GALAXY FORMATION*

By G. Börner

MPI für Astrophysik, 8046 Garching, FRG

(Received November 3, 1989)

Basic problems of formation of large structures in the universe are discussed.

PACS numbers: 98.60.Ac

1. Introduction

Astronomical observations during the last few decades have established a few basic features of the universe. All the galaxies that can be observed (except the very near ones) are moving away from our position. Obviously then they all must have been much closer to us at earlier epochs. The universe was much more densely filled with matter at an early epoch than at present. The detection of an all-pervading smooth radiation field corresponding to a temperature of 3 K of a thermal black-body radiation shows that the early stages in the cosmic history must have been hot and dense. This is also fairly obvious since an expanding radiation cools, while a contracting one heats up.

Thus our present environment of stars, galaxies, and very cold radiation has evolved from a very different state of a hot and dense mixture of matter and radiation. This experimental fact can be accomodated perfectly by the simple and highly symmetric Friedmann-Lemaître solutions of Einstein's theory of general relativity. These models describe the isotropic and homogeneous overall expansion of the matter and radiation contents in terms of an ideal fluid characterized by its density ϱ and pressure p. For a detailed introduction see e.g. [8].

In these models a time coordinate can be introduced such that the surfaces of constant time are 3-spaces of constant curvature. The line element can be written

$$ds^2 = dt^2 - R^2(t)d\sigma^2.$$

The 3-space line element $d\sigma^2$ can be written

$$d\sigma^2 = d\gamma^2 + f^2(\gamma)\,(d\vartheta^2 + \sin^2\vartheta d\varrho^2),$$

where $f(\gamma)$ depends on the value of the normalized constant curvature $K\left(\frac{K}{R^2(t)}\right.$ is the

* Presented at the XXIX Cracow School of Theoretical Physics, Zakopane, Poland, June 2–12, 1989.

Gaussian curvature of the 3-space)

$$f(\gamma) = \begin{cases} \sin\gamma & \text{for} \quad K = +1; \\ \gamma & \text{for} \quad K = 0; \\ \sinh\gamma & \text{for} \quad K = -1. \end{cases}$$

This form of the metric requires that the energy-momentum tensor is of the perfect fluid type

$$T_{\mu\nu} = \varrho V_\mu V_\nu - p(g_{\mu\nu} - V_\mu V_\nu).$$

The density ϱ and pressure p are functions of t only, and the flow lines with tangent V_μ are the curves $(\gamma, \vartheta, \varphi) = \text{const}$. The coordinates chosen in (2) are "comoving". The cosmic time t corresponds to the proper time of observers following the mean motion V_μ. The redshift $z = \dfrac{\Delta\lambda}{\lambda}$ of light from a comoving galaxy is given by $1+z = \dfrac{R(t_0)}{R(t_e)}$, where t_e: time of emission, t_0: present time.

The expansion factor $R(t)$ changes with time like the spatial distance of any two particles of the perfect fluid, i.e. of 2 "galaxies" without peculiar motions. Such comoving galaxies are considered to be the representative particles following the perfect fluid motion. It must be tested observationally, of course, whether a system of real galaxies can be approximated well by a perfect fluid. Einstein's field equations give the "Friedmann" equations for the expansion factor $R(t)$:

$$\dot{R} = \frac{8\pi G}{3}\varrho R^2 + \tfrac{1}{3}\Lambda R^2 - K;$$

$$\frac{d}{dt}(\varrho R^3) + p\frac{d}{dt}(R^3) = 0.$$

These equations determine the dynamics of the cosmological model, if an equation of state, e.g. $p = f(\varrho, t)$ or $p = f(\varrho)$ is given.

The rate of change $\dfrac{\dot{R}}{R}$ at the present time t_0 is the Hubble constant $H_0 \equiv \dfrac{\dot{R}(t_0)}{R(t_0)}$ which can be measured, if the distance and the redshift of some comoving object is known precisely. In practice reliable distance measurements are only possible for local regions ($z < 0.001$), but the galaxies follow the smooth Hubble expansion probably only at distances beyond the Virgo cluster ($z > 0.01$). Therefore the uncertainties are still quite large. Values between

$$H_0 = (50 \pm 7)\ \text{km s}^{-1}\ \text{Mpc}^{-1} \qquad [6]$$

and

$$H_0 = (100 \pm 10)\ \text{km s}^{-1}\ \text{Mpc}^{-1} \qquad [7]$$

are found from different interpretations of the observations. The inverse H_0^{-1} has the dimensions of time, the characteristic time of the expansion. Values are in the range (1 to 2) $\times 10^{10}$ years. For $K = 0$ models the present epoch $t_0 = \dfrac{2}{3H_0}$.

The Hubble constant gives the scale of the universe, and it can be used to express all other physical parameters as dimensionless quantities. Thus $\varrho_c = \frac{8\pi G}{3H_0^2}$ corresponds to a "critical" density, and the density parameter

$$\Omega_0 = \frac{\varrho_0}{\varrho_c}$$

is often used instead of the mean density. The sign of the curvature is determined by Ω_0 (and the cosmological constant Λ):

$$\frac{K}{H_0^2 R_0^2} = \Omega_0 - 1 + \frac{\Lambda}{H_0^2}.$$

For $\Lambda = 0$ the relation is especially simple.
Then

$$K = +1 \quad \text{for} \quad \Omega_0 > 1;$$

$$K = 0 \quad \text{for} \quad \Omega_0 = 1;$$

$$K = -1 \quad \text{for} \quad \Omega_0 < 1.$$

Values of the mean density are not known as precisely, as would be necessary to decide which of the 3 possibilities actually occurs. Number counts of galaxies lead to a value $\Omega_{0,G} \simeq 0.01$. Application of the virial theorem to rich clusters of galaxies shows that there must be a certain amount of nonluminous, dark matter. If this is taken into account (extrapolating the mass-to-luminosity ratio from rich clusters to all galaxies) then $\Omega_0 = 0.1$ to 0.3.

This is rather close to $\Omega_0 = 1$, the critical value (for $\Lambda = 0$) giving $K = 0$. We may conclude that the observations point to $\Omega_0 < 1$, but $\Omega_0 = 1$ or even $\Omega_0 \geqslant 1$ cannot be excluded, since any background density distributed more homogeneously than the large clusters is not accounted for.

In the context of such simple models there remains the fundamental task to explain the extreme inhomogeneities observed at present: the sizes and shapes of galaxies, and the clustering hierarchy of galaxies into groups, clusters, and superclusters. Generally these structures are thought to have evolved from initially small perturbations of a smooth background density which have grown by gravitational interaction. Clearly, there must be some scale, where local physics is the dominant effect in shaping the objects. At present it is not clear what this scale is.

Baryonic fluctuations below a certain size will be strongly damped by photon diffusion. The size — the "Silk damping-mass" M_D — corresponds approximately to the mass of a galaxy or of a group of galaxies. There is no further scale information, however, which could, for example, tell us why a fluctuation of $100 M_D$ should eventually break up into smaller pieces of galactic size. The case of star formation is understood much better. We have at least a good idea of the mechanism — an interplay of gravitation and nuclear

physics — which lead to the fragmentation of interstellar clouds of widely different sizes into stars with masses in a narrow range of 1 to 60 solar masses.

Two different general schemes have been proposed. In some theories, the formation of structures starts on small scales and proceeds step by step to the larger structures. Galaxies cluster hierarchically to build groups, groups form clusters, and clusters merge to form superclusters [9] (the "hierarchical" model). In other models, the largest structures formed first via gravitational instability, followed by fragmentation processes to shape the smaller structures. It seems plausible that the largest structures collapse anisotropically forming thin, dense sheets of matter. This aspect has given to this scenario its popular name, the "pancake" picture [10, 8, 2]. In Section 4 below I want to show you a few details of simulations making use of the pancake scenario, because it is in this field that some work has been done at the MPI für Astrophysik in München. Before that in Section 2 some visual impressions will be discussed, and in Section 3 a few basic theoretical concepts. Finally Section 5 deals with a few detailed statistical investigations of the galaxy distribution.

2. *Visual impressions of the galaxy distribution*

Already the projection of the galaxies on the sky shows that the distribution is very inhomogeneous. There are double system, multiples, groups with up to 100 members, rich clusters with more than 1000 members, and superclusters encompassing many rich clusters.

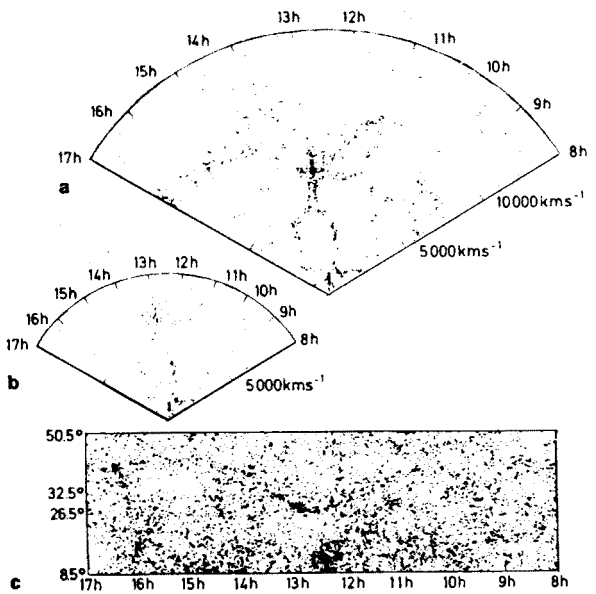

Fig. 1. A map of the redshifts (expressed as a velocity cz) of 1061 galaxies plotted versus Right Ascension in a wedge of declination δ, within the interval $26.5° < \delta < 32.5°$. $m_B \leq 15.5$ and $v < 15000$ km/s (a). (b) contains the 182 galaxies with $m_B < 14.5$, $v = 10000$ km/s in the same region. In (c) the projected map is shown of 7031 objects with $m_B < 15.5$, listed by Zwicky et al. in the region $8^h < \alpha < 17^h$; $8.5° < \delta < 50.5°$ [11]

The Virgo cluster near the north galactic pole, as well as the Local Supercluster can be discerned in the maps of galaxies brighter than 13th photographic magnitude prepared as early as 1932 by Harlow Shapley and Adelaide Ames.

The great number of redshifts that have been measured with large telescopes and modern equipment help to get an idea of the 3-dimensional spatial structure. The assumption of an undisturbed Hubble expansion relates the redshift directly to distance ($cz = H_0 d$). Fig. 1 displays a famous result obtained from a systematic collection of the redshifts of all galaxies of apparent magnitude $m \leqslant 15.5$ in a region of the sky measuring 6° in latitude

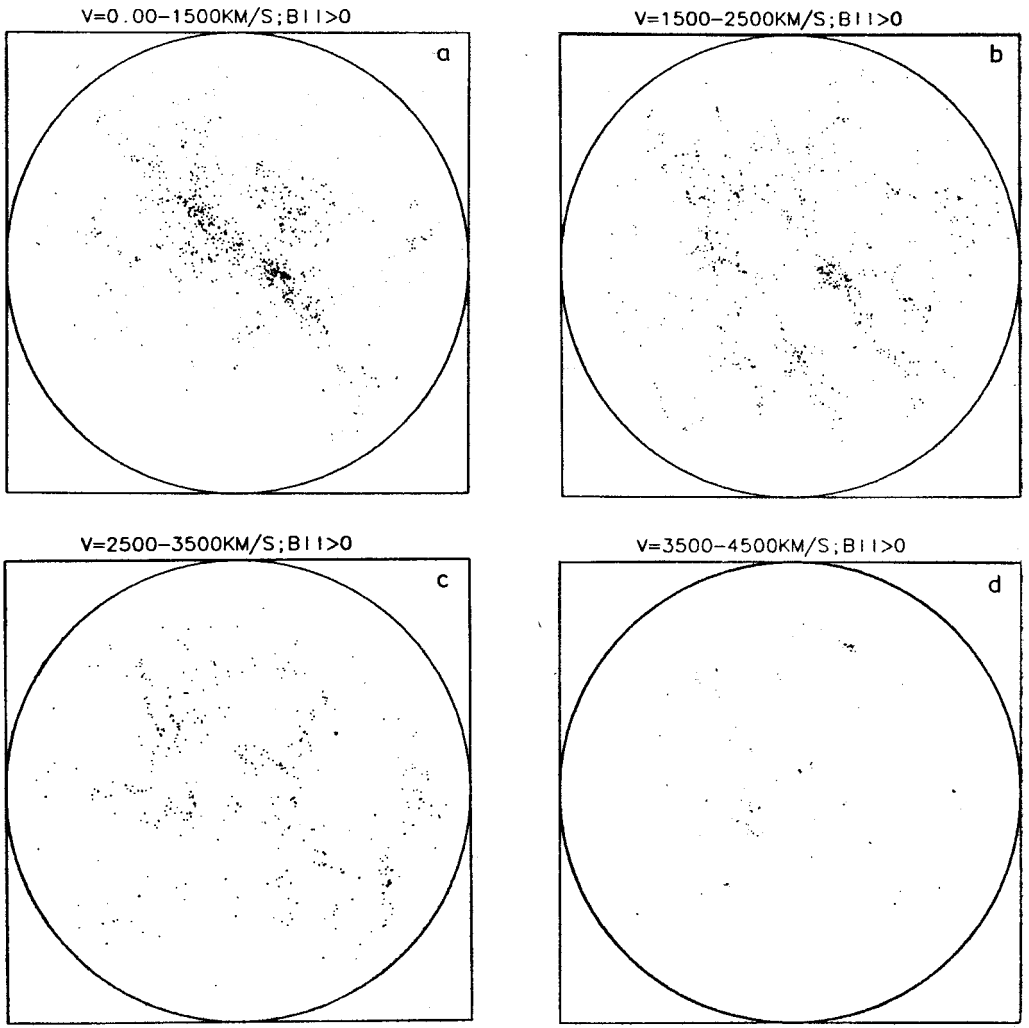

Fig. 2. Projection of the galaxies at a certain distance onto the plane of the galactic equator. Coordinates are galactic latitude and longitude. The galactic pole is at the center, concentric circles around it define galactic latitude. Maps are constructed from the ZCAT catalogue by Houjun Mo. Redshift ranges are indicated

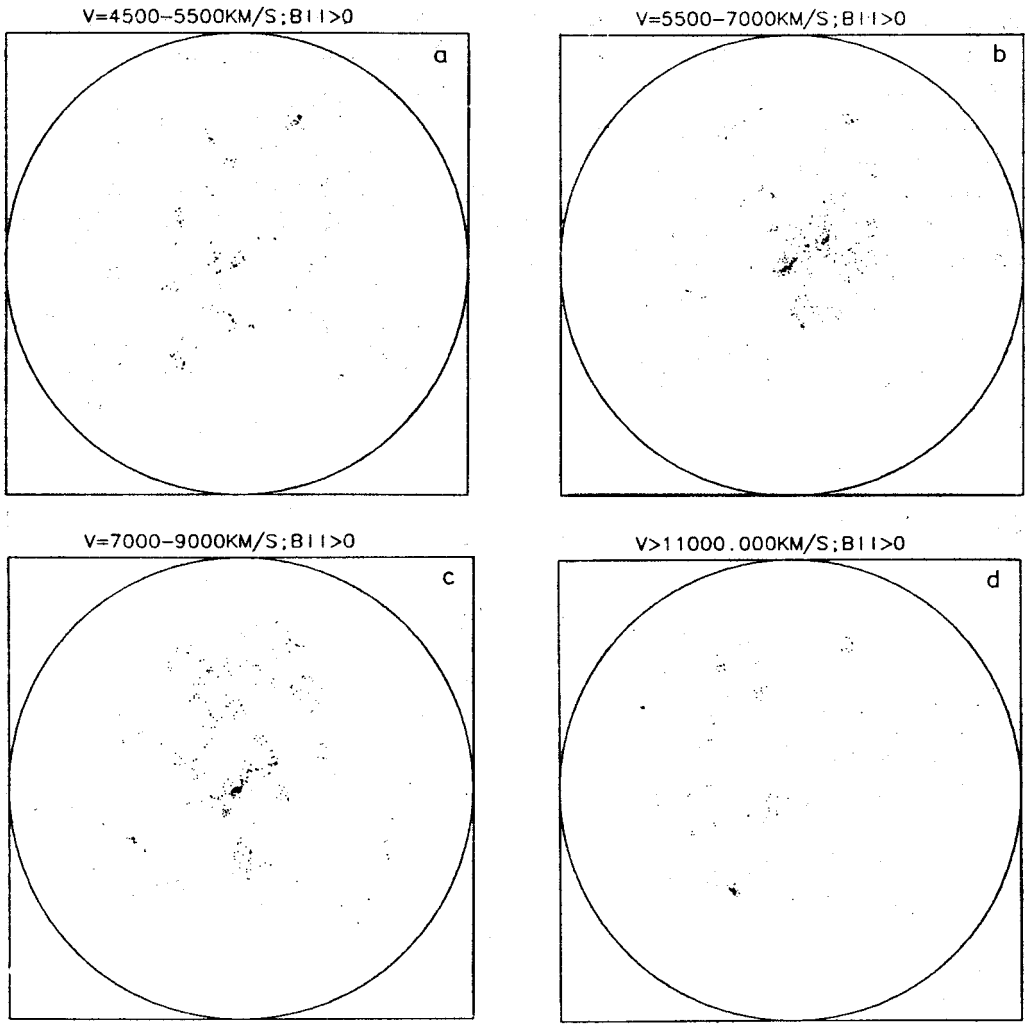

Fig. 3. Same as 2, but different redshift ranges

and 117° in longitude [11]. The 6° wide strip between galactic latitude +26.5° and +32.5° is projected onto the plane of the wedge. Along the edges the measured redshift, and the galactic longitude (given as right ascension) are used as coordinates. The map of this slice gives a vivid impression of the clustering of galaxies into thin sheets or filaments, whilst space is dominated by large, apparently empty (i.e. containing much less galaxies) quasispherical voids. The universe appears to have a bubbly, foam-like structure. The man-shaped figure in the center is the Coma cluster. The elongation pointing towards the observer ("finger of God") is due to peculiar velocities in the cluster, where Doppler shifts add to the general expansion redshift.

In Figs 2 and 3 several maps are shown which have been constructed by Houjun Mo (MPI f. Astrophysik) from the ZCAT catalogue. (The ZCAT is a continuing compilation

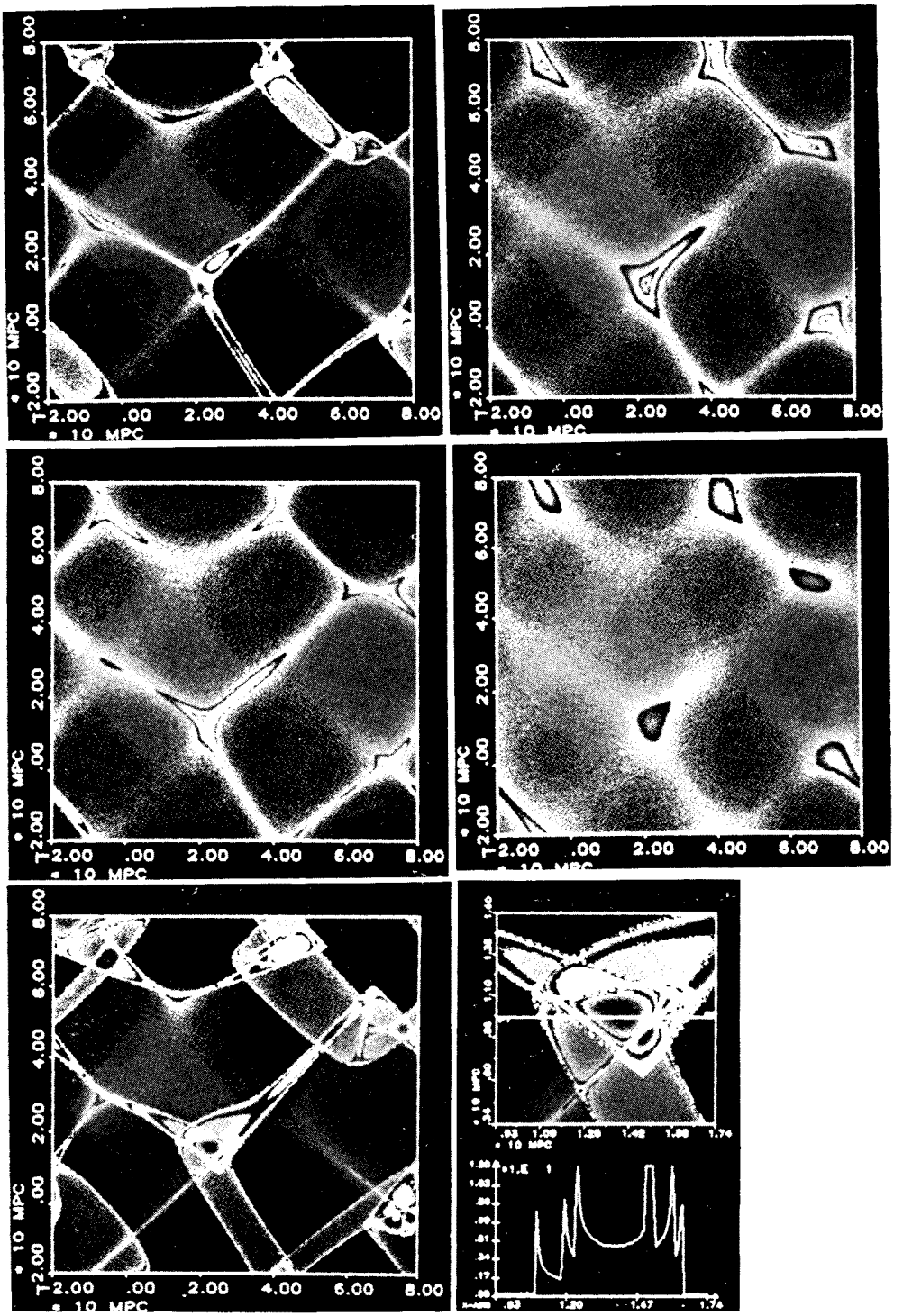

Fig. 5. High-resolution 2-dimensional simulation of a pancake-model [2]. Different shades from black to white give the range from "voids" ($\frac{1}{2}$ of the mean density) to high density "caustics". Interesting singularity structures can be seen (e.g. pancakes penetrating each other). Initial data are Fourier spectra with Gaussian amplitudes. Density and velocity perturbations are prescribed independently with 25 modes each. Initial density amplitude $= 10^{-3}$, initial redshift 1000, $\Omega = 1$. 10^8 particle trajectories are collected in $(512)^2$

Fig. 6. 3-dimensional simulations of the Zel'dovich solution [3]. The pictures are projections of contours in Lagrange space. From left to right and top to bottom the surfaces surround regions with successively smaller positive eigenvalues of the matrix $\frac{\partial^2 \psi}{\partial \xi_j \partial \xi_k}$. Surfaces with smaller eigenvalues enclose those belonging to larger ones. Pancake structures develop inside these contours. Top right and middle left show two different views at rigt angles of the same structure

of all galaxies with measured redshifts, set up by M. Geller and P. Huchra of the CfA). This catalogue contains about 20000 galaxies and about 8000 of those have already measured redshifts. H. Mo has made stereographic maps of the celestial sphere for galaxies in different redshift ranges. The galactic pole is at the center, concentric circles around it define galactic latitude, and longitude (or right ascension) is measured along the circles. The different redshift ranges should be viewed like slices through space at different distances. This

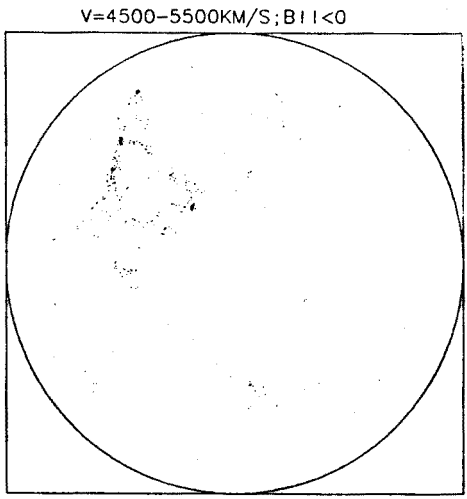

Fig. 4. Map from the ZCAT for the southern hemisphere (bII < 0). The redshift range is 4500 to 5500 km/s. Note the large filamentary feature un the upper left

presentation clearly exhibits the dominance of the Local Supercluster close by (Fig. 2, v = 0–1500 km/s). But it also shows nicely how the concentration of galaxies on the surfaces of spherically shaped voids is a dominating aspect of the more distant objects. It comes out especially clearly for the northern hemisphere in the redshift range of 3500 km/s to 5500 km/s. At large redshifts there are again some of these features (the voids must be expected to be smaller by a factor 2 between 5000 km/s and 11000 km/s), but less pronounced because there are not so many bright galaxies (cf. Figs 2 and 3). In Fig. 4 a map from the southern sky (latitude $b < 0$) is shown of the slice v = 4500 to 5500 km/s. There is a large filamentary of sheetlike feature in the upper left part which looks remarkably like a coherent structure. Do we see here a pancake or a cosmic string?

3. Basic theoretical concepts

A natural way to explain the formation of galaxies and clusters in a smooth universe is to assume that initially small density perturbations grow and eventually collapse due to gravitational instability. This idea meets the following difficulty: In linear approximation small density contrasts $\delta = \frac{\delta\varrho}{\varrho}$ grow after the epoch of recombination as a power of the

cosmic time $t^{2/3}$. Recombination, i.e. the formation of hydrogen atoms occurs at a redshift $z \approx 1500$.

The density contrast at the present epoch $\delta(t_0)$ is related to the amplitude at the recombination time t_R by

$$\delta(t_0) = (1+z)\delta(t_R), \tag{1}$$

where $z \sim 10^3$ for $\Omega_0 = 1$, $z \sim 10^2$ for $\Omega_0 \ll 1$.
Initial perturbations $\delta(t_R)$ can thus increase their value by factors 100 to 1000, depending on Ω_0.

The initial amplitudes are limited, on the other hand, by the high degree of isotropy of the 3 K background. Observations of the radiation temperature in different directions give upper limits of variations at various angular scales:

$$\frac{\delta T}{T} \leqslant 4\times 10^{-5} \quad \text{at } 3',$$

$$\frac{\delta T}{T} \quad 1.5\times 10^{-5} \quad \textit{at } 7',$$

$$\frac{\delta T}{T} \leqslant 3.7\times 10^{-5} \quad \text{at } 8^0.$$

A mass M at recombination time subtends an angle

$$\theta = 40''(h\Omega_0)\left(\frac{M}{10^{11}M_\odot}\right)^{1/3}, \quad (h = H_0/100) \tag{2}$$

i.e. these limits exist on the scale of clusters. For adiabatic perturbations $\frac{\delta n}{n} = -3\frac{\delta T}{T}$, and limits on the density contrast are obtained from the 3 K background isotropy: e.g $\delta(t_R) \leqslant 5\times 10^{-5}$ at 7', hence on the scale of 7'

$$\delta(t_0) \leqslant 5\times 10^{-2} \quad \text{for} \quad \Omega_0 = 1,$$

$$\delta(t_0) \leqslant 5\times 10^{-3} \quad \text{for} \quad \Omega_0 \ll 1.$$

There is not enough growth in the small perturbations! The universe stays homogeneous, and therefore the natural model of galaxy formation from adiabatic perturbations has to be abandoned. There can be several different ways out from that difficulty, none without its own, peculiar problems:

(*i*) *isothermal fluctuations*: these require $\delta T = 0$ by some (more or less artificial) constraint on the fluctuations. Isothermal fluctuations start to grow only, when adiabatic fluctuations of the same magnitude have developed.

(*ii*) *reionization*: an epoch of reheating can lead to reionization of the matter, and can smear out fluctuations of the photon field on small angular scales. The bulk motion of the scattering electrons will be appreciable, however, in most models, and will introduce new disturbances of similar magnitude as the original ones.

(*iii*) *nonbaryonic dark matter*: Nonluminous matter of nonbaryonic form is a possibility with the drawback that experimentally no such object as a massive neutrino, an axion, or a cosmic string has been found. But a background of these objects does not interact with photons, can have larger initial perturbations, larger growth times, and provide the gravitational wells where baryonic matter accumulates.

Despite the difficulties mentioned galaxy formation is an active field of research which proceeds on different levels with respect to problems of detail without a complete and consistent picture at present. In the following I want to mention two highlights of research at our institute.

4. The "pancake scenario" of Zel'dovich

This model describes in an approximate, analytic way even some nonlinear features of the collapse of collisionless matter.

It turns out that an approximate solution to the Newtonian equations of gravity can be found by considering the integral curves of the particle trajectories in Lagrange space.

The relation between Euler (x_i) and Lagrange (ξ_i) coordinates ($i = 1, 2, 3$) is

$$x_i = R(t)\xi_i + b(t)\frac{\partial \psi}{\partial \xi_i}. \tag{3}$$

$R(t)$ is the expansion factor, and $b(t)$ is obtained from the linear evolution equations $b(t) \sim t^{4/3}$.

The function $\psi(\xi)$ describes the initial perturbation spectrum of the particle motion. Mass conservation

$$dm = \varrho dx_1 dx_2 dx_3 = \varrho_0 d\xi_1 d\xi_2 d\xi_3 R^3(t)$$

leads to the relation

$$\varrho = \varrho_0 \left\{\det\left(\frac{dx_i}{d\xi_k}\right)\right\}^{-1} \tag{4}$$

between the homogeneous background density ϱ_0, and the real density. Clearly something drastic must happen at the zeros of the determinant! Transformation to principal axes gives eigenvalues α_i ($i = 1, 2, 3$) with

$$\varrho = \varrho_0 \left(1 - \frac{b}{R}\alpha_1\right)^{-1}\left(1 - \frac{b}{R}\alpha_2\right)^{-1}\left(1 - \frac{b}{R}\alpha_3\right)^{-1}. \tag{5}$$

Let $\alpha_1 > \alpha_2 > \alpha_3$, then ϱ becomes infinite for $(b/R)\alpha_1 \to 1$!

This approximation by Zel'dovich describes the formation of regions of very high density. The generic collapse seems to be preferentially one dimensional, i.e. onto a plane. Therefore the name "pancake" scenario was coined for the model.

The mathematical background of this structure formation is the formation of caustics by particle trajectories (described and classified in the context of "catastrophe theory"

by Arnold). We have been working on some aspects of this model for some time. Exact analytic solutions for the pancake picture have been found [2]. High resolution numerical simulations of particle trajectories have also been done [2, 3].

Fig. 5 shows a 2-dimensional simulation, where voids (black regions), the linear regime (faint), and high-density objects (strongly white) can be seen. Pancake structures penetrating each other can be discerned. The singularity classification given by Arnold can be applied to the complicated structures at the knots in this picture.

Fig. 6 gives an impression of a 3-dimensional simulation of particle trajectories. Pancakes have been determined by the magnitude of the eigenvalues of $\det \frac{dx_i}{d\xi_k}$. The function $\psi(\xi)$ is always taken as

$$\psi(\xi) = \sum_{K}^{k_{max}} A(K) \sin(K \cdot \xi) + B(K) \cos(K \cdot \xi),$$

with random numbers $A(K)$ and $B(K)$.

In calculations a few hundred Fourier components are used, and a cut-off K_{max} is designed such that all perturbations below a certain scale are eliminated. This should model the effect of the coherence length present in hot dark matter scenarios.

So these simulations give a general impression of nonlinear effects, but not yet a realistic scenario that can be compared to the observations. To do this, we shall have to implement initial conditions derived from the linear perturbation theory of a hot or cold DM model.

5. What are the largest scales of clustering?

From an analysis of Abell clusters a single, very large object of angular extent $\sim 40°$ has been found [4]. This corresponds to a scale of 270 h^{-1} Mpc. We have carried out an analysis of Abell and Zwicky clusters which has revealed the existence of such large scale objects to be quite generic [5]. The analysis of the cluster-distribution is carried out as follows: The sample area (on the celestial sphere) is divided into pixels which are assigned a number 1 or 0 depending on whether or not their cluster density exceeds some threshold density (Clusters are approximated by a density distribution $\sim \exp(-\theta^2/\lambda^2)/\pi\lambda^2$ where λ is the mean distance $\sim \langle n \rangle^{-1/2}$ in the sample). Pixels with a 1 assigned are considered to belong to one cluster, if they touch each other. A percolation analysis of these pixel clusters is then carried out. The real sample is compared to randomly generated samples (Figs 7, 8). We find that significant differences from a random distribution exist up to scales of 50° corresponding to a length of $\sim 350h^{-1}$ Mpc with a density contrast of $\delta \sim 0.5$. It can be seen from Fig. 8 that scales increase with decreasing density contrast indicating the possibility that the structures become larger and larger, but also gradually blend into a uniform background.

Can the large scale structures be due to a chance alignment of small scale clusters $\sim 15°$? A test is carried out by division of the sample into cells of size 15° (20° and 30° tests are also done), and a random reshuffling of cells. If the reshuffled sample is signifi-

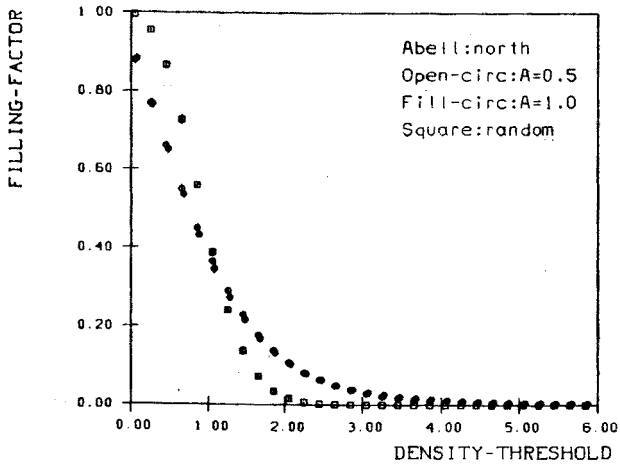

Fig. 7 The filling factor is the fraction of the sample volume occupied by pixel clusters compared to the total. $A = 0.5$ and $A = 1.0$ are two different models for galactic extinction. 1σ-bars are plotted, derived from 10 randomly generated samples

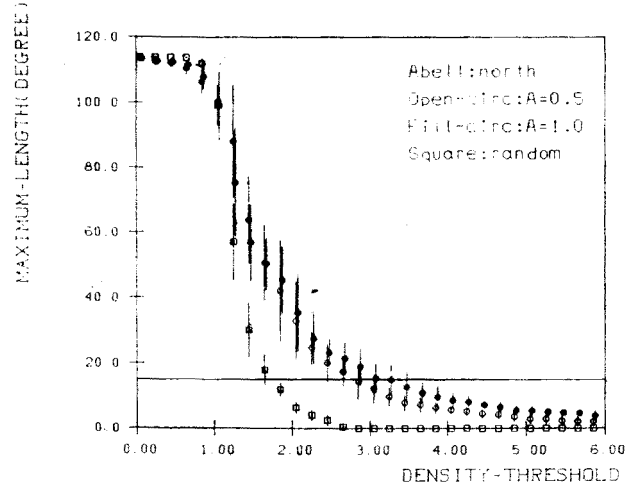

Fig. 8. The maximum length of clusters is plotted against the density threshold. The horizontal line indicates angular scales of 15°. Significant differences from the random samples are apparent for scales up to ~50°

cantly different from the original one, then the large scale structure is real and not due to a chance alignment. We find (Fig. 9) that such smaller scale alignment below 15° cannot be responsible for the big structures. For scales of 20° or 30° the evidence is not so clear as yet.

There is one big worry with these very large structures: The gravitational potential well redshifts photons, and a signal in the radiation background emerges (Sachs-Wolfe effect) of magnitude $\frac{\Delta T}{T} \approx 5 \times 10^{-4} \left(\frac{L}{300\ \text{Mpc}}\, h^{-1} \right)^2$ on a scale of $\left(\frac{L}{300} \right)$ degrees. Observa-

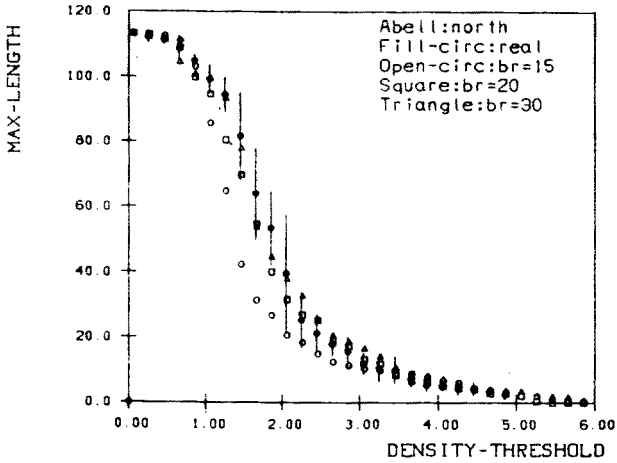

Fig. 9. A test for the effect of chance alignment and projection. Three artificial samples (reshuffled) with cell sizes 15°, 20°, 30° are compared to the original sample

tional limits are below that: $\frac{\Delta T}{T} \leqslant 10^{-4}$. Can this be reconciled? Perhaps the discrepancy can be removed, if the real mass distribution is much smoother than the distribution of the luminous objects.

REFERENCES

[1] Redhead et al., Proc. IAU Symp. No. 130.
[2] T. Buchert, Ph. D. Thesis, MPI f. Astrophysik, 1988; T. Buchert, G. Götz, *J. Math. Phys.* **28**, 2714 (1987); T. Buchert, *Astronomy* and *Astrophysics*, 1989, in press.
[3] R. Klaffl, Diploma Thesis, MPI f. Astrophysik.
[4] R. B. Tully, *Ap. J.* **303**, 25 (1989).
[5] H. Mo, G. Börner, MPA-preprint, 1989.
[6] A. R. Sandage, G. A. Tammann, *Nature* **307**, 326 (1984).
[7] G. de Vaucouleurs, Proc. 10th Texas Symp., N. Y. Acad. Sin. 1981, p. 90.
[8] G. Börner, *The Early Universe*, Springer Verlag 1988.
[9] P. J. E. Peebles, *Large scale of the Universe*, Princeton University Press, 1980.
[10] Y. B. Zel'dovich, *Astron. Astrophys.* **5**, 84 (1970).
[11] V. de Lapparent, J. M. Geller, J. P. Huchra, *Ap. J.* **302**, L1 (1986).