Fermi National Accelerator Laboratory

FERMILAB-Conf-84/135
December 1984

FORMATION OF GALAXIES

Alexander S. Szalay

Institute for Theoretical Physics
University of California
Santa Barbara, CA

and

Theoretical Astrophysics Group
Fermi National Accelerator Laboratory
Batavia, Illinois 60510, U.S.A.

Paper to be published in the Proceedings of the International Symposium 'Phase Transitions in the Very Early Universe', University of Bielefeld, June 1984

FORMATION OF GALAXIES

Alexander S. SZALAY

Institute for Theoretical Physics, University of California, Santa Barbara CA, and Astrophysics Group, Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL U.S.A.

The present theories of galaxy formation are reviewed. The relation between peculiar velocities and the correlation function of galaxies points to the possibility that galaxies do not form uniformly everywhere. Scale invariant properties of the cluster-cluster correlations are discussed. Comparing the correlation functions in a dimensionless way, galaxies appear to be stronger clustered, in contrast with the comparison of the dimensional amplitudes of the correlation functions. Theoretical implications of several observations as Lyman-α clouds, correlations of faint galaxies are discussed. None of the present theories of galaxy formation can account for all facts in a natural way.

1. INITIAL CONDITIONS

The universe contains a wide dynamic range of objects: from stars (1 M_o) all the way to superclusters ($10^{16} M_o$). A major question that we are unable to answer yet is whether the formation of structure has started with smaller masses clustering on ever larger scales[1], or whether extremely large structures formed first, then subsequently fragmented into smaller ones.[2] If we knew the precise initial conditions then the present structure of the universe could be derived by applying the laws of physics. Let us summarize, what has to be known about the initial conditions for this ambitious project.

The fluctuations are likely to be adiabatic, since the specific entropy of the universe, n_B/n_γ is tied to microscopic parameters of particle physics. In the inflationary theories quantum fluctuations arise in a natural way. However, the necessary amplitude seems to require rather special prescriptions for the effective potential.[3] The initial perturbations are expected to be scale free, therefore their Fourier amplitude depending on the wavenumber k can be well described by a power law, $\delta_k^2 \sim k^n$. If the spectral index is n=1, the amplitude of the different perturbations is the same when their wavelength equals the horizon size. This 'double scale-invariant' is called the Zeldovich spectrum, and is known to arise in inflationary scenarios.[4] There are severe constraints on the fluctuation amplitudes. If the fluctuations were adiabatic, the perturbations of the metrics generate fluctuations in the temperature of the microwave background. On small angular scales (4.5 arc min) these limits are extremely small[5]:

$\delta T/T < 2.9 \times 10^{-5}$

The standard growth rate of fluctuations in a flat universe is $(1+Z)^{-1}$. The H-He plasma becomes gravitationally unstable only after recombination, at $Z \approx 1000$. At this point the density and temperature fluctuations are similar, $3\delta T/T=\delta \rho/\rho$. This does not leave enough margin for fluctuation growth, the fluctuations cannot reach the nonlinear stage our universe seems to be in today. Present calculations confirm[6,7] that if the universe is baryon dominated, only prohibitively high initial fluctuation amplitudes can result in the formation of galaxies. If the universe is dominated by some form of collisionless dark matter, the dark matter fluctuations are unaffected by pressure, therefore grow even before recombination. After recombination these curvature perturbations caused by the dark matter will accelerate fluctuation growth in the baryons, so the $\delta T/T$ constraints are less stringent.

Though the initial spectrum is a power law, by the time it becomes nonlinear it will be considerably modified. When the universe is radiation-dominated, fluctuations within the horizon have a minimal increase[8], whereas the ones outside the horizon grow. This effect will bend the slope of the spectrum from n to n^{-4} for wavenumbers higher than k_{eq}, corresponding to the size of the horizon when the matter and radiation energy densities were equal. The presence of the collisionless dark matter results in distortions of a different kind: the free motion of particles erases structures smaller than the free streaming scale.[9,10,11,12] The mass scale of this collisionless damping process can be expressed in terms of the mass and entropy of the particles the dark matter consists of.

$$M_x \approx 2.2\ m_p^3\ m_x^{-2}$$

In the case of neutrinos this mass takes the value of $M_{\nu m} \approx 3.2 \times 10^{15}\ m_{30}^{-2}\ M_\odot$, corresponding to the comoving length scale $\lambda_{\nu m}=41\ m_{30}^{-1}$ Mpc. Depending on what the 'temperature' of the dark matter is, this damping scale can change from the above 41 Mpc to extremely small values. The neutrinos are 'hot' particles, since their average momentum is close to that of the background radiation photons. Most other candidates for the dark matter like axions and photinos - yet undiscovered - would have decoupled much before the neutrinos, having a lower entropy or temperature, so they are called 'cold'. They hardly move at all, their damping scale is negligible. Intermediate candidates, like a gravitino of 1 keV mass would be 'warm'.

A major underlying assumption in calculating most consequences of a given fluctuation spectrum is that the phases of the individual Fourier components are random, i.e. the perturbations are a random Gaussian process. One can envisage scenarios, where this will not be the case, like perturbations originating from strings.[13] For a given spectrum combined with the assumption of random phases one can calculate the distribution of mass fluctuations, density of local peaks, density profiles around local peaks, the distribution of peaks of a given size, etc.

The expansion of the universe is characterized by $\Omega=\rho/\rho_{crit}$, the density parameter, H_o, the Hubble constant, Λ_0, the cosmological constant. $\Lambda_0=0$ and $\Omega=1$ corresponds to the flat universe, which appears to be necessary for inflation. Λ_0 is generally assumed to be negligible. However, $\Omega=1$ and $H_o=50$ km/s. Mpc with $\Lambda_0=0$ imply uncomfortably low values for the present age of the universe. Calculations of the primordial ^4He and D + ^3He abundance indicate[14], that the baryon density of the universe at the time of primordial nucleosynthesis lies in the range of $0.01 < \Omega_B < 0.1$. This suggests that if baryons dominate the mass density then the universe is open by a large margin.

Fluctuation growth also depends on the density of the universe. If $\Omega < 1$, the growth of perturbations effectively stops at the redshift $z \approx \Omega^{-1}$. The detailed predictions of $\delta T/T$ are below the current limits if the dark matter consists of neutrinos with about 30 eV mass, but restrict Ω if the cold particles dominate the universe.[6,7] $\Omega \gtrsim 0.2 \times h^{-4/3} \approx 0.5 \times h_{50}^{-4/3}$ ($h=H_o/100$ km/s.Mpc and $h_{50}=H_o/50$ km/s.Mpc, dimensionless). In deriving this limit it was assumed that galaxies follow the mass distribution: the amplitude of the fluctuations today was normalized to J_3, the integral of the galaxy-galaxy correlation function $\xi(r)$.

2. NONLINEAR STRUCTURE

Here we would like to discuss the expected structure of the universe if the dark matter is either hot, warm or cold. Once the first mass scale in a spectrum with a large damping cutoff (hot) reaches nonlinearity, particle trajectories cease expanding away from each other and converge, resulting in the temporary formation of caustics. The density becomes very high and a flat 'pancake' is formed.[2] At first they arise at isolated spots where the initial velocity perturbations had the largest gradient. Soon these regions grow, turning into huge surfaces which intersect, forming the walls of a cell-structure which is itself gravitationally unstable.

In this nonlinear phase mode-mode coupling among Fourier components sends power to short wavelengths, and correlates the phases even though the initial fluctuation spectrum may have had random ones. The methods of catastrophe theory were applied[15] to analyze structure that develops in such potential motion. It was found that the two dimensional pancakes are only the lowest order singularities; other singular topological structures should also appear. String-like features are one example, and they can be seen in the N-body simulations.

When the intersection of trajectories takes place, gas pressure builds up, the velocity of the collapsing gas exceeds the sound speed and a shock wave is formed.[2] The gas is shock-heated up to keV temperatures and cools by emitting radiation over a broad spectrum. Recently several authors[16] have calculated the cooling of collapsing neutrino-baryon pancakes, the details of which are considerably different from those in a pure baryon pancake[17]: the baryon density is lower, infall velocities are higher, thus the cooling rate is much slower. It is evident that the fraction that can cool significantly is a sensitive function of the mass of the collapsing region. This cooling is necessary, since only cold gas is able to form the seeds of galaxies, so the local column density modulates the rate of galaxy formation.

The UV and soft X-ray emission can photoionize the intergalactic medium, making galaxy formation in regions that have not yet formed pancakes more difficult, which would accentuate the contrast in galaxy density between the strings and pancakes vs. voids, even though the density contrast may be only 3-10.

If the dark matter is warm, it will still form pancakes, though of galactic size. There the cooling is much more efficient[18], those timescales will determine the fate of each object. If the dark matter is cold, the spectrum $k^{3/2}\delta_k$ is substantially different from the hot and warm case. It has no peak at all, but it is slowly increasing towards the smallest scales. These small scales will collapse first, but later the larger systems are also going nonlinear, forming a clustering hierarchy. Due to the complicated nature of these many-body interactions only numerical N-body simulations are able to follow the evolution of such systems.

3. N-BODY SIMULATIONS AND GALAXY CORRELATIONS

If we knew all the parameters listed above, it would be relatively easy to calculate the evolution of the universe. Only gravitational forces act on collisionless dark matter so one can numerically solve the transport equations, even in the nonlinear regime. This has indeed been done, as we discuss here. Given the initial conditions, these numerical experiments will tell us the mass distribution in the universe. One can hope, that the structure obtained this way will resemble the real universe. i.e. galaxies trace the mass distribution.

Starting from the above mentioned initial conditions extensive N-body simulations consisting of more than 32000 particles[19,20] were made. These projects all used some version of a particle/mesh Fourier code, 64^3 in size. The calculations were started, when $\delta\rho/\rho$ was about 0.2, and the approximate Zeldovich solution[2] corresponding to the growing mode of perturbations was used to determine initial positions and velocities, then the trajectories of the particles were integrated. The free parameters of the calculations are Ω, H_o and the initial amplitude of the fluctuations. For a given Ω one can use conservative limits for the age of the universe to obtain a value of H_o. If $\Omega=1$, then $t_0 > 12$ Gy requires $H_o < 54$ km/s.Mpc.

The initial amplitude can be defined in different ways. For simulations with hot dark matter, neutrinos, the epoch of galaxy formation, Z_{GF} was defined as the redshift when 1 percent of all particles have gone through a 'caustic'. Unless $Z_{GF} \approx 0.4$, the correlation function disagrees with that of the galaxies. For cold dark matter the initial amplitude is determined in a different way. Due to the growth of nonlinearity, $\xi(r)$ is rapidly increasing both in slope and amplitude, just like for hot dark matter. One can define 'today' when the correlation function of the particles most resembles that of the galaxies, i.e. a power law with a slope -1.8.

$$\xi(r) = (r/r_0)^{-1.8}$$

However, at this point the amplitude is too small. One can resolve this difficulty by choosing $H_o = 22$km/s.Mpc, but this is hardly the way out.

There is one more difficulty: the random velocity dispersion of galaxies is well known[21]:

$$\langle v_{12}{}^2\rangle^{1/2} \approx 300\text{-}400\text{km/s}.$$

In the neutrino simulations, if $Z_{GF} > 1$ the corresponding velocity dispersions are in the 1200 km/s range, clearly too high. In linear perturbation theory

$$\langle v_{12}{}^2\rangle \approx (180\ \text{km/s})^2\ f(\Omega)\ \xi(0)\ (H_o\lambda_{\nu m})^2.$$

Today, $\xi(0) \approx (1+Z_{GF})^2$, so either $\Omega << 1$, forbidden by the $\Delta T/T$ constraints, or Z_{GF} is small[22].

For cold dark matter a similar problem exists, but the high velocities arise from small scale nonlinearities; and since a low Ω model is ruled out, the only remaining possibility is to have

$$\xi(0) = |\delta\rho/\rho|^2$$

fairly small. Then we are in an even sharper contradiction with the observed galaxy autocorrelation.

On the other hand, the galaxies consist mostly of baryonic gas capable of emitting and absorbing radiation. These dissipative processes, strongly density and temperature dependent, occur at a different rate at different places.[18] All these effects, combined with possible shock waves due to the finite pressure in the H-He gas, may have an important role in determining where galaxies form. As a result, the galaxies may not follow the light at all, so the mass autocorrelation should not be compared to the galaxy autocorrelation. Galaxy formation, as long as it is a random process, initiated by gravitational infall will be likely to start at the regions of highest densities. One can therefore associate the particles in these regions with galaxies. This 'biasing' of galaxy formation towards these high densities is a heuristic procedure, but probably a fair approximation to what really happens. The detailed numerical procedure Davis et. al. used to select these 'biased' particles in the cold dark matter models involved a smoothing of the densities before the actual selection was made. The physical explanation of where the threshold of the selection should be is much less clear, it can only be adjusted to the observed number density of galaxies. This 'biasing' process will enhance the correlations, without the large peculiar velocities. This enhancement makes the cold particles actually work, as far as the agreement with ξ_g is concerned.

In all these simulations the correlation function of the mass is evolving rather rapidly both in slope and amplitude. Comparing the correlation functions today to that at the redshift 0.5 one can see significant evolution even over that small redshift range. From angular correlations $w(\theta)$ of very faint J=24 galaxies limits were obtained by Koo and Szalay[25] how ξ behaves at that redshift, and these limits are incompatible with the results of the simulations. Since ξ for the 'biased' galaxies is much more stable, the data do not rule out biased galaxy formation.

In the neutrino dominated universe the particles which have crossed the caustics at any given time are the ones associated with 'galaxies'. This selection also dramatically increases the correlation, making the model incompatible with observations. The particles selected this way contain all the regions, where galaxies may form, but not all these particles need be galaxies. Most of the contribution to the small scale correlation is coming from the regions where tight clumps are. If the rate of galaxy formation could be biased against those regions, the correlation could be somewhat decreased.

4. OBSERVATIONS OF THE LARGE SCALE STRUCTURE

Redshift surveys, filaments and voids

Redshift surveys seem to be the best way to determine the real distribution of galaxies in 3 dimensions. The first such surveys caused a lot of excitement, because they indicated, that galaxies are not uniformly distributed over space, but rather they occupy a few percent of the available volume. They are often found in long chain-like filaments, like the Perseus Supercluster. They leave behind large voids, like the 60 Mpc diameter one in Bootes. Such structures can easily arise in the neutrino dominated picture, but they are hard to form with cold dark matter. Presently there is not enough data to assess the statistical significance of the presence of the voids.

Cluster-cluster correlations

It has been known for some time, that Abell clusters have a similar correlation function, but much larger correlation length, than galaxies do[26]. It was realized only recently, how hard it is to explain this feature. The theory of rare Gaussian events[27], higher order correlation functions[28] are recent attempts at explanation.

However, the cluster correlation data have a peculiar property. The richer (thus rarer) clusters have an even higher correlation amplitude. Clusters are defined as the peaks of a given height in the galaxy distribution. The centers of these peaks are selected as the new point catalog, and the correlation is calculated. This selection resembles a renormalization transformation. Whenever systems of different physical dimensions are compared, it is preferential to use dimensionless quantities. The correlation function is dimensional, related to a length scale. The only natural length that appears in these point-catalogs is the mean separation, determined by the density. The value of the correlation function at this distance (i.e. expressing distance in units of the mean separation) is thus dimensionless. The surprising thing is, that this dimensionless number for the cluster data is 0.35, while for galaxies it is 1.1. Compared in this way galaxies are stronger clustered than the Abell clusters are.

If the bias to galaxy formation occurred in a scale-invariant way (at least on scales from a few Mpc and up), one would expect, that all dimensionless correlation amplitudes are equal. The slope of the correlation function would be related to the geometry of the pattern, essentially to its fractional (or 'fractal') dimension. Small scale gravitational clustering may break the scale invariance, and increase the dimensionless galaxy correlation amplitude, in agreement with the data. A detailed analysis will be published elsewhere.[29]

5. LY-α ABSORPTION SYSTEMS

In the spectra of high redshift QSO's many Ly-α absorption systems were found[23]. The typical characteristics of this Ly α-forest are narrow (10-40 km/s) absorption lines with typical neutral H column densities of $N_{HI} \approx 10^{13-14}$ cm^{-2}. The line width sets an upper limit to the temperature of the clouds, $T_{cloud} < 30000$ K. The neutral H number density in the clouds is about $10^{-3.5-4.5}$, about a 10 times overdensity at those redshifts. There are about 40 clouds in a unit redshift interval, between redshifts 2 and 3. This corresponds to a mean separation along the line of sight to more than 100 Mpc, but this is $\langle n R^2 \pi \rangle^{-1}$, the mean free path, not the real mean separation $\langle n \rangle^{-1/3}$. The cloud sizes are about 10-30 kpc,[24] which tells us that the comoving number density of clouds has to be 100 to 1000 times higher than that of bright galaxies. These clouds appear to be unclustered,

$$\xi_{clouds} < 0.1\ \xi_g.$$

If the universe is dominated by cold dark matter, this provides another clue, that we need 'biased' galaxy formation. If galaxies, when formed, were evenly distributed just as the Ly α clouds are, and their present correlation is due to their gravitational motion, one would expect the same correlation for the clouds. The clouds could not be destroyed sufficiently when near to a galaxy to explain the lack of correlations. If the galaxies are clustered anomalously 10 times stronger, then the clouds have just the ordinary clustering properties of the mass.

If the universe is neutrino-dominated, the hot gas in the pancakes has extremely high temperature and pressure, so the clouds cannot exist in those regions, the external pressure would compress them. The only place, where they could survive, would be in the voids. This would explain their low correlation, but it seems to be extremely difficult to regenerate the required small scale fluctuations within the voids.

The formation process of the Ly α clouds is still not clear, and the major uncertainty is the state of the intergalactic medium, its temperature and density. It can be shock heated, but then the release of the energy driving the shocks must be associated with explosions related to galaxy formation processes. It is not easy to see, why the correlations of Ly α clouds are so much weaker than the galaxy correlations in this case. The IGM can be photoionized, but in the latter case the source of the necessary photons is somewhat unclear.

6. CONCLUSION

All the present theories of galaxy formation fail to explain the observed universe in its full complication. The recent experimental development on the microwave background fluctuations provides the strongest constraints on the present theories yet. The details of the galaxy correlation properties are a new challenge, indicating that galaxies are unlikely to be tracers of the mass distribution. The particular process which will create a local 'biasing' of galaxy formation will be likely to be related to details of cooling and gas dynamics. The correlations of clusters of galaxies suggest, that whatever is the source of 'biasing', it is likely to be scale invariant over large scales. The origin and clustering properties of Ly α clouds are related to the formation of galaxies, and this relation needs further study.

7. ACKNOWLEDGEMENTS

I would like to acknowledge useful discussions with Simon White, Dick Bond, Nick Kaiser, David Koo and Dave Schramm.

8. REFERENCES

1) P.J.E. Peebles, The Large Scale Structure of the Universe, (Princeton University Press, Princeton, N.J. 1980).

2) Ya.B. Zeldovich, Astron. Astrophys. 5, (1970) 84; R.A. Sunyaev and Ya.B. Zeldovich, Astron. Astrophys. 20 (1972) 189.

3) A. Guth and S.-Y. Pi, Phys. Rev. Lett. 49 (1982) 1110; P. Steinhardt and M.S. Turner, Phys. Rev. D29 (1984) 2162; R. Brandenberger, this volume.

4) R. Brandenberger and R. Kahn, Phys. Rev. D29 (1984) 2172; J. Bardeen, P. Steinhardt, and M.S. Turner, Phys. Rev. D28 (1983) 679; R. Brandenberger, Rev. Mod. Phys. (1984) in press.

5) J.M. Uson and D.T. Wilkinson, Ap. J. Lett. 277 (1984) 1; J.M. Uson and D.T. Wilkinson, Proc. Inner Space/Outer Space Workshop, Fermilab (U. Chicago Press, 1984 in press).

6) J.R. Bond and G.P. Efsthathiou, Proc. Inner Space/Outer Space Workshop, Fermilab (U. Chicago Press, 1984 in press).

7) N. Vittorio and J. Silk, Ap. J. Lett. (1984) in press.

8) M. Guyot and Ya.B. Zeldovich, Astr. Ap. 9 (1970) 227; P. Meszaros, Astr. Ap.

9) S.S. Gershtein and Ya.B. Zeldovich, JETP Lett. 4 (1966) 174; R. Cowsik and J. McClelland, Phys. Rev. Lett. 29 (1972) 669; G. Marx and A.S. Szalay, Proc. Neutrino '72, 1 (1972) 123; A.S. Szalay and G. Marx, Astron. Astrophys. 49 (1976) 437; D.N. Schramm and G. Steigman, Ap. J. 243 (1981) 1; J.R. Bond, G. Efsthathiou, and J. Silk, Phys. Rev. Lett. 45 (1980) 1980; A.G. Doroshkevich, M.Yu. Khlopov, R.A. Sunyaev, A.S. Szalay and Ya.B. Zeldovich, Ann. N.Y. Acad. Sci. 375 (1980a) 32; A.G. Doroshkevich and M.Yu. Khlopov, Sov. Astron. 25 (1981) 521.

10)J.R. Bond and A.S. Szalay, Proc. Neutrino 81 1 (1981) 59; J.R. Bond and A.S. Szalay, Ap. J. 274 (1983) 443; J.R. Bond, A.S. Szalay, and M.S. Turner, Phys. Rev. Lett. 48 (1982) 1636.

11)P.J.E. Peebles, Ap. J. 258 (1982) 415; P.J.E. Peebles, Ap. J. Lett. 263 (1982) 1.

12)H. Pagels and J. Primack, Phys. Rev. Lett. 48 (1982) 223; G.R. Blumenthal, H. Pagels, and J. Primack, Nature 299 (1982) 37; G.R. Blumenthal, S.M. Faber, J.R. Primack, and M.J. Rees, Nature (1984) in press.

13)A. Vilenkin, Proc. Inner Space/Outer Space Workshop, Fermilab, (U. Chicago Press (1984) in press).

14)K.A. Olive, D.N. Schramm, G. Steigman, M.S. Turner, and J. Yang, Ap. J. 246 (1981) 557.

15)V.I. Arnold, S.F. Shandarin, and Ya.B. Zeldovich, Geophys. Astrophys. Fluid Dynamics 20 (1982) 111.

16)J.R. Bond, J. Centrella, A.S. Szalay, and J.R. Wilson, M.N.R.A.S. (1984) to be published; P.R. Shapiro, C. Struck-Marcell, and A. Melott, Ap. J. 275 (1983) 413.

17)A.G. Doroshkevich, S.F.Shandarin, E. Saar, M.N.R.A.S. 184 (1978) 643.

18)M.J. Rees and J. Ostriker, M.N.R.A.S. 179 (1977) 541; J. Silk, Ap. J. 211 (1977) 638.

19)M. Davis, G.P. Efsthathiou, C.S. Frenk, and S.D. M. White, submitted to Ap. J. (1984); C.S. Frenk, S.D.M. White and M. Davis, Ap. J. 271 (1983) 417; S.D.M. White, Proc. Inner Space/Outer Space Workshop, Fermilab (U. Chicago Press (1984) in press).

20)J. Centrella and A. Melott, Nature 305 (1983) 196; A.A. Klypin and S.F. Shandarin, M.N.R.A.S. 204 (1983) 891; A. Melott, M.N.R.A.S. 202 (1983) 595.

21)M. Davis and P.J.E. Peebles, Ap. J. 267 (1983) 465.

22)N. Kaiser, Ap. J. Lett. 273 (1983) 1.

23)P. Young, W.L.W. Sargent, and A. Boksenberg, Ap. J. 252 (1982) 10.

24)C.B. Foltz, R.J.Weymann, H.J. Röser, and F.H. Chaffee, Ap. J. Lett. 281 (1984) 1.

25)D.C. Koo and A.S. Szalay, Ap. J. 282 (1984) 390.

26)P.J.E. Peebles, and M.G. Hauser, Ap. J. Suppl. 28 (1974) 19; N.A. Bahcall and R.M.Soneira, Ap. J. Lett. 258 (1982) 17.

27)N. Kaiser, Proc. Inner Space/Outer Space Workshop, Fermilab (U. Chicago Press (1984) in press).

28)J.N. Fry, Proc. Inner Space/Outer Space Workshop, Fermilab (U. Chicago Press (1984) in press).

29)A.S. Szalay and D.N. Schramm, (1984) in preparation.