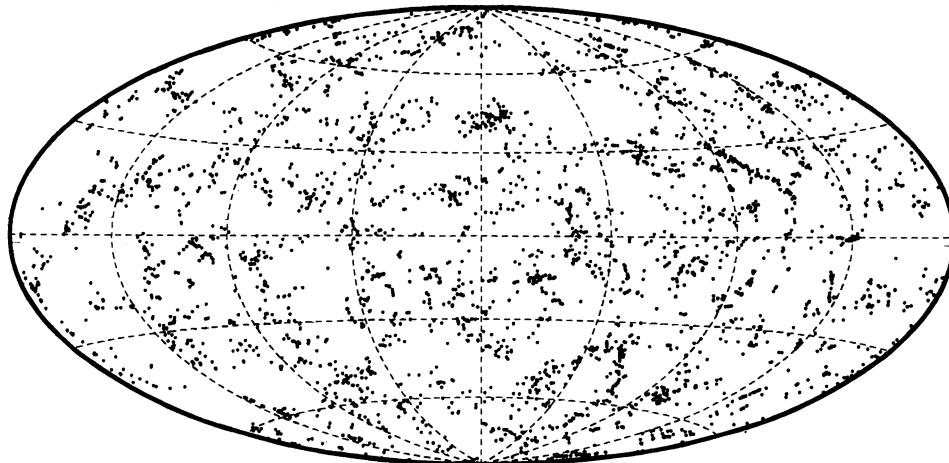

FIG. 4.—Equal-area Hammer's projection of the Voronoi vertices within a radius of $50h^{-1}$ Mpc; 3185 points have been plotted
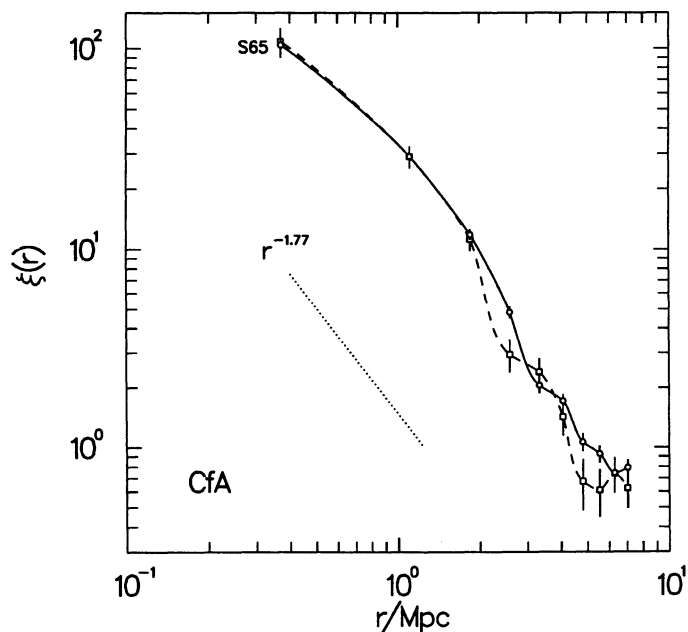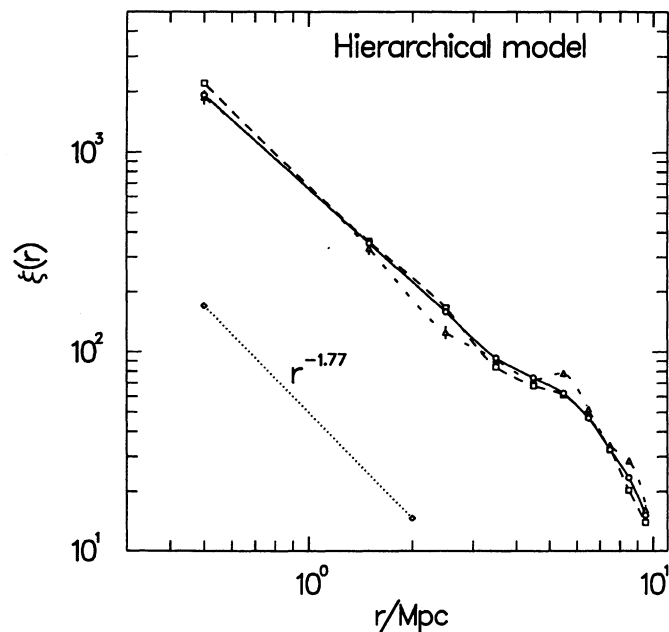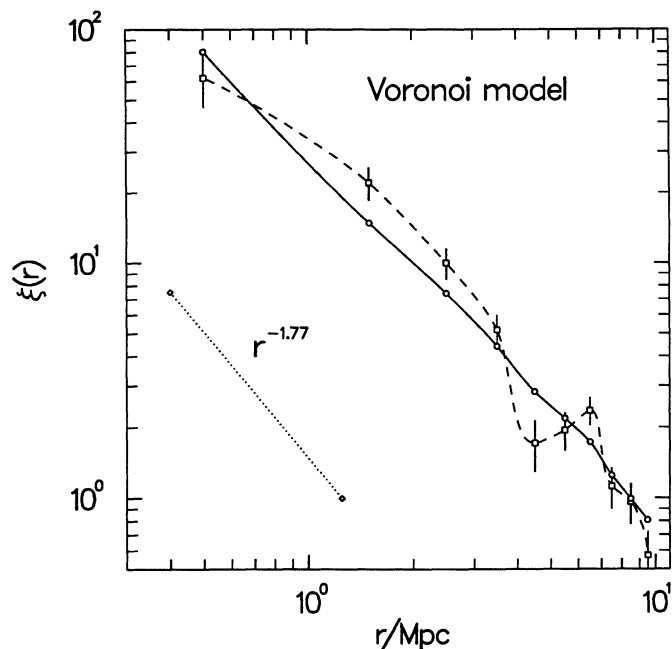
Fig. 5a

Fig. 5b

Fig. 5c

FIG. 5.—The spatial two-point correlation function: (*a*) for the CfA sample (*solid line*), and 50% randomly selected subsample (*dashed line*). (*b*) Soneira-Peebles model (*solid line*), 10% (*dashed line*), and 2% (*dotted line*). (*c*) Voronoi model (*solid line*), 10% (*dashed line*).

There are two features of importance in relation to this definition of the Hausdorff dimension: the fact that we need the infinum of all possible covering sets, and the fact that we need to take a limit as the size of the covering elements tends to zero. We want to have an estimation of the dimensionality of support of the point distribution using only the point set. Many methods of estimating dimensionality of point sets do not directly address this problem. Box-counting methods provide the "capacity dimension" which is frequently used as

an estimate of the Hausdorff dimension, even though examples are known of sets for which the Hausdorff and capacity dimensions are not the same.

The problem of identifying the covering set that gives the infinum is solved by using the distribution of the edge lengths of a construction called the "minimal spanning tree." The minimal spanning tree is a graph-theoretical construct that was introduced by Kruskal (1956) and Prim (1957). The minimal spanning tree of a set of N points is the *unique* network of $N-1$ edges (each linking two points) providing a route between any pair of points while minimizing the sum of the lengths of the edges. It defines a set of disks whose diameters are the edge lengths, and which when taken together provide the required covering. The covering is minimal in the sense of equation (5) and comes as close as possible to realizing the definition of the Hausdorff dimension for a discrete point set.

The minimal spanning tree was introduced into astronomy by Barrow, Bhavsar, and Sonoda (1985) to describe intrinsic patterns in the distribution of galaxies and was used by Bhavsar and Ling (1988) to argue that the filaments in the galaxy distribution are real and not merely due to chance alignments.

Formally stated in graph theoretical terms the minimal spanning tree can be formulated as follows:

The data set is a *graph G*, consisting of a *vertex set V* (the points) and *edge set E* (an edge is a straight line connecting two points; E is a subset of $V \times V$), each edge having a "length" or "weight" (in our case the Euclidean distance between the two vertices). A sequence of edges joining vertices is a *path*, a closed path is called a *circuit*. If there is a path between any pair of vertices the graph is called *connected*. A connected graph containing no circuits is called a *tree*. If the tree of a connected graph contains all the vertices of the vertex set then it is called a *spanning tree*. The length of a tree is defined to be the sum of the weights of the component edges. The *minimal spanning tree* (MST) is the spanning tree of minimal length.

We calculated the minimal spanning tree of the considered

point sets using the algorithm described by Prim (1957), in particular the implementation of Whitney (1972).

The algorithm starts by choosing an arbitrary point of the set and finding its nearest neighbor. These two points and the corresponding edge form the subtree T_1. For each *isolated point* (point not yet in the subtree) the identity and distance to its nearest neighbor within the subtree is stored; by definition this distance is called the distance of the isolated point to the subtree. These potential MST edges are called *links*. The *M*th subtree, T_M, is formed by adding to T_{M-1} the *nearest neighbor of the subtree*, being the isolated point whose distance to the subtree is minimal, together with the corresponding link. After adding this new point to the subtree the links are updated, i.e., the distance from each isolated point to the new subtree vertex is calculated to see whether it is smaller than the previous distance of the point to the subtree. The resulting T_{N-1} is the minimal spanning tree. If there are M points in the subtree the updating requires computation proportional to $N - M$, for M going from 1 to $N - 1$, so that the total computation time is proportional to $N(N - 1)/2$.

In Figure 6, we have plotted the minimal spanning tree of the galaxy subsample selected by Gott *et al.* (1986). The sample contains 161 galaxies in a cube with sides of $28.9h^{-1}$ Mpc, defined by the spherical triangle $A(\delta = 90°)$, $B(\delta = 0°, \alpha = 9^h49^m)$ and $C(\delta = 0°, \alpha = 15^h49^m)$. In order to have a volume-limited complete subsample only galaxies brighter than $M_B = -18.99$ have been included. In the left plot all the galaxies in the sample are plotted, while the right plot shows all the MST links between the galaxies in this sample.

Determining the Hausdorff dimension using the set of minimal spanning tree edge lengths $\{l_i\}_{i=1}^m$, in which $m = N_R - 1$ and N_R is the number of randomly selected points from the sample, is related to the method given by Badii and Politi (1984) by using the nearest neighbor distances:

$$H^\beta(\mathscr{A}) \equiv \sum_{i=1}^m l_i^\beta(m) = K(\beta)m^{1-\beta/h(\beta)} , \qquad (8)$$

The fixed point of the function $h(\beta)$, $h(\tilde{D}_H) = \tilde{D}_H$ is a good estimator of the Hausdorff dimension.

The correlation dimension D_2 can be formally defined as

$$D_2 = \lim_{r \to 0} \frac{\log C(r)}{\log r} , \qquad (9)$$

where $C(r)$ is the correlation integral (Grassberger and Procaccia 1983) defined by

$$C(r) = \int_0^r 4\pi s^2[1 + \xi(s)]ds . \qquad (10)$$

In a finite sample of a homogeneous fractal D_2 and D_H are equal. Therefore the slope γ of the correlation function is related to the Hausdorff dimension D_H of the fractal by $D_H \approx 3 - \gamma$.

In the CfA sample the correlation dimension is $D_2(\text{CfA}) \simeq 1.3 \pm 0.1$, while the Hausdorff dimension calculated from the minimal spanning tree is $D_H(\text{CfA}) \simeq 2.1 \pm 0.1$ (Martínez and Jones 1990).

For the Soneira-Peebles hierarchical model both values are nearly equal: $D_H \simeq D_2 \simeq 1.2 \pm 0.1$. In the Voronoi model there is a clear difference between the two: D_2 (Voronoi) $\simeq 1.4 \pm 0.1$ and D_H (Voronoi) $\simeq 2.0 \pm 0.1$. The fact that these two values are different indicates that the distribution is not well represented by a simple fractal. A generalization of this concept, the multifractal, is needed to characterize such kind of sets.

VI. THE MULTIFRACTAL FORMALISM

The multifractal formalism was first formulated to characterize nonlinear phenomena, such as dynamical systems, strange attractors and turbulence (Mandelbrot, 1974, 1982; Frisch and Parisi 1985; Jensen *et al.* 1985; Halsey *et al.* 1986). Jones *et al.* (1988) initiated the application of this method to describe the galaxy distribution. In that paper they also compared their results for the real data with some cold dark matter N-body simulations.

Consider a set $\mathscr{A}$ embedded in a Euclidian space R^N. For each $r > 0$ we can consider all possible coverings of $\mathscr{A}$ formed by sets with diameters $r_i \le r$; $\Upsilon^r_{\mathscr{A}}$. For any general measure μ we can calculate its counterpart in each of the sets of the covering:

$$\mu_i = \int_{i-\text{set}} d\mu , \qquad (11)$$

Now we can define the following general partition sum:

$$\Gamma(q, \tau) = \lim_{r \to 0} \inf_{\Upsilon^r_{\mathscr{A}}} \sum_i \frac{\mu_i^q}{r_i^\tau} . \qquad (12)$$
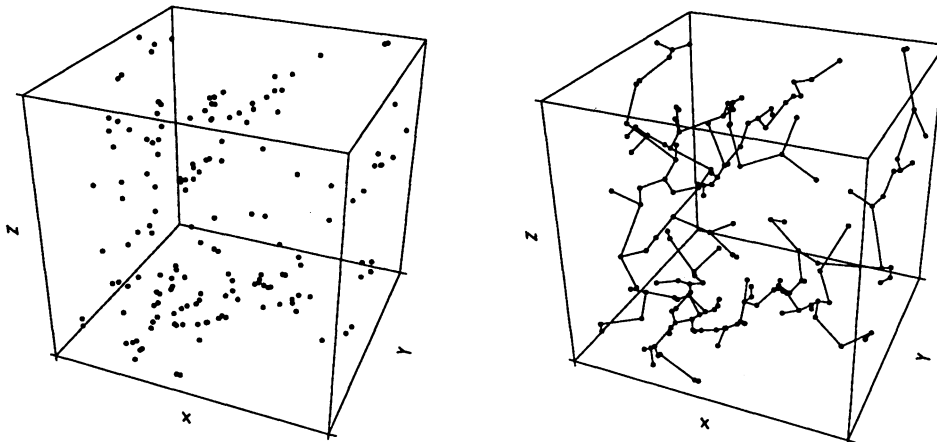

Fig. 6.—The minimal spanning tree of a cubical complete subsample of the CfA catalog. The length of the side is $28.9h^{-1}$ Mpc. Our Galaxy is at the bottom left corner.

Then a new function $\tau(q)$ is defined by the requirement

$$\Gamma(q, \tau) = \text{constant} . \qquad (13)$$

It is easily seen that, for $q = 0$, equations (12) and (13) are exactly the definition of the Hausdorff dimension (see eqs. [5]–[7] above), $D_H = D_0 = -\tau(0)$. In general, the set of numbers defined by

$$D_q = (q - 1)^{-1} \tau(q) \qquad (14)$$

forms a family of relevant dimensions which have been used to characterize many physical systems (Renyi 1970; Hentschel and Procaccia 1983; Halsey *et al.* 1986). Because D_q is a generalization of the concept of dimension provided by Hausdorff in 1919 to sets that are clearly not homogeneous fractals, these quantities are called *generalized dimensions*. The dimensions D_1 and D_2 are called the "information" and "correlation" dimension.

Because we are dealing with sets containing a finite number of points, so that limits $r \to 0$ cannot be taken, we have to resort to approximating methods. Three methods are as follows.

Algorithm 1: Box-counting.—Consider a point set G which represents the positions of a sample of galaxies in three-dimensional Euclidean space. In this method the assumption (or simplification) is that all the domains forming the covering of G have the same size $r_i = r$, leading to a partition of size r of the volume in which the N galaxies are embedded. Then a probability measure is assigned to each cell:

$$\bar{p}_j(r) = \frac{\bar{n}_j(r)}{N} , \qquad (15)$$

where the subscript j labels the cells and $\bar{n}_j(r)$ is the total number of galaxies in the cell j. We are interested in the behavior of the partition sum,

$$Z(q, r) = \sum_{j=1}^{N_{\text{cell}}} \bar{p}_j(r)^q , \qquad (16)$$

where N_{cell} is the number of occupied cells.

A new variable τ is introduced as the exponent of the scale factor in the following way,

$$\Gamma_1(q, \tau, r) = r^{-\tau} Z(q, r) = r^{-\tau} \sum_{j=1}^{N_{\text{cell}}} \bar{p}_j(r)^q . \qquad (17)$$

This function behaves in the limit $r \to 0$ as

$$\Gamma(q, \tau) = \lim_{r \to 0} \Gamma_1(q, \tau, r) = \begin{cases} \infty & \text{if } \tau > \tau(q) \\ \text{constant} & \text{if } \tau = \tau(q) . \\ 0 & \text{if } \tau < \tau(q) \end{cases} \qquad (18)$$

This defines a new function $\tau(q)$ for which the function Γ is finite. The variable $\tau(q)$ is related to the generalized dimensions D_q (eq. [14]).

Algorithm 2: Correlation and density reconstruction method.—In the box-counting method only a small interval of r-values is significant when one computes the partition sum (16), due to the discreteness of the sample. If r is too small there is only one galaxy per cell; if r is too large, the number of galaxies per cell is rather constant and the statistics are very poor. These considerations lead one to consider other estimators of the partition sum (12). One method is the correlation

algorithm (Grassberger and Procaccia 1983). For each galaxy labeled by i, we count the number of points $n_i(r)$ lying within a sphere of radius r centered on galaxy i. We can define a measure at scale r by

$$p_i(r) = \frac{n_i(r)}{N} = \frac{1}{N} \sum_{j=1}^{N} \Theta(|r_i - r_j| - r) , \qquad (19)$$

with N the total number of galaxies in the sample, r_i and r_j the positions of galaxies i and j, respectively, and Θ the Heaviside step function. This set of values p_i represents a sample of the underlying fractal measure. The partition sum is now

$$Z(q, r) = \frac{1}{N} \sum_{i=1}^{N} p_i(r)^{q-1} . \qquad (20)$$

If these moments behave as a power law of scale r we say that the measure is multifractal:

$$Z(q, r) = r^{\tau(q)} \times \text{const} , \qquad (21)$$

in which the constant depends only on q; consequently the function $\tau(q)$ is obtained from

$$\tau(q) = \frac{d \log Z(q, r)}{d \log r} . \qquad (22)$$

Equivalently one can obtain the function $\tau(q)$ by fitting

$$W(\tau, p) = \frac{1}{N} \sum_{i=1}^{N} r_i(p)^{-\tau} = p^{1-q} \times \text{const} \qquad (23)$$

where $r_i(p)$ is the radius of the smallest sphere centered at galaxy i whose probability is p. Equation (21) works better for $q > 1$, whereas equation (23) converges quickly for $q \leq 1$ (Grassberger, Badii, and Politi 1988). This method is rather efficient, and it is the one used in § VIII.

Algorithm 3: Minimal spanning tree.—The minimal spanning tree with m edges produces a set $\{l_i\}_{i=1}^{m}$ of edge lengths, with $m = N_R - 1$; N_R is the number of galaxies in a subsample formed by random selection from the whole sample. Now, we define the partition function by

$$S(\tau, m) = \frac{1}{m} \sum_{i=1}^{m} l_i(m)^{-\tau} . \qquad (24)$$

If this sum behaves as a power law of the number of edges m in the subsample we have

$$S(\tau, m) = m^{q-1} \times \text{const} , \qquad (25)$$

with the constant depending only on τ. By fitting this scaling relation for different values of $m(m + 1$ points selected randomly from the whole sample) we obtain q as a function of τ. Notice that m has to be large enough for the corresponding MST branches to be small enough to be within the scaling region.

This method is promising because it remains close to the original definitions as expressed in equation (12). The three algorithms explained above give consistent results in those cases where, all of them can be applied (van de Weygaert, Jones, and Martínez 1990; Domínguez-Tenreiro, Roy, and Martínez 1990).

The $f(\alpha)$ curve.—Alternatively, we can write the fractal measure (19) as a power law of the scale. These scaling indices or point-wise dimensions (Grassberger *et al.* 1988) are defined by the equation

$$p_i(r) = r^{\alpha_i} , \qquad (26)$$

The distribution of the scaling indices characterizes the dimensionality of the set $\mathscr{A}$. The distribution is quantified by means of the α-spectrum, $f(\alpha)$, defined by the *Ansatz*,

$$n(\alpha)d\alpha \sim N|\ln r|^{1/2}r^{\alpha - f(\alpha)}\,d\alpha \, , \qquad (27)$$

where $n(\alpha)d\alpha$ is the number of times that α takes values in the interval $(\alpha, \alpha + d\alpha)$. The Hausdorff dimension of the set of points that have the same value for the scaling index α is $f(\alpha)$. In a homogeneous fractal the curve $f(\alpha)$ degenerates into a single point: $\alpha_0 = f(\alpha_0) = D_0$. The statistical properties of a broad class of point sets are equivalently described by either the $f(\alpha)$ function or by the generalized dimensions D_q. To make this evident, let us consider the partition sum as a continuous integral,

$$Z(q, r) = \frac{1}{N}\sum_{i=1}^{N}p_i(r)^{q-1} = \frac{1}{N}\int n(\alpha)r^{(q-1)\alpha}\,d\alpha$$

$$= \int |\ln r|^{1/2}r^{\alpha q - f(\alpha)}\,d\alpha \, . \qquad (28)$$

The last integral can be solved by using the Laplace integral approximation theorem (Copson 1967):

$$\int |\ln r|^{1/2}r^{\alpha q - f(\alpha)}\,d\alpha = r^{\alpha(q)q - f[\alpha(q)]}\left\{\frac{\pi}{2f''[\alpha(q)]}\right\}^{1/2} \qquad (29)$$

the conditions of this theorem defining the function $\alpha(q)$ by

$$\left.\frac{df(\alpha')}{d\alpha'}\right|_{\alpha' = \alpha(q)} = q \qquad (30)$$

and

$$\frac{d^2f(\alpha')}{d\alpha'^2} < 0 \, . \qquad (31)$$

Combining equations (21) and (29) gives

$$\tau(q) = \alpha q - f(\alpha) \, , \qquad (32)$$

while equations (30) and (32) lead to

$$\alpha(q) = \frac{d\tau}{dq} \, . \qquad (33)$$

These last two equations relate the pairs (q, τ) and (α, f) through a Legendre transformation, and consequently the information on the set $\mathscr{A}$ is described equivalently by both.

The function $f(\alpha)$ is convex and has a single maximum at α_0, which is the most frequent value of the scaling indices. Equations (30) and (32) tell us that this occurs at $q = 0$, where $f(\alpha_0) = -\tau(0) = D_0$. Also the scaling indices take values in a finite range $[\alpha_{min}, \alpha_{max}]$:

$$\alpha_{min} = \lim_{q \to \infty} D_q \, , \qquad (34)$$

$$\alpha_{max} = \lim_{q \to -\infty} D_q \, . \qquad (35)$$

Important information about the degree of inhomogeneity of the point set can be obtained from the second derivate of $f(\alpha)$.

$$\frac{d^2f}{d\alpha^2} = \left(\frac{d^2\tau}{dq^2}\right)^{-1} = \frac{1}{\tau''(q)} \, . \qquad (36)$$

The curvature of $\tau(q)$ is a measure of its departure from the

linear behavior expected for homogeneous fractals; if $\tau''(q) = 0$, the set is a simple fractal, and the value of $|\tau''(q)|$ measures the richness of different structures observed in a multifractal set.

The $f(\alpha)$ spectrum can be used to fully characterize the structural features of the distribution of galaxies. In the next section we will present several realizations of a random multiplicative multifractal to illustrate the interpretation of the $f(\alpha)$ spectrum.

Once we have $\tau(q)$ we can obtain $f(\alpha)$ by simply taking the Legendre transform. Note that while D_q is mathematically equivalent to $f(\alpha)$, the numerical evaluation of $f(\alpha)$ involves an additional differentiation of the data and so is inherently less well determined unless the sample is large enough that D_q itself is well determined.

VII. THE MULTIFRACTAL SPECTRUM ILLUSTRATED

To give a feeling of the discriminating power of the $f(\alpha)$ spectrum we show some realizations of a hierarchical clustering model in two dimensions. The model we use is of a kind discussed frequently by Meakin (1987) and has the virtue of being analytically calculable.

The hierarchy is constructed by dividing a square into four equal square pieces. We assign four numbers, p_1, p_2, p_3, and p_4, $p_i \in [0, 1]$ to each of the subsquares. Each subsquare is then subdivided in the same way as the parent, and the subdivisions are randomly assigned the same four numbers p_i. The value attached to a square is the product of the p_i's of the square, its parent and all its other ancestors. Going to L levels of the hierarchy produces a $2^L \times 2^L$ array of numbers that are each products of L of the p_i. The distribution of cell values therefore depends on the initial choice of the p_i values. It can be shown that in the limit $L \to \infty$

$$D_q = (1 - q)^{-1}\log_2 (f_1^q + f_2^q + f_3^q + f_4^q) \, , \qquad (37)$$

where

$$f_i = \frac{p_i}{\sum_{i=1}^{4}p_i} \, . \qquad (38)$$

In our examples the process is continued until a lattice with 512×512 cells has been constructed. The point process realization is generated by distributing 10^4 particles on the lattice with probabilities that are simply proportional to the cell value. Because the random numbers are generated multiplicatively the process can be referred to as a "multiplicative random process." It is manifestly non-Gaussian.

For the purposes of illustration we have chosen three models with p_i values showed in Table 1. In Figure 7 we have plotted these models with their $f(\alpha)$ curves derived by box-counting method (*solid line*). We have also shown the theoretical curves derived from equation (37) through a Legendre transformation (*dashed line*). Model I is clearly a homogeneous fractal with dimension $D_q = D_0 = \log 3/\log 2$ for any q. In this case the $f(\alpha)$ curve degenerates into a single point. Instead, models II and III are multifractal sets (inhomogeneous fractals). Notice that

TABLE 1

PARAMETERS OF THE MULTIPLICATIVE
RANDOM MODELS

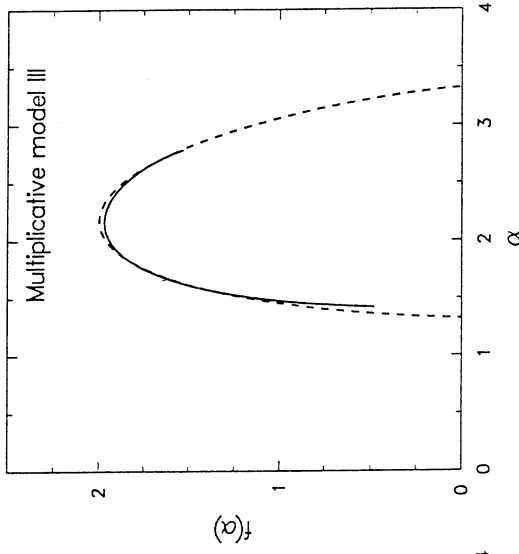| Model	p_1	p_2	p_3	p_4
I	1	1	1	0
II	1	1	0.5	0.5
III	1	0.75	0.5	0.25
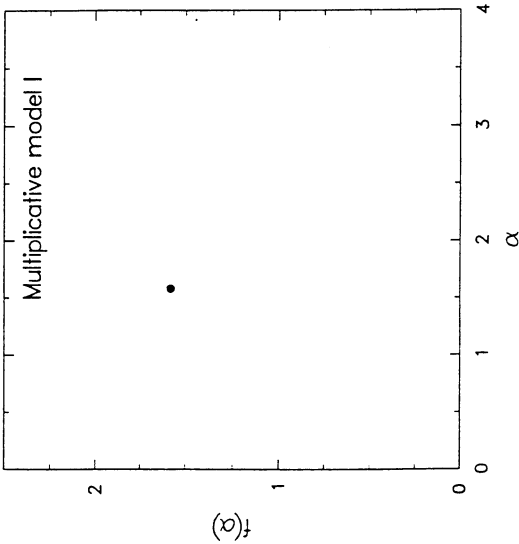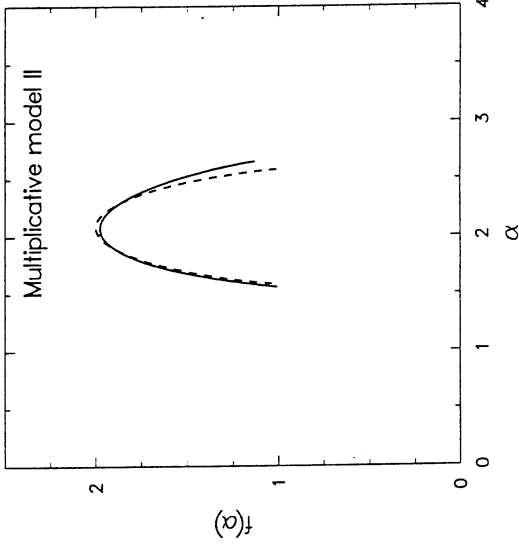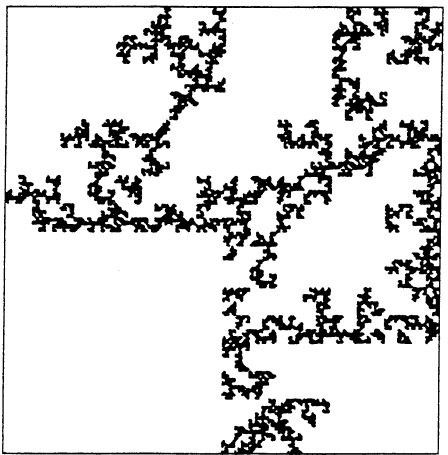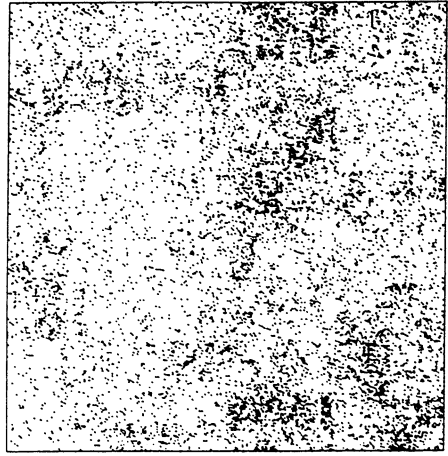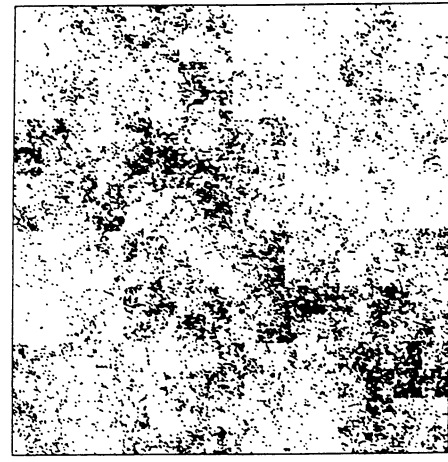
FIG. 7.—Three realizations of the multiplicative random fractal with the corresponding $f(\alpha)$ curve. Model I is a homogeneous fractal and $f(\alpha)$ degenerates into a single point, while models II and III are multifractals. Increasing the inhomogeneity, the range of α values also increases. Solid lines are the computational $f(\alpha)$ curves calculated from the realizations, and dashed lines are the expected theoretical curves.

the degree of nonuniformity has been increased from model II to model III, and therefore the spread in the $f(\alpha)$ is wider in III than in II.

VIII. THE MULTIFRACTAL ANALYSIS

In this section we present the results of the calculations of the $f(\alpha)$ spectrum for the two stochastic models and for the real data. Calculations for these models have been performed for the whole model ($\sim 10^3$–10^4 points), and for a subsample taken from the whole by randomly selecting 10% of the points. This has been done in order to check the stability of the statistical descriptors when having only relatively small data sets, which is the case with the available data in redshift space.

a) The CfA Data

We have performed the multifractal analysis for the real data. The partition sum (see eq. [20]) has been calculated for the subsample S65 described in § II. In Figure 8 we show log $Z(q, r)$ as a function of log r (see eq. [22]) for $q = 2, 3, 4, 5$, $\tau(q)$ has been calculated as the slope of the regression lines over the range $[r_1, r_2]$, where r_1 is the mean nearest neighbor distance and r_2 is the mean interparticle distance $(V/N)^{1/3}$. The fact that there may be no scaling outside of this range may be simply a finite sample effect or it could be a fingerprint of the breakdown in the fractal range (Martínez and Jones 1990). The distribution is dominated by discreteness noise on the smallest scales, while the distribution looks Poisson-like on the larger scales. As the sample size increases, we would expect to see an amplification of the scaling range if the galaxies did indeed follow the kind of generalized fractal scaling law suggested here.

For $q \leq 1$ the scaling has been found by fitting equation (23) for small values of the probabilities p, ($p \in [0.01, 0.1]$). In Figure 9 we show how to get the fractal dimension $D_0 = -\tau(0)$
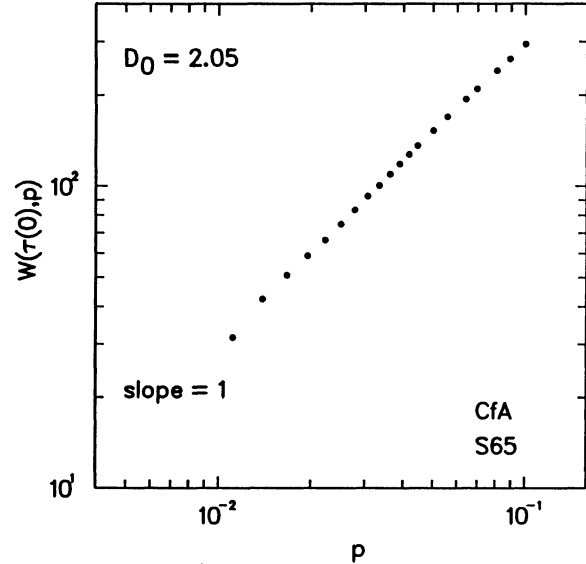
FIG. 9.—The fractal dimension D_0 of the CfA subsample obtained from the scaling of the partition sum $W[\tau(q), p]$ vs. p when $q = 0$.

for the data sample by using this procedure. The $f(\alpha)$ curve for the CfA sample is shown in Figure 10, and following the discussion so far, we would conclude that the galaxy distribution is well represented by a multifractal. It is interesting to remark that the left branch of the $f(\alpha)$ curve remains always stable for different complete volume-limited subsamples drawn from the CfA. The right branch is not so universal due to the different distribution of the low-density regions (voids) in each subsample.

b) Clustering Models

In Figure 11 we show the D_q curve for the Soneira-Peebles clustering model. (*solid line*: whole model; *dashed line*: 10% random subsample, *dotted line*: 2% random sample). The model is a homogeneous fractal and the D_q curve is virtually independent of q, as would be expected.
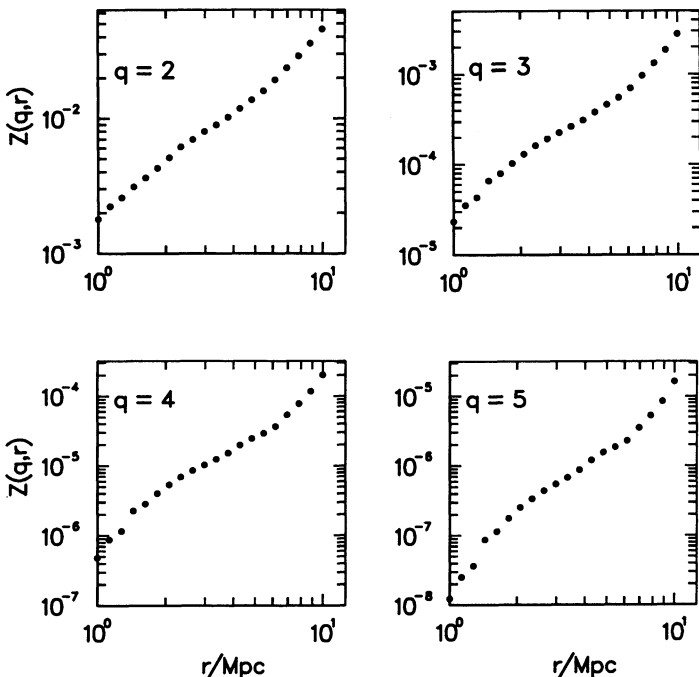
FIG. 8.—The scaling behavior of the partition sum $Z(q, r)$ of the CfA subsample for different values of q. The slope in the scaling region is $\tau(q)$.
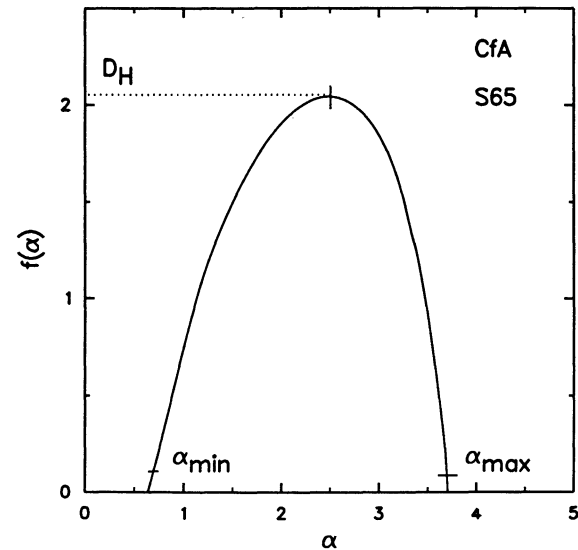
FIG. 10.—The $f(\alpha)$ multifractal spectrum of scaling indices for the CfA subsample. The width of $f(\alpha)$ shows that the distribution is a complex mixture of subsets corresponding to many different scaling indices.
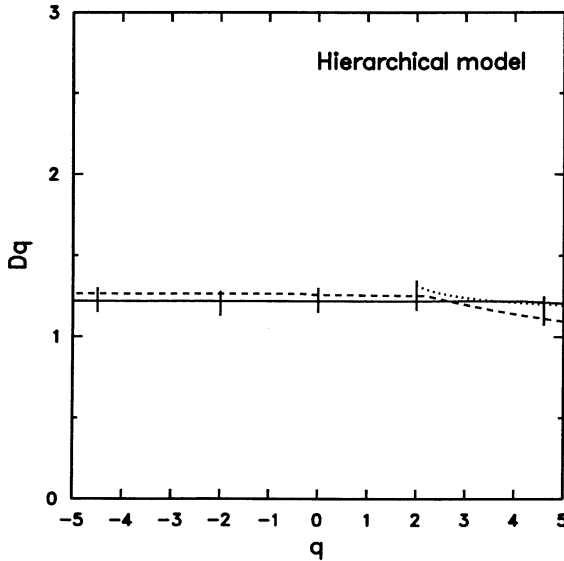
FIG. 11.—D_q curve for the Soneira-Peebles hierarchical model. It is a simple fractal, and therefore the curve remains roughly constant for any q. The $f(\alpha)$ will degenerate into the point, $\alpha = f(\alpha) \approx 1.23$. The solid line corresponds to the whole model; dashed and dotted lines correspond to random selection of 10% and 2% of the particles.

In this model the density scales as a power law of the distance (Peebles 1980) $n(r) \sim r^{-\gamma}$. The measure (19) is $p_i(r) \sim r^{3-\gamma}$ and therefore the partition sum (20) behaves as

$$Z(q, r) = \frac{1}{N} \sum_{i=1}^{N} p_i(r)^{q-1} \sim r^{(q-1)(3-\gamma)} \sim r^{\tau(q)} . \qquad (39)$$

The value of $D_q = 3 - \gamma$ is around 1.23, as would be expected from the input parameters. (The hierarchy was fixed so as to give the correct slope to the two-point correlation function). The $f(\alpha)$ curve degenerates to a point.

For the Voronoi model, we show in Figure 12 the $f(\alpha)$ curves for the whole sample (*solid line*) and a random subsample with
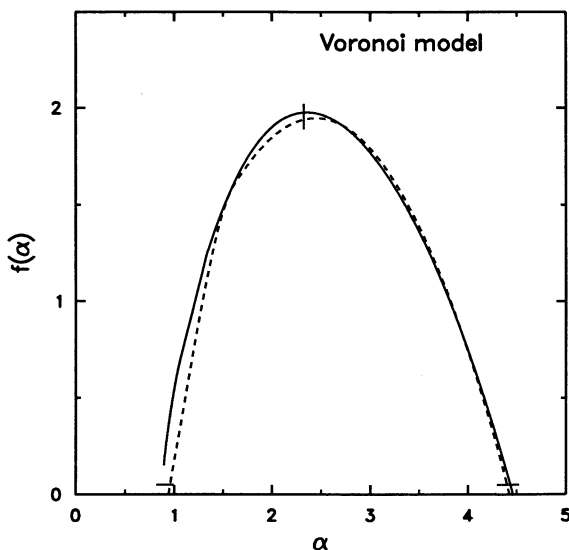

FIG. 12.—The $f(\alpha)$ curve of the Voronoi model. It is a multifractal, as the real galaxy distribution. The agreement between the solid line (whole model) and the dashed line (10% random sample) shows the stability of this clustering descriptor.

10% of the points (*dashed line*). The agreement between the curves etablishes our ability, for this sample, to derive a meaningful $f(\alpha)$ curve even from small samples. The model is clearly not a homogeneous fractal, as we would have expected since we have already found that the correlation and Hausdorff dimension are different for this set. The differences between Figures 10 and 12 [right branch of the $f(\alpha)$ curve] show how the Voronoi model and the CfA catalog, being both multi-fractals and having both some similar scaling properties, are not equivalent as point processes. This fact gives a good idea about the discriminating power of the multifractal formalism.

IX. CONCLUSIONS

It has long been realized that the clustering of galaxies is a highly complex process. The two-point correlation function is a first-order descriptor that owes much of its importance to its intimate connection with the dynamical aspects of the clustering process (Peebles 1980). The possibility that there may be a simple relationship between all orders of correlation function would allow a more complete description of the clustering process (Fry 1984; Schaeffer 1984), but there is no evidence for the kind of extrapolation that has been suggested apart from the rather tentative studies of the projected three- and four-point correlation functions.

The most useful hypothesis is that there may be a rather complex kind of scaling underlying the galaxy distribution. So, although we cannot know the high-order correlation functions in detail, we can nevertheless know something about their amplitudes. (The simplest such hypothesis, that the universe is a homogeneous fractal manifestly fails, so we have to go deeper than that). The clustering information is contained in the D_q [or $f(\alpha)$] curves. If the scaling is of the multifractal type, then that information is sufficient to specify the entire clustering process. This follows from the intimate connection between D_q and the moment generating function.

Multifractal measures can be computed reliably even for samples of a small number of objects, and therefore we can apply this formalism to analyze real galaxy catalogs. The evidence that the multifractal description is a correct description comes entirely from the CfA catalog, and expanding that survey will clarify the situation. Until then, it is the best hypothesis we have.

Turning to the paradigms, we see that they play an important role in assessing our understanding of clustering processes in general and of our ability to describe the processes quantitatively. The simple fractal hierarchy fails because the Hausdorff dimension of the galaxy distribution is not related to the correlation dimension in the way that would be expected in a homogeneous fractal distribution.

The Voronoi model is especially remarkable in this respect, since, despite its somewhat abstruse manner of generation, it has very interesting scaling properties. It is a multifractal, and it has a rich enough dimensionality structure. What is surprising perhaps is that it is based in the first instance on a Poisson process (the initial tessellation centers).

It should be interesting to apply the multifractal formalism to cosmological N-body simulations. In each time step of this kind of simulations, the exponent of the correlation function is calculated by fitting a power law, and it is compared with the observed value $\gamma = 1.77$ (White *et al.* 1987). Probably, it will be more reliable to do the comparison not only with $D_2 \sim 3 - \gamma$, but with the whole spectrum D_q or $f(\alpha)$.

It is not to be expected that any such scaling would extend over all cosmic scales. Indeed, the only evidence we have for any scaling is in regions where the two-point correlation function is larger than unity. We know that even that scaling must break down on the larger scales where the universe becomes homogeneous. It will also break down on the very smallest scales where star formation rather than galaxy formation has been the dominant process. So such discussions are inevitably confined to discussions about what happens on scales $1h^{-1}$–$10h^{-1}$ Mpc. We could not expect to make statements about what is observed on larger scales since we expect the scaling to have broken down by then. However, the initial conditions that led to the observed scaling behavior on the $1h^{-1}$–$10h^{-1}$ Mpc scales may have influenced the largest scale structures. That is a question of dynamics and is as yet largely unsolved.

The suggestion from pancake-like models (Buchert 1989) is that we may be close to seeing that link.

REFERENCES

Badii, R., and Politi, A. 1984 *Phys. Rev. Letters*, **52**, 1661.
Bahcall, N., and Soneira, R. 1983, *Ap. J.*, **270**, 20.
Barrow, J. D., Bhavsar, S. P., and Sonoda, D. H. 1985, *M.N.R.A.S.*, **216**, 17.
Bhavsar, S. P., and Ling, N. 1988, *Ap. J. (Letters)*, **331**, L63.
Bonometto, S. A., and Sharp, N. A. 1980, *Astr. Ap.*, **92**, 222.
Bouchet, F. R., and Lachièze-Rey M. 1986, *Ap. J. (Letters)*, **302**, L37.
Buchert, T. 1989, *Astr. Ap.*, **223**, 9.
Coleman, P. H., Pietronero, L., and Sanders, R. H. 1988, *Astr. Ap.*, **200**, L32.
Coniglio, A. 1986, in *Fractals in Physics*, ed. L. Pietronero and E. Tosatti. (Amsterdam: North Holland), p. 165.
Copson, T. E. 1967, *Asymptotic Expansions* (Cambridge: Cambridge University Press).
Davis, M., and Peebles, P. J. E. 1983, *Ap. J.*, **267**, 465.
de Lapparent, V., Geller, M., and Huchra, J. 1986, *Ap. J. (Letters)*, **302**, L1.
Domínguez-Tenreiro, R., and Martínez, V. J. 1989, *Ap. J. (Letters)*, **339**, L9.
Domínguez-Tenreiro, R., Roy, L. J., and Martínez, V. J. 1990, preprint.
Efstathiou, G., Fall, S. M., and Hogan, G. 1979, *M.N.R.A.S.*, **189**, 203.
Einasto, J., Klypin, A. A., Saar, E., and Shandarin, S. F. 1984, *M.N.R.A.S.*, **206**, 529.
Einasto, J., Klypin, A. A., and Saar, E. 1986, *M.N.R.A.S.*, **219**, 457.
Falconer, K. J. 1985, *The Geometry of Fractal Sets* (Cambridge: Cambridge University Press).
Frisch, U., and Parisi, G. 1985, in *Turbulence and Predictability of Geophysical Flows and Climatic Dynamics*, ed. N. Ghil, R. Benzi and G. Parisi (Amsterdam: North Holland), p. 84.
Fry, J. N. 1984, *Ap. J. (Letters)*, **277**, L5.
Gilbert, E. N. 1962, *Ann. Math. Stat.*, **33**, 958.
Giovanelli, R., Haynes, M. P., and Chincarini, G. L. 1986, *Ap. J.*, **300**, 77.
Gott, J. R., Melott, A. L., and Dickinson, M. 1986, *Ap. J.*, **306**, 341.
Grassberger, P., Badii, R., and Politi, A. 1988, *J. Stat. Phys.*, **51**, 135.
Grassberger, P., and Procaccia, I. 1983, *Phys. Rev. Letters*, **50**, 346.
Halsey, T. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I., and Shraiman, B. I. 1986, *Phys. Rev. A.*, **33**, 1141.
Hausdorff, F. 1919, *Math. Ann.*, **79**, 157.

Hentschel, H. G. E., and Procaccia, I. 1983, *Physica D*, **8**, 435.
Icke, V. 1984, *M.N.R.A.S.*, **206**, 1P.
Icke, V., and van de Weygaert, R. 1987, *Astr. Ap.*, **184**, 16.
Jensen, M. H., Kadanoff, L. P., Libchaber, A., Procaccia, I., and Stavans, J. 1985, *Phys. Rev. Letters*, **55**, 2798.
Jones, B. J. T., Martínez, V. J., Saar, E., and Einasto, J. 1988, *Ap. J. (Letters)*, **332**, L1.
Kirshner, R. P., Oemler, A., Schechter, P. L., and Shectman, S. A. 1981, *Ap. J. (Letters)*, **248**, L57.
———. 1987, *Ap. J.*, **314**, 493.
Kruskal, J. B. 1956, *Proc. American Math. Soc.*, **7**, 48.
Mandelbrot, B. B. 1974, *J. Fluid. Mech.*, **62**, 331.
———. 1977, *Fractals* (San Francisco: Freeman).
———. 1982, *The Fractal Geoemetry of Nature* (San Francisco: Freeman).
Martínez, V. J., and Jones, B. J. T. 1990, *M.N.R.A.S.*, **242**, 517.
Meakin, P. 1987, *Phys. Rev A.*, **36**, 2833.
Meyering, J. L. 1953, *Philips Res. Rept*, **8**, 270.
Miles, R. E. 1970, *Mathematical Biosciences*, **6**, 85.
Peebles, P. J. E. 1980, *The Large-Scale Structure of the Universe* (Princeton: Princeton University Press).
Pietronero, L. 1987, *Physica A*, **144**, 257.
Prim, R. C. 1957, *Bell System Tech. J.*, **36**, 1389.
Renyi, A. 1970. *Probability Theory* (Amsterdam: North Holland).
Schaeffer, R. 1984, *Astr. Ap.*, **134**, L15.
Soneira, R., and Peebles, P. J. E. 1978, *A.J.*, **83**, 845.
Tully, R. B., and Fisher, J. R. 1978, *IAU Symposium 79, The Large-Scale Structure of the Universe*, ed. M. S. Longair and J. Einasto (Dordrecht: Reidel), p. 214.
van de Weygaert, R., and Icke, V. 1989, *Astr. Ap.*, **213**, 1.
van de Weygaert, R., Jones, B. J. T., and Martínez, V. J. 1990, in preparation.
White, S. D. M., Davis, M., Efstathiou, G., and Frenck, C. S. 1987, *Nature*, **330**, 451.
Whitney, V. K. M. 1972, *Comm. Assoc. Comput. Mach.*, **15**, 273.

ROSA DOMÍNGUEZ-TENREIRO: Departamento de Física Teórica, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

BERNARD J. T. JONES and RIEN VAN DE WEYGAERT: Sterrewacht Leiden Postbus 9513, 2300 RA Leiden, The Netherlands

VICENT J. MARTÍNEZ: Departament de Matemàtica Aplicada i Astronomia, Universitat de València, Burjassot, 46100 València, Spain