PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Galaxy Formation: The First Million Years [and Discussion]

J. D. Barrow and C. J. Hogan

Phil. Trans. R. Soc. Lond. A 1980 **296**, 273-288
doi: 10.1098/rsta.1980.0170

Phil. Trans. R. Soc. Lond. A **296**, 273–288 (1980) [273]
Printed in Great Britain

Galaxy formation: the first million years

By J. D. Barrow
Department of Astrophysics, University Observatory, South Parks Road, Oxford OX1 3RQ, U.K.

An overview is given of the non-equilibrium behaviour conceivable during the first million years of the Universe's life. We describe various classical and quantum processes of this type which might constrain and fashion the irregularity spectrum on large cosmic scales. The likely fate of adiabatic, isothermal and vortical fluctuations is traced through the radiation dominated phase of a hot, big-bang Universe. Both the linear and nonlinear behaviour of these fluctuations is discussed and it is indicated how different preferred mass scales and irregularity spectra might be expected to arise in association with them.

'Singularity is almost invariably a clue. The more featureless and commonplace a crime is, the more difficult . . .' (Sherlock Holmes: *The Boscombe Valley mystery*, by Arthur Conan Doyle).

1. Ends and beginnings

One of the most longstanding and outstanding problems of modern cosmology is to account for the structural organization in the Universe as manifested by the presence of stars, galaxies and galaxy clusters. Unfortunately a conspiracy of gravitational and hydrodynamical complexity forestalls any direct analytical attack upon this problem in its full generality. Instead a combination of approximation, intuition and faith must be employed to isolate those physical processes that could be predominantly responsible. From this pursuit, two paradigms have gradually emerged and generally polarized theoretical approaches to the subject. On the one hand, those of the fragmentation persuasion pursue the idea that the early proto-structures surviving the radiation era were at least of cluster size, exceeding *ca.* $10^{13}\ M_{\odot}$. Subsequently, substructures on galactic scales and below are envisaged to develop via fragmentation of these clusters, the details being principally determined by scale-dependent, non-gravitational processes. By way of contrast, those of the aggregation persuasion seek to demonstrate that the larger structures were constructed in hierarchical fashion from smaller subunits of *ca.* $10^{5}\ M_{\odot}$, say, in a scale-independent manner by gravitational instabilities alone. An extremely promising statistical tool that may eventually discriminate between these contrasting ideas is the analysis of galaxy clustering with scale. The absence of structure in the two-point correlation function over a large range of mass scales within which clustering occurs has been interpreted by some as suggesting that gravitational instability and aggregation is the dominant mechanism at work (Peebles & Groth 1975; Groth & Peebles 1977). However, the fragmentation scheme has produced extremely convincing explanations for the characteristic masses and sizes of galaxies (Rees & Ostriker 1977; Silk 1977; Binney 1977).

Regardless of which of these approaches is correct it seems inevitable that physical processes operating during the first million years of the Universe's life will have had a significant effect upon the irregularity spectrum emerging at the recombination era $t \approx 10^{13}$ s. It is the aim of this paper to review some of these processes and in order to provide a concrete cosmological framework within which to do so we shall confine our attention to deviations from a hot,† isotropically expanding, general relativistic space–time – the Friedmann Universe. Observations of light elements around the galaxy together with the presence of the microwave background radiation provide evidence for its validity, and there are additional thermodynamical reasons for believing that the Universe was at least as isotropic as it appears to be today on large scales at all times in the past (Barrow & Matzner 1977; Barrow 1978; Penrose 1979).

Within this context we shall briefly examine a number of potentially interesting non-equilibrium processes of both a gravitational and non-gravitational nature in the radiation era. In this way we may discover if such processes are capable of imprinting a well defined structure upon the matter distribution in an expanding model Universe. In some cases these ideas have been worked out in a fairly detailed fashion in the research literature and are moderately well known, while others are more recent, sketchy or conjectural in nature.

Unfortunately, as yet, the deep insights obtained into the development of structure within physical systems far from equilibrium by Prigogine and his coworkers (Glansdorff & Prigogine 1971) have not been extended to gravitational systems in a rigorous way. A promising attempt along these lines has recently been made by Liang (1978*a*) and will doubtless be followed up in the future.

Naïvely speaking, the distortions from the isotropic Friedmann metric of interest to us may be decomposed at the linear level into scalar (density), vector (vortical) and tensor (gravitational wave) modes.‡ All of these metric potential fluctuations can ultimately lead to structural organization in an otherwise featureless self-gravitating system. We shall look sequentially at the influence of quantum and classical processes having gravitational or material origin in an attempt to elucidate the likely residue of such influences upon the primordial fluctuation spectrum.

2. Gravitational processes

2.1. *Quantum effects*

Classical gravitation physics begins at a time $t_{\text{pl}} \sim (Gh/c^5)^{\frac{1}{2}} \approx 10^{-43}$ s after the singularity; before this time the gravitational field takes upon itself a quantum nature. Nothing is known with any confidence of physics during this era, although some purely dimensional considerations can in fact make interesting predictions about irregularities. It seems, for example, that certain preferred irregularity spectra may even be imprinted upon the global structure of the Universe at these remotely early times by spontaneous vacuum fluctuations in the gravitational field energy. In this quantum era the geometry everywhere experiences unavoidable, zero-point fluctuations on the Planck length scale, $\lambda_{\text{pl}} \sim (Gh/c^3)^{\frac{1}{2}} \sim 10^{-33}$ cm. Wheeler (1957) was the first to sketch how we might dimensionally estimate the most probable form of such fluctuations.

† The remarks that I shall make concerning isothermal fluctuations have direct analogues in cold Universes, which have been studied for example by Carr (1977). However, adiabatic perturbations are not relevant to such models nor, one suspects, will vortical motions be important since in the absence of radiation they rapidly decay with the expansion, $v \propto R^{-1}$.

‡ No exact, inhomogeneous, radiation-filled cosmological solutions to Einstein's equations are yet known and so approximation methods must be employed to follow the early evolution of irregularities. Fortunately, techniques in functional analysis exist to tell when approximate solutions correspond to small parameter limits of true nonlinear solutions to Einstein's equations.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES THE ROYAL SOCIETY A PHILOSOPHICAL TRANSACTIONS OF

Whereas classical variations in the Hilbert action $S = (c^4/16\pi G) \int R(-g)^{\frac{1}{2}}\,d^4\boldsymbol{x}$ are required to vanish ($\delta S = 0$), for geometries to be allowed, this is no longer so in the quantum picture. All geometrical configurations are now possible, only some are more probable. Feynman's approach associates a probability amplitude $\exp(iS/h)$ with any given geometry. A functional propagator $K\{G_2, \sigma_2; G_1, \sigma_1\}$ then gives the probability of passing from one geometrical configuration G_1 on a hypersurface σ_1 to G_2 on another σ_2 as $\sum \exp(iS/h)$; the sum being taken over all 'histories', that is over all various possible geometries consistent with the specified initial and final states. Now, dimensionally, since $R \approx \partial\Gamma + \Gamma^2$ and $\Gamma \sim \partial g$ we have $S \sim \lambda^4 O\{\partial^2 g/\partial\lambda^2, (\partial g/\partial\lambda)^2\}$. Using the fact that the most significant contributions to the sum over histories arise at the poles $\delta S \sim \frac{1}{2}\pi$, we can estimate the most probable metric fluctuation spectrum as

$$\delta g \equiv \Delta \sim \frac{\lambda_{\rm pl}}{\lambda} + \left(\frac{\lambda_{\rm pl}}{\lambda}\right)^2 \sim \left(\frac{M_{\rm pl}}{M}\right)^{\frac{1}{3}} + \left(\frac{M_{\rm pl}}{M}\right)^{\frac{2}{3}}$$

over scales $\lambda > \lambda_{\rm pl}$. The latter term represents the uncertainty in the potential $g \sim GM/\lambda$ over a Compton wavelength $\lambda \sim h/Mc$.

2.2. *Classical considerations*

Metric perturbations may be decomposed into scalar, vector and tensor modes as first shown in Lifshitz's classic paper of 1946. Conversely, we can associate with any fluctuation in density, spin or shear anisotropy a corresponding metric distortion. For density perturbations $\rho \to \rho + \delta\rho$, this metric distortion to the space–time geometry is

$$\Delta \sim \frac{\text{gravitational potential of fluctuation}}{\text{gravitational potential at horizon}} \sim \frac{G\delta\rho\lambda^2}{G\rho(ct)^2} \sim \left(\frac{\delta\rho}{\rho}\right)\lambda^2$$

for a perturbation with overdensity $\delta\rho$ and scale λ at fixed time t; the metric gravitational potential is *ca.* GM/R.

Analogous metric perturbations may be associated with vortical and shear perturbations. Harrison (1970) and Zeldovich (1973) first pointed out the preferred nature of so-called *constant curvature fluctuations* having scale-independent Δ with $\delta\rho/\rho \propto \lambda^{-2} \propto M^{-\frac{2}{3}}$ for fluctuations in the total density. In this case, every scale possesses the same overdensity when it first enters the horizon, whereas in general a fluctuation spectrum of the form $\delta\rho/\rho \propto M^{-\alpha}$ contributes metric distortions of amplitude $\Delta \sim M^{\frac{2}{3}-\alpha}$. Thus if α exceeds $\frac{2}{3}$ then $\Delta \to \infty$ as $M \to 0$ (although we might imagine a cut-off at $M_{\rm pl} \sim (hc/G)^{\frac{1}{2}} \sim 10^{-5}$ g) and a chaotic, non-Friedmannian initial state is required when the smallest scales enter the horizon, rendering the small perturbation analysis inconsistent there. Now, if α is less than $\frac{2}{3}$, for example $\alpha = \frac{1}{2}$ corresponds to random, uncorrelated 'white noise' of statistical fluctuations in an equilibrium ideal gas, then although the space–time is initially unperturbed by the fluctuations it will ultimately become non-Friedmannian owing to gravitational effects of the fluctuations on large scales if it expands forever. The intermediate 'Zeldovich hypothesis' with $\alpha = \frac{2}{3}$ enables small amplitude density fluctuations to have a small gravitational effect upon the space–time at all times and weights all mass scales equally. It is thus essentially a hypothesis of cosmic simplicity ensuring that the metric perturbations to the Friedmann background are always negligible if they are ever negligible. This covers the spatial distribution of $\delta\rho/\rho$ at fixed t; let us now briefly consider the time evolution.

It is now well known that density perturbations are only *linearly* unstable in time in an expanding medium (Lifshitz 1946; Bonnor 1957). In general, for deviations from a perfect fluid background with equation of state $p = (\gamma - 1)\,\rho$ and t comoving proper time, we have, for $\lambda > ct$,

$$\delta\rho/\rho = A(\boldsymbol{x})\; t^{2(3\gamma-2)/3\gamma} + B(\boldsymbol{x})\; t^{(\gamma-2)/\gamma}.$$

The growing 'A' mode originates from variations in the spatial curvature from place to place while the decreasing 'B' mode derives from non-simultaneity of the big-bang and spatially varying anisotropies in the expansion flow. All mass scales are predicted to grow at the same rate while their size exceeds that of the horizon. Thus an initial mass spectrum $\delta\rho/\rho \sim (M/M_0)^{-\alpha}$ imprinted at t_i in the radiation era would have amplitude $\delta\rho/\rho \sim (M/M_0)^{-\alpha}\, t/t_i$ at subsequent times on large scales. Although $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ are arbitrary functions of scale in this expression, a deeper analysis reveals that not all such linearized solutions really correspond to genuine nonlinear solutions of the field equations. For example, an open background universe would be incompatible with choices of A and B that required so much matter on large scales that compact space sections would result. This 'linearization stability' problem of ensuring that the perturbation modes close to exact solutions with symmetries really do correspond to linearizations of a true nonlinear solution has been investigated for Friedmann universes by D'Eath (1976) who deduces various general constraints upon $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ for this to be so.

A long-standing controversy is the question of whether the condensations now observed on galactic scales with large amplitudes ($\delta\rho/\rho \gg 1$) could have attained such overdensities by enhancing only linearly in time from a purely statistical level, say $(\delta\rho/\rho)_{\mathrm{init}} \propto N^{-\frac{1}{2}} \propto M^{-\frac{1}{2}}$, when beginning their growth at a 'reasonable 'starting time. It is clear, of course, that the necessary fluctuations can grow from such amplitudes as required so long as they start early enough (Peebles 1968), but the real logical problem is that by such a methodology one can 'explain' *any* present amplitude. It is therefore of questionable information content. For example, 'white noise' ($\delta\rho/\rho \propto N^{-\frac{1}{2}}$) arising at $t_{\mathrm{pl}} \sim 10^{-43}$ s predicts $\delta\rho/\rho \gg 1$ on galactic scales at red shifts of *ca.* 10, thus yielding only black holes today, while if these fluctuations arise at the Compton time $t_c \sim h/Mc^2 \sim 10^{-23}$ s, we obtain only $\delta\rho/\rho \ll 1$ today on galactic scales. Harrison (1970) has argued that constant curvature fluctuations arising at the Planck time would be just right, with $\delta\rho/\rho \sim 1$ on galactic scales by $z \sim 10$. Many commentators have felt that the slow growth problem sprang from the lack of an instability growing exponentially with time; however, it should be clear that the presence of such an exponentially amplifying mode would be catastrophic for the theory. Initial fluctuations could not have exceeded the level $\exp(-10^{16})$ or all galactic scale overdensities would have long since evolved into black holes!

The evolution of small rotational velocity perturbations in the matter distribution of an isotropically expanding flow can be traced by using the conservation of angular momentum; thus $I\omega \sim (\rho R^3)\, R^2\omega \sim \rho R^5\omega$ is constant in the absence of non-equilibrium processes. If the rotation is large ($\omega^2/\rho \gtrsim 1$), a significant metric distortion and nonlinear coupling with the shear motions will occur.

Other more exotic processes of a purely gravitational nature deserve a brief mention since they have attracted some interest recently. Liang (1978*a*) has attempted to formulate the statistical mechanics of a gas of self-gravitating particles, after early attempts by Terletskii (1955), Genkin (1970), Bonnor (1956) and Saslaw (1968). In naïve terms, these earlier studies attempted to improve upon the kinetic theory for an ideal gas by including the long range gravitational attraction of the molecules. This inclusion yields a 'non-ideal' gas law,

$$PV = NkT - \beta V^{-\frac{1}{3}},$$

similar to the relation for a van der Waals gas. This system exhibits an apparent phase transition when $\partial p/\partial V = 0$, although this is not a true critical point since $\partial^2 p/\partial V^2$ is non-zero there. Although larger fluctuations might be expected there ($(\partial\rho/\rho)_c \sim M^{-\frac{1}{6}}$), they appear to diverge elsewhere and the analysis is of doubtful physical meaning. Liang has adopted a more rigorous approach to assessing the gravitational contributions to the thermodynamics. He calculates the most probable irregularity spectrum that minimizes the free energy under assumptions of stationariness and small deviations from homogeneity. Remarkably, he was able to predict a most probable baryon irregularity spectrum as $(\partial\rho/\rho)_b \sim (M/M_{Jb})^{-\frac{1}{3}}$, where M_{Jb} ($\sim 10^5\, M_\odot$) is the matter Jeans mass during the radiation era. This is virtually identical to the spectrum deduced from analyses of the observed correlation and multiplicity functions for galaxies by Gott & Rees (1975) over a wide range of scales, both in slope and amplitude. Unfortunately, a fairly unmotivated screened potential was introduced to incorporate the many-body effects accurately and the end result is quite sensitive to its form, but future work may be able to give this idea a firmer foundation. This approach to the problem has strong links with the work of Prigogine and his coworkers, who have long stressed the fact that many of the organized structures familiar to us in Nature are not consequences of equilibrium fluctuations at all but rather are due to the development of non-equilibrium order – so-called 'dissipative structures'. Our intuition is familiar with the destruction of clustering in the neighbourhood of thermal equilibrium, so, conversely, structural organization might be expected far from equilibrium as a result of the competition between entropy maximization and energy minimization within a physical system.

3. Material processes

3.1. *Quantum effects*

The behaviour of quantum fields in a classical gravitational field is a topic that has received much attention since Parker (1968) and Sexl & Urbantke (1967) revived the question of super-adiabatic wave amplification in an expanding Universe investigated first by Schrodinger (1939). Particularly important was the realization by Zeldovich (1970) that particle creation could occur very efficiently in anisotropically expanding universes at the expense of the 'anisotropy energy'. This occurs because they possess an ellipsoidal gravitational potential which is repulsive along directions that implode as the whole volume expands. Despite this realization, no estimates have yet been made of its likely effect upon the inhomogeneity spectrum. Approximately, we expect the dominant particle production to occur in modes with frequency less than the cosmic expansion rate ($\nu < t^{-1}$), where the gravitational field changes rapidly enough to prevent virtual pairs reannihilating into the vacuum. If the production begins at t_i then the resulting number density of gravitons produced is estimated by integrating over the modes between ν and $\nu + \delta\nu$,

$$n_g \sim \int_0^{t_i^{-1}} \nu^2 \, d\nu \sim t_i^{-3} \sim t_{pl}^{-3},$$

and thus the energy density is $\rho_g \sim h/c^5t_i^4$. This greatly exceeds the ambient density (*ca.* $(Gt_i^2)^{-1}$ for $t \lesssim t_{pl} \sim (Gh/c^5)^{\frac{1}{2}}$), and indicates that strong physical effects should be expected as a result of this process.

Zeldovich & Starobinskii (1978) have noticed that the rate of graviton production in slightly anisotropic universes is explicitly linked to the Weyl tensor invariant $C_{ijkl}C^{ijkl}$ via

$$(-g)^{-\frac{1}{2}}\frac{d}{dt}((-g)^{\frac{1}{2}}n_g) = \left(\frac{1}{960\pi}\right)C_{ijkl}C^{ijkl}.$$

These results allow one to summarize the important consequences of these quantum processes as follows: gravitational wave energy producing anisotropy in the expansion dynamics ($\sigma^2 \sim R^{-6} \sim t^{-2}$) near a singularity can be dominated by the stresses contributed by more energetic created particles $\rho_g \sim t^{-4}$ near $t \sim 0$. In addition, it seems that the gravitons are created at rest relative to the local fundamental observers and have a dramatic damping effect upon any primordial vorticity spectrum. Since the inertia of the created particles (*ca.* t^{-4}) so greatly exceeds that of any initially present (*ca.* t^{-2}), those moving with high velocities must be dramatically slowed to conserve linear and angular momentum (Lukash *et al.* 1975; Barrow 1977). This creates a severe problem for the turbulent scenario of galaxy formation unless the equation of state is very stiff near the singularity (Barrow 1977). As adiabatic fluctuations carry velocity perturbations with them, we might also expect an indirect constraint upon them by this mechanism. In addition, since the created graviton energy exceeds that contained in curvature or shear fluctuations ($\delta\rho/\rho$), near the singularity we would expect a significant effect upon any primordial inhomogeneity spectrum. With a creation rate directly dependent on the Weyl curvature invariant we might expect production of gravitons to occur preferentially in regions where this conformal curvature is larger, suggesting that constant curvature fluctuations would finally result. This would require the particle creation to be non-local or else overdensities in gravitational wave energy, for example, would merely be transferred into corresponding overdensities in created particles.

3.2. *Classical processes*

The gravitational instabilities described in §2.2 will, of course, only occur on mass scales large enough to allow the gravitational force to overcome the influences of both pressure and material transport processes. Since the Jeans mass at the end of the radiation era is extremely large (*ca.* $10^{15}\,\Omega^{-2}\,M_\odot$, where Ω is the total density in units of the critical density), we might expect that these non-gravitational processes were extremely significant participants in the fashioning of precursors to galactic or cluster-sized structures. Here I shall try to summarize the main ideas concerning the influence of such processes upon a fairly arbitrary initial spectrum of irregularities (for more details see Rees 1971; Jones 1970, 1976). We imagine such a spectrum to contain principally a combination of adiabatic, isothermal and vortical modes; gravitational wave modes are ignored since observations indicate their effects to be small. In the interests of simplicity I shall consider the fate of these modes separately although I shall not confine their amplitudes to lie only in the linear régime.

One feature of entirely relativistic origin that is common to all these fluctuation modes is the role of the particle horizon. Roughly, the presence of an initial space–time singularity indicates that at any subsequent time t (measured in seconds), regions separated by a scale encompassing a mass exceeding $M_H \sim 10^5\, tM_\odot$ are causally incoherent. No classical transport processes occurring at time t can erase inhomogeneities on scales exceeding M_H.†

Let us now consider the fates of the three above-mentioned fluctuation types to discover if preferred mass scales and irregularity spectra can be imprinted upon the Universe during the radiation era.

3.2.1. *Isothermal fluctuations*

3.2.1.1. *Linear régime.* These are often termed 'entropy' fluctuations since they are associated with variations in the baryon density, $(\delta\rho_b \neq 0)$, in a smooth radiation bath, $(\delta\rho_\gamma = 0)$ and this results in spatial variations of the specific radiation entropy $S_b \sim T_\gamma^3 \rho_b^{-1} \sim \rho_\gamma^{\frac{3}{4}} \rho_b^{-1}$. However, although the radiation field is therefore not participating directly in the build-up of inhomogeneities it has two very strong indirect effects which almost entirely govern their behaviour once they enter the horizon. First, matter fluctuations smaller than the Jeans mass, which is $0.2\, M_H$ in the radiation era when the sound speed is *ca.* $\frac{1}{\sqrt{3}}c$, find themselves embedded in an extremely viscous medium. Peebles (1965) first pointed out that if an electron within an overdensity were moving at velocity v relative to the isotropic radiation background it would feel a radiation flux *ca.* aT^4v in its direction of motion and would eventually be stopped. The drag force is $F_d \sim -(v/c)\, aT^4/\lambda$, where $\lambda \sim (\sigma n_e)^{-1}$ is the scattering length. Electrostatic forces ensure that the electrons move with the protons and the gravitational force acting upon them is $F_g \sim GMm_p/L^2 \sim Gm_p\,(\rho L^3)/L^2$ and thus the ratio of the two competing forces is $|F_d/F_g| \sim \sigma_T T^4/Gm_p ct \sim 10^{-8}(1+z)^{\frac{5}{2}}$ for the $\Omega = 1$ Friedmann model. Thus, before recombination, $z \gtrsim 10^3$, $F_d > F_g$ and the matter fluctuations are 'frozen' into the radiation sea, unable to enhance or attenuate. At the end of the radiation era an isothermal spectrum will directly mirror the amplitudes present when scales less than *ca.* $10^{15}\, M_\odot$ entered the horizon. Secondly, it has been pointed out in different contexts by several authors (Chernin 1973; Meszaros 1975; Guyot & Zeldovich 1970) that small baryon fluctuations are unable to grow or collapse while the dynamics are radiation-dominated because the expansion takes place too rapidly. For, when ρ_γ exceeds ρ_b the free-fall time in the matter, $t_{mf} \sim (G\rho_b)^{-\frac{1}{2}}$, greatly exceeds the expansion time (*ca.* $(G\rho_\gamma)^{-\frac{1}{2}}$). This constraint is quite distinct from that mentioned above, being of gravitational rather than electromagnetic origin, and would even constrain the development of isothermal modes originating from number fluctuations in a population of small black holes (Meszaros 1975; Carr 1976).

If larger scales have smaller initial density amplitudes, the first isothermal modes to become gravitationally unstable at the end of the radiation era will be *ca.* $10^5\, \Omega^{-\frac{1}{2}}\, M_\odot$, the matter Jeans mass, which remains constant during the radiation era. This fact was first exploited by Peebles & Dicke (1968), who argued that it could explain the existence of globular star clusters which have approximately this mass. Others, in particular Press & Schecter (1974), have appealed to the nonlinear statistical clustering of objects of this size to explain the development of structure on much larger scales.

Gott & Rees (1975) have claimed that if one adopts the constant curvature hypothesis, that fluctuations have the same *total* overdensity $(\delta\rho/\rho)_T$ on entering the horizon scale, then this predicts a characteristic *baryon* density spectrum emerging at recombination resulting from a purely isothermal fluctuation spectrum. In the radiation era, $\rho_T \sim \rho_\gamma$, but $\delta\rho_T = \delta\rho_b$

† It is not known whether the presence of particle horizons is a generic feature of space–time, but it does not appear to be generic to spatially homogeneous cosmologies.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

since $\delta\rho_\gamma \equiv 0$, so $\delta\rho_b/\rho_b = (\delta\rho/\rho)_T\,(\rho_T/\rho_b) = (\delta\rho/\rho)_T\,R^{-1} \propto M^{-\frac{1}{3}}$ because $(\delta\rho/\rho)_T$ is constant on all scales when they enter the horizon, by the hypothesis. Thereafter it is frozen in by the radiation drag until recombination when only those smaller than *ca.* $10^5\,\Omega^{-\frac{1}{2}}\,M_\odot$ can be supported by gas pressure alone. Those exceeding this scale are then gravitationally unstable. It is intriguing to notice that this predicts a spectral form similar to that deduced by Gott & Rees (1975) from observations of galaxy clustering,

$$\left(\frac{\delta\rho}{\rho}\right)_{\rm rec} \sim 10^{-4}\left(\frac{10^{17}\,M_\odot}{M}\right)^{\frac{1}{3}} \sim \left(\frac{10^5\,\Omega^{-\frac{1}{2}}\,M_\odot}{M}\right)^{\frac{1}{3}} \sim \left(\frac{M_{\rm Jb}}{M}\right)^{\frac{1}{3}}$$

for $M < 10^{15}\,\Omega^{-2}\,M_\odot$, where $M_{\rm Jb}$ $(\sim 10^5\,\Omega^{-\frac{1}{2}}\,M_\odot)$ is the matter Jeans mass. This relation would suggest that the observed fluctuations have a statistical origin and are of an isothermal nature since $M_{\rm Jb}$ is epoch dependent in the radiation era and so plays the role of an invariant scale.

These isothermal perturbations will carry with them no associated velocity perturbations because of the radiative drag and therefore require galaxies to gain any rotation via tidal effects. Unfortunately, though, small isothermal fluctuations cover their tracks very effectively. They produce only very small changes in the primordial synthesis of light elements (Epstein & Petrosian 1975) and it is impossible to extricate the consequences of these small baryon fluctuations from those of anisotropies or background density changes. The gravitational 'fingerprints' that should have been left upon the fine scale structure of the 3 K radiation (unless other sources of opacity have subsequently erased them) differ slightly from those predicted by adiabatic variations and might eventually enable one to decide if this component is the dominant part of the primordial irregularity spectrum.

3.2.1.2. *Nonlinear régime.* It has often been suggested that star formation is primordial, or at least that some population of explosive stars formed early in the Universe's history before galaxy formation, so enabling pregalactic nucleosynthesis to occur. This idea has received considerable backing from recent evidence for an intergalactic iron abundance at almost a solar level (Mitchell *et al.* 1976). Such stars might arise naturally if isothermal fluctuations had very large $(\delta\rho_b/\rho_b \gg 1)$ primordial amplitudes. This idea is also computationally very appealing because the early radiation domination of the dynamics enables one to trace the evolution of matter fluctuations of huge amplitude, up to $\delta\rho_b/\rho_b \lesssim 10^8\,(M_\odot/M)^{\frac{1}{2}}$ without having to worry about nonlinear gravitational effects and perturbations of the background expansion. The most detailed analysis to date is that of Hogan (1978) and it reveals quite a variety of possible evolutionary tracks for clumps according to their position in the 'mass–overdensity plane'. Briefly, there are two key processes that control the fate of matter fluctuations with large $(1 \lesssim \delta(\rho/\rho)_b \lesssim 10^8\,(M_\odot/M)^{\frac{1}{2}})$ amplitude and they are important over scales delineated by the two characteristic masses that are time-independent during the radiation era, the matter Jeans mass, $M_{\rm Jb}$, and the Jeans mass determined by radiation pressure support, $M_{\rm J\gamma} \sim (P_\gamma/P_{\rm gas})^{\frac{3}{2}}\,M_{\rm Jb} \sim 10^{17}\,M_\odot$. Clouds smaller than $M_{\rm Jb}$ with large amplitude have sufficient internal pressure to inflate them slowly. This continues until radiation has had sufficient time to random walk into the interior where it then contributes the predominant inflationary pressure; the clump then rapidly disperses. Clouds larger than $M_{\rm Jb}$ but smaller than $M_{\rm J\gamma}$ need to be hotter than the ambient radiation background to provide radiation pressure support against their own self-gravity. However, eventually this radiation will be able to diffuse out of the cloud into the cooler background and the cloud will subsequently collapse. Hogan (1978) has traced these evolutionary features in some detail and points out how the large abundances of heavy elements synthesized within these isothermal inhomogeneities might enable observational restrictions to be imposed upon their existence. For example, regions with $(\delta\rho/\rho)_b \sim 10^5$ would generate z ($\gtrsim$ Mg) $\sim 3\times10^{-5}$ during primordial nucleosynthesis. Considerable dilution of the deuterium abundance would also undoubtedly occur on average and might be problematic if the observed interstellar abundance, $X(\mathrm{D}) \sim 10^{-5}$, is primordial.

A more detailed examination of this scenario tracing the recombination history and the early phases of star formation would be extremely interesting.

3.2.2. *Adiabatic fluctuations*

3.2.2.1. *Linear régime.* If a spectrum of small amplitude adiabatic fluctuations ($\delta\rho_b/\rho_b = \delta\rho_\gamma/\rho_\gamma < 1$) were present before recombination, then two separate dissipative effects could erase parts of this spectrum. In the earliest stages, $z > 10^4$, when the fluctuations are totally radiation-dominated, a radiative viscosity arises because of the slight anisotropy in the radiative pressure induced by the tidal field of the density irregularity. This viscosity removes the velocity perturbations and thereby the associated potential fluctuations. At the end of the radiation era when fluctuations are becoming matter-dominated, diffusive damping becomes important and is now able to attenuate overdensities via dissipation of their associated temperature perturbations. This effect arises because photons tend to random walk out of regions of high pressure when they no longer significantly contribute to the inertia. This erases high pressure regions when the diffusion time therein is less than the age of the Universe. If the linear extent of a fluctuation is N 'steps' then in general N^2 'steps' will be required for a photon to escape by a random walk. The diffusion time, t_d is thus

$$t_d \sim n_s\, t_{\mathrm{scatt}} \sim \left(\frac{\lambda}{L_1}\right)^2 \frac{L_1}{c} \sim \lambda^2 (L_1 c)^{-1},$$

where the number of scatterings is n_s, the clump size is λ and the path length of unit optical depth $L_1 = (\sigma_T n_e)^{-1}$. Thus the damping time varies with the square of the fluctuation size and all clumps smaller than $\lambda_D \sim (ctL_1)^{\frac{1}{2}}$ in scale are damped. This corresponds to mass scales $M_D \lesssim 10^{12}\, \Omega^{-\frac{5}{4}}\, M_\odot$ (Silk 1968; Peebles & Yu 1970; Chibisov 1972). A rigorous calculation increases M_D slightly by accurately incorporating the long mean free path effects which occur just before recombination when the mean free path approaches the horizon size. The proximity of M_D to the mass of small galaxy clusters rather than, say, of galaxies has led many to suggest that galaxies form from the fragmentation of cluster-sized objects of mass M_D, the smallest adiabatic modes surviving the radiation era. This scheme is quite well developed and has many appealing features (Doroshkevich *et al.* 1974; Rees 1977).

It has been pointed out by Sunayev & Zeldovich (1970) that a discontinuous jump in the amplitude of adiabatic fluctuations smaller than the Jeans mass should take place at recombination. Considering the flow conservation for the perturbed fluid gives us the following relation between velocities v and overdensity amplitudes δ ($\equiv \delta\rho/\rho$) before ($-$) and after ($+$) decoupling of matter from the radiation, $v_+/v_- \propto \lambda/c_s t \propto \delta_+/\delta_-$. For the velocity field to remain continuous we expect an amplification $\delta_+/\delta_- \propto \lambda^{-1} \propto M^{-\frac{1}{3}}$. Alternatively, we may view this in the following way. If adiabatic fluctuations obey the constant curvature hypothesis, then their associated velocity fields will be constant as each scale enters the horizon

$\delta\rho/\rho \sim v/c$ and remain frozen in by the radiation background. As recombination occurs, the matter is freed from the grip of the radiation field and can move unencumbered a distance *ca.* vt_{rec} during recombination. These motions will generate baryon irregularities, varying with scale $\lambda \sim vt_{\text{rec}}$ as $\delta\rho_b/\rho_b = 3\delta\lambda/\lambda \sim vt_{\text{rec}}/\lambda \propto M^{-\frac{1}{3}}$. Gott & Rees (1975) argue that this spectrum should extend up to the Jeans scale before decoupling, $M_J \sim 10^{15}\,\Omega^{-2}\,M_{\odot}$, but with two breaks expected in it: one at the mass M_D, below which adiabatic irregularities are diffusively attenuated by a factor *ca.* $\exp\{-(M_D/M)^{\frac{2}{3}}\}$, and the other at $M_{J,\,\text{rec}} \sim 10^{15}\,\Omega^{-2}\,M_{\odot}$ above which the primordial constant curvature form $\delta\rho/\rho \propto M^{-\frac{2}{3}}$ would be preserved since these large scales avoid the action of transport processes during the pre-recombination era.

3.2.2.2. *Nonlinear régime.* There is again, of course, no *a priori* reason why the adiabatic fluctuation spectrum should be of such an amplitude as to avoid the influence of nonlinear processes on interesting mass scales unless it has our computational ineptitude in mind. There are two particular nonlinear processes that we might expect to be of importance if adiabatic fluctuations attain an amplitude exceeding about 5% before entering the horizon. The first and most drastic of these is the possibility of forming primordial black holes (Zeldovich & Novikov 1966; Hawking 1971), and is specifically associated with large amplitudes in the A-mode curvature fluctuations of the linear theory.†

Gravitational collapse of objects within the Universe should have occurred when metric potential fluctuations were sizeable; in particular, adiabatic overdensities exceeding *ca.* 5% on a scale above the Jeans length *ca.* $\frac{1}{\sqrt{3}}ct$ but less than the horizon size *ca.* ct would suffice. A complete picture of all the possibilities would be rather lengthy; see Carr (1975) and Barrow & Carr (1978) for detailed discussions. The main conclusions may be summarized as follows: if black holes are not to be overproduced on either very small or very large scales, constant curvature fluctuations $\delta\rho/\rho \propto M^{-\frac{2}{3}}$, with constant metric distortion on all scales, are required. The fraction of the cosmic matter density allowed to undergo black hole collapse at early times $(10^{22} > z > 10^{4})$ is very small (Zeldovich & Novikov 1966). If a fraction f of the total mass energy is allowed to collapse at z_* then $f \lesssim 10^{5}\,z_*^{-1}$ when $z_* \lesssim 10^{22}$. This constraint arises since once the radiation energy density falls inside the holes it is no longer red-shifted away as rapidly as the external radiation background; rather it evolves as a pressureless component whose current value is constrained by the allowed age of the Universe compatible with the radioactive age of the Earth. Analogous ideas can be used to constrain the possibility of black hole formation after $z \sim 10^{4}$ (Barrow & Silk 1979; Carr 1978). Finally, statistical clustering of small holes may be an important factor in creating potential wells into which protogalactic material can later fall (see, for example, the scenario of White & Rees 1978). If a spread of hole masses is present in a number density fluctuation then the largest hole in a cluster will have the predominant effect and capture material by its gravitational Coulomb field (Ryan 1972; Meszaros 1975; Carr 1977; Barrow & Carr 1978). No characteristic mass scales really emerge from these solely gravitational considerations, although other speculative considerations associated with anisotropies or the equation of state in the hadron era admit interesting possibilities (Barrow & Carr 1978).

The second nonlinear adiabatic process of possible relevance to galaxy formation is less exotic but considerably more complex. If a curvature fluctuation has an overdensity, $\delta\rho/\rho$,

† It also seems possible for primordial black holes to form indirectly via classical accretion and quantum particle creation processes around white holes or regions of delayed bang-time associated with B-mode fluctuations (Eardley 1974; Lukash *et al.* 1975; Lake & Roeder 1976).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

of order 1 %, say, when it is encompassed by the particle horizon, so that it is not large enough to undergo gravitational collapse before pressure forces intervene (or alternatively if it is associated with a decaying shear fluctuation of *B*-mode type), then it may interact with its own subharmonics to form a shock wave. We know that simple waves will develop a pressure discontinuity in this way after *ca.* $(\delta\rho/\rho)^{-1}$ oscillations. After shock formation we expect such a process to damp out the motion producing in a time $t_d \sim \lambda(\delta\rho/\rho)^{-1}$, where $\delta\rho/\rho$ is of order Δv_0, the velocity jump or shock strength. This indicates that large amplitude, short wavelength fluctuations are most efficiently damped by this means; recall that in §3.2.2.1 we saw how *linear* fluctuations were diffusively damped in an amplitude-independent time $t_{lin} \propto \lambda^2$. However, for velocity 'overshoot' and shock formation to occur, sufficient time must be available; that is, the shocking time t_d must be less than the age of the Universe, t_{ex}. Thus, $\lambda(\delta\rho/\rho)^{-1} < t_{ex}$, which requires amplitudes such that $\delta\rho/\rho \lesssim \lambda/\lambda_J \sim (M/M_J)^{\frac{1}{3}}$ (since the Jeans length λ_J is related to t_{ex} in the radiation era by $\lambda_J \sim \frac{1}{\sqrt{3}} c\, t_{ex}$) (see Peebles 1970; Jones 1970). Now as Jones (1970) first pointed out, because the sound speed changes rapidly following $t_{eq} \sim 10^{10}\, \Omega^{-2}$ s, the criteria for shock formation will differ before and after this time. If $\delta\rho/\rho$ must exceed Δ^* for shocks to form then $\Delta^* \propto t^{-\frac{1}{2}}$ for $t < t_{eq}$ but Δ^* is constant after t_{eq} and so any nonlinearity destined to shock in this way must do so before t_{eq}.†

The residual spectrum is then expected to obey $(\delta\rho/\rho)_{rec} < (M/M_J)^{\frac{1}{3}}$ on scales up to $M_{J,\,rec} \sim 10^{15}\, \Omega^{-2}\, M_\odot$. This result was first obtained by Peebles (1970), who performed a second order perturbation analysis upon a Newtonian cosmology assuming a random distribution of wave phases to close the perturbation scheme. A more precise analysis in a relativistic cosmology has recently been performed by Liang (1977), who exploited the conformal invariance of the sound wave equations in an isotropically expanding radiation Universe to derive solutions from the exact special relativistic waves of Taub (1948). This approach solves the relativistic hydrodynamic equations exactly in a uniformly expanding background but ignores the reaction of the waves' own self-gravity changes upon the expansion dynamics. Such an approximation is valid for wavelengths and amplitudes such that $\lambda \ll \lambda_J(\delta\rho/\rho)^{-\frac{1}{2}}$, enabling large amplitudes to be discussed accurately on appropriate scales. Since the linear damping time is *ca.* λ^2 while the nonlinear damping time is *ca.* λ, nonlinear effects will predominate at shorter wavelengths, which enter the horizon earlier. Coordinating these ideas we see that the fate of a nonlinear adiabatic fluctuation avoiding black hole collapse will be to grow by gravitational instability before entering the horizon, then to oscillate *ca.* $(\delta\rho/\rho)^{-1}$ times before forming a shock after entering the Jeans length.‡ The wave energy will then be dissipated as heat, causing its amplitude to decay, $\delta\rho/\rho \propto t^{-\frac{1}{2}}$. Eventually the time scale for this decay will exceed the age of the Universe, t_{ex} (when its amplitude is, say, $(\delta\rho/\rho)_s$) and linear damping processes then intervene just before recombination to reduce the amplitude via an exponential damping factor to $(\delta\rho/\rho)_{rec} \sim (\delta\rho/\rho)_s \exp\{-(M_D/M)^{\frac{2}{3}}\}$. After recombination, all scales exceeding the new Jeans scale, *ca.* $10^5\, \Omega^{-\frac{1}{2}}\, M_\odot$ can recommence their amplitude growth by gravitational instability, $\delta\rho/\rho \propto t^{\frac{2}{3}}$.

Liang (1977) has further argued that these considerations indicate there to be a characteristic

† In reality, the details will be considerably more complex because of wave collisions and the positive feedback on the sound speed caused by the heat input and reionization from non-equilibrium processes in the shock fronts.

‡ These nonlinear adiabatic fluctuations will yield isothermal fluctuations over the scales where nonlinear effects are important. Shocks will lead to the clumping of baryons in the residual plasma. Also extremely nonlinear fluctuations will generate a collection of black holes that behave collectively as an isothermal mode that interacts only gravitationally with the expansion flow. Nonlinear adiabatic modes generate isothermal ones.

mass scale below which no fluctuation could survive shocking to attain the observed Gott–Rees amplitude at recombination of $\delta\rho/\rho \sim (10^5\,\Omega^{-\frac{1}{2}}\,M_\odot/M)^{\frac{1}{3}}$, no matter how large its initial amplitude. Taking into account linear damping, it is found that this critical mass scale singled out by the nonlinear processes is *ca.* $10^{11}\,\Omega^{-1}\,M_\odot$. A further twist has been added by further work of Baker & Liang (1978), who argue that all these features could be completely changed for ultra-relativistic ($\Gamma > 1$) waves. The dimensional analysis yielding the shock damping time *ca.* $\lambda\,(\delta\rho/\rho)^{-1}$ for Newtonian and mildly relativistic shocks breaks down completely in this limit and the damping disappears as v approaches the speed of light. The kinetic energy of the wave motion grows much faster than can be randomized as heat at the shock front, and forestalls the damping. They estimate $t_d \sim \lambda(1-(\Delta v_0)^2)^{-n}$, where n is positive and of order one; thus $t_d \rightarrow \infty$ as $v \sim \Delta v_0 \rightarrow c$ and the waves do not damp out.†

In a cosmological setting this might indicate that a break should appear in the irregularity spectrum if galaxies and clusters arose as a shock-damped residue of ultra-relativistic nature. In the subrelativistic region the power law residue should appear, but the primordial distribution would be mirrored in the ultra-relativistic region unless collisions and wave interactions provided efficient damping on short time scales.

It is conceivable that the inclusion of the self-gravity of these waves in a fully self-consistent fashion would alter these predictions somewhat. This feature has so far only been examined in Newtonian Universes by Asano (1974) and Liang (1978*b*), and both indicate that the stable irregularity spectra that result seem to have amplitudes increasing with mass, quite the opposite situation to that observed.

3.2.3. *Vortical fluctuations*

One of the problems with the gravitational instability schemes for galaxy formation that I have reviewed so far is that they generally involve predictions resting upon an unknown, and possibly unknowable, initial time at which fluctuations began to amplify as well as the initial amplitude on each scale. This unsavoury feature, coupled with the manifestly vortical appearance of many spiral galaxies, provoked many imaginative attempts to explain the development of galactic structures from vortical motions alone (Gamow 1952; Von Weiszacker 1957; Nariai 1956; Ozernoi & Chernin 1968). At first sight such a conception is extremely appealing. Small vortical velocity perturbations from the Friedmann background remain constant relative to a radiation background to conserve linear and angular momentum. Thus no *ad hoc* initial time seems necessary to explain the development of large amplitude perturbations at later epochs; only an amplitude $(v/c)_{eq}$ appropriate to the end of the radiation era at $t_{eq} \sim 10^{10}\,\Omega^2$ s seems required. Thereafter the velocities adiabatically decay ($v \propto R^{-1}$) and so the maximum scale of cosmic turbulence is attained at t_{eq}. In addition the sound speed during the radiation era is extremely large, *ca.* $(1/\sqrt{3})\,c$, and so long as $(v/c) \lesssim 0.4$, these relativistic velocities will be subsonic. After t_{eq} the sound speed falls off much more rapidly than the velocities decay and they inevitably become supersonic. From this chaotic state, density fluctuations are expected to arise with amplitude $\delta\rho/\rho \sim (v/c_s)^n$ (where $n \approx 2$, but extremely uncertain). Such seed fluctuations might then be amplified by gravitational instability in the usual manner although Peebles (1971) has argued that without some unusual

† This unusual and counterintuitive behaviour seems to be confirmed by recent numerical studies of Anile *et al.* (1979), but is probably only a consequence of the one-dimensional physics.

pressure support, for example magnetic fields, shocks would compress the material into superdense cores more akin to quasars or galactic nuclei than galaxies themselves.

Such a theory picks out a characteristic turbulent scale $\lambda \sim vt$ at cosmic time t. On smaller scales ($\lambda < vt$), the eddy turnover time will be shorter than the age of the Universe, and nonlinear hydrodynamical processes will lead to rapid energy transfer to increasingly shorter wavelengths where the vortical energy will be damped out by radiative viscosity. By way of contrast, those eddies of scale $\lambda > vt$ will, at time t, be 'frozen' into the expansion flow unable to evolve by hydrodynamic means because their turnover times exceed the age of the Universe (see Olson & Sachs 1973; Barrow 1977*a*; Drury & Stewart 1976). The largest scales becoming turbulent will thus be of order $\lambda_{eq} \sim (vt)_{eq}$ which neither freeze in nor have enough time to damp out by nonlinear energy transfer. This scale encompasses a mass $M_{eq} \sim \rho_{eq}(vt_{eq})^3 \sim 3\times 10^{15}(v/c)^3_{eq}\,\Omega^{-2}\,M_\odot$, which will be eroded by acoustic and viscous decay in any subsequent time interval before recombination when the eddy scale will be $M_{rec} \sim 1.5\times 10^{12}(v/c)^3_{eq}\,\Omega^{-\frac{17}{4}}\,M_\odot$. The one free parameter is v_{eq}, assumed independent of M for simplicity, which yields fluctuations of galactic scale for any Ω value determined observationally. Since highly developed turbulence is suspected to possess an 'inertial' range of scales over which the rate of energy transfer is constant, it was expected that an irregularity spectrum could be generated from this region of the turbulent motion with no memory of either generating or dissipative stresses on large or small scales. The simple Kolmogorov spectrum $v \propto \lambda^{\frac{1}{3}}$ yields $\delta\rho/\rho \propto M^{-\frac{2}{3}}$ over a range of scales exceeding the characteristic mass M_{rec}, according to Kursov & Ozernoi (1974*a–d*).

The apparent simplicity of this scheme is, however, rather illusory. In the radiation era, velocities do not remain constant with respect to a set of freely falling observers but rather relative to a set of non-geodesic coordinates also feeling the perturbation effects (Novikov & Zeldovich 1970; Barrow 1977*a*). Thus the metric distortions associated with eddies of galactic scale diverge as $t \to 0$ when they increasingly exceed the horizon size. In general, the small perturbation description fails and a strongly anisotropic space–time structure is required to accommodate the velocity fields at red shifts exceeding *ca.* $2\times 10^5\,\Omega(v/c)^2_{eq}$ (Barrow 1977*a*). When the background geometry becomes anisotropic in this way the velocity field is no longer constant and $v \to c$ as $t \to 0$. This indicates that the theory requires an initial time and v amplitude to specify the model uniquely just like the gravitational instability theories.†

Drury & Stewart (1976) have claimed that the level of turbulence required to create eddies of galactic scale is in fact too low to develop an inertial range of significant scale. Highly relativistic and supersonic motions would be required in the radiation era $(v/c)_{eq} \sim 1$. This needs anisotropic dynamics throughout the entire pre-recombination era with the accompanying problematic side effects.

Viscous dissipation will also take place over certain length scales and the eddy motion will be damped out in this way by the end of the radiation era over mass scales less than $M_T \sim 10^{11}\,\Omega^{-\frac{11}{4}}\,M_\odot$, the highly anisotropic momentum distribution within the spinning regions giving rise to a significant viscous stress (Harrison 1973). Jones (this symposium) has also given a very clear discussion of the damping due to radiative drag during the recombinative era itself. He claims that only vorticity on scales exceeding *ca.* $5\times 10^{11}\,\Omega^{-\frac{7}{2}}\,M_\odot$ will survive such

† It is possible for the velocities to decrease to zero as the singularity is approached ($v \propto R^2 \propto t^{\frac{2}{3}}$) if the equation of state in its vicinity is maximally stiff ($p = \rho$). Such an initial state could potentially explain the presence of any large scale vorticity in the Universe, although not of course the angular momentum (Barrow 1977*b*).

braking and this generates larger scale irregularities of too great a density to be associated with either galaxies or clusters.

Finally, the discordant predictions of this theory when confronted with light element abundances (Barrow 1977*a*) and the microwave background isotropy (Anile *et al.* 1976) now make this once attractive theory increasingly difficult to defend.

4. Fluctuation amplitudes

We shall conclude this survey by remarking upon one aspect of the galaxy formation problem that cannot look to post-recombination processes of an astrophysical nature for its resolution. This problem concerns the value of the fluctuation *amplitude* which needs to be relatively large (*ca.* 0.1%) on the relevant scales when they enter the horizon in order for the observed structure to subsequently develop. What explanations might be possible for this relatively high level of inhomogeneity?

It seems clear that any unique theoretical prediction relating the amplitude of a fluctuation to its wavelength must emerge from considerations of *nonlinear* physics. For example, in §2.1 I have outlined how naïve quantum gravitational effects predict metric and density fluctuations of unit amplitude on the Planck scale *ca.* 10^{-5} g. This is possible because quantum physics forges a unique link between the energy amplitude E and wavelength via the Uncertainty Principle limit $E \sim hc/\lambda$. At the classical level, more detailed investigations into the nonlinear hydrodynamics within an expanding radiation-dominated Universe may reveal intrinsically nonlinear phenomena like solitary waves or solitons whose amplitudes are determined by their size. Such structures might arise through the stabilization of nonlinear waves by gravitation (Asano 1974; Liang 1978*b*). In addition the recent discovery by Belinskii & Zhakarov (1979) of exact, inhomogeneous cosmological vacuum solutions to Einstein's equations of a soliton nature may enable one to calculate the expected amplitude of gravitational and density inhomogeneities of given wavelength if matter-filled analogues of these solutions can be found.

Finally, one interesting attempt to predict the fluctuation amplitude in the Universe is that of Liang mentioned in §2.2. Specifically, he considers a *stationary* gas of inhomogeneously distributed baryons in a smooth radiation bath. An expression for the gravitational free energy is obtained taking into account interaction between the non-uniformity distributed baryons. Remarkably a most probable inhomogeneous baryon distribution appears to exist close to the completely homogeneous equilibrium state and has the form $\delta\rho_b/\rho_b \sim (M/M_{Jb})^{-\frac{1}{3}} \sim (M/10^5\,\Omega^{-\frac{1}{2}}\,M_\odot)^{-\frac{1}{3}}$. The appearance of the matter Jeans mass M_{Jb} is natural since it is a time-invariant characteristic scale of a baryon–radiation mixture. It would be expected to emerge in a similar way from an analogous time-dependent analysis. However, it is likely that the spectral gradient $M^{-\frac{1}{3}}$ is entirely an artefact of the time-independence. In general, a system with time evolution would possess different amplitudes of inhomogeneity on different scales when they became causally coherent. The stricture of time-independence would enforce the situation where these amplitudes are all identical; that is, they must represent constant curvature fluctuations in the total density fluctuation and as described in §3.2.1.1 this ensures a *baryon* fluctuation spectrum $\delta\rho/\rho \propto M^{-\frac{1}{3}}$. There seems no reason to expect that this prediction would emerge from a *time-dependent* analysis of the gravithermodynamics.

I should like to thank Christ Church, Oxford, for their support, Professor M. J. Rees for his remarks which motivated the inclusion of §4, and Professor J. Silk for helpful comments.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

References (Barrow)

Anile, A., Danese, L., De Zotti, G. & Motta, S. 1976 *Astrophys. J.* **205**, L59.
Anile, A., Miller, J. & Motta, S. 1979 *Phys. Rev. Lett.* (submitted).
Asano, N. 1974 *Prog. theor. Phys. Suppl.* **55**, 52.
Barrow, J. D. 1977*a* *Mon. Not. R. astr. Soc.* **178**, 625.
Barrow, J. D. 1977*b* *Mon. Not. R. astr. Soc.* **179**, 47P.
Barrow, J. D. 1978 *Proc. 4th General Conference of European Physical Society*, York, ch. 4. Heinemann.
Barrow, J. D. & Carr, B. J. 1978 *Mon. Not. R. astr. Soc.* **182**, 537.
Barrow, J. D. & Matzner, R. A. 1977 *Mon. Not. R. astr. Soc.* **181**, 719.
Barrow, J. D. & Silk, J. 1979 *Gen. Relativ. Gravit.* **10**, 633.
Belinskii, V. A. & Zakharov, V. E. 1979 *Soviet Phys. JETP.* (In the press.)
Binney, J. J. 1977 *Astrophys. J.* **215**, 483.
Bonnor, W. B. 1957 *Mon. Not. R. astr. Soc.* **117**, 104.
Bonnor, W. B. 1956 *Mon. Not. R. astr. Soc.* **116**, 351.
Carr, B. J. 1975 *Astrophys. J.* **201**, 1.
Carr, B. J. 1977 *Astron. Astrophys.* **56**, 377.
Carr, B. J. 1978 *Mon. Not. R. astr. Soc.* **189**, 123.
Carr, B. J. 1977 *Astron. Astrophys.* **60**, 13.
Chernin, A. D. 1973 *Soviet Astr.* **18**, 441.
Chibisov, G. V. 1972 *Soviet Astr.* **16**, 235.
Doroshkevich, A. G., Sunyaev, R. A. & Zeldovich, Y. B. 1974 In *Confrontation of cosmological theory and observational data* (ed. M. S. Longair), pp. 213–225. Dordrecht: D. Reidel.
D'Eath, P. D. 1976 *Ann. Phys.* **98**, 237.
Drury, L. O'C. & Stewart, J. M. 1976 *Mon. Not. R. astr. Soc.* **177**, 377.
Eardley, D. M. 1974 *Phys. Rev. Lett.* **33**, 442.
Epstein, R. I. & Petrosian, V. 1975 *Astrophys. J.* **197**, 281.
Gamow, G. 1952 *Phys. Rev.* **86**, 251.
Glansdorff, P. & Prigogine, I. 1971 *Thermodynamic theory of structure, stability and fluctuations*. New York: Wiley.
Genkin, I. I. 1970 *Soviet Astr.* **14**, 5.
Gott, J. R. 1977 *A. Rev. Astron. Astrophys.* **15**, 235.
Gott, J. R. & Rees, M. J. 1975 *Astron. Astrophys.* **45**, 465.
Groth, E. J. & Peebles, P. J. E. 1977 *Astrophys. J.* **217**, 385.
Guyot, M. & Zeldovich, Y. B. 1970 *Astron. Astrophys.* **9**, 227.
Harrison, E. R. 1970 *Phys. Rev.* **D1**, 2726.
Harrison, E. R. 1973 *Mon. Not. R. astr. Soc.* **165**, 185.
Hawking, S. W. 1971 *Mon. Not. R. astr. Soc.* **152**, 75.
Hogan, C. 1978 *Mon. Not. R. astr. Soc.* **185**, 889.
Jones, B. J. T. 1970 D.Phil. thesis, University of Cambridge.
Jones, B. J. T. 1976 *Rev. mod. Phys.* **48**, 107.
Jones, B. J. T. 1977 *Mon. Not. R. astr. Soc.* **180**, 151.
Kursov, A. A. & Ozernoi, L. M. 1974*a* *Soviet Astr.* **18**, 157.
Kursov, A. A. & Ozernoi, L. M. 1974*b* *Soviet Astr.* **18**, 300.
Kursov, A. A. & Ozernoi, L. M. 1974*c* *Soviet Astr.* **18**, 700.
Kursov, A. A. & Ozernoi, L. M. 1974*d* *Soviet Astr.* **19**, 569.
Lake, K. & Roeder, R. C. 1976 *Nuovo Cim. Lett.* (2) **16**, 17.
Liang, E. P. 1975 *Phys. Lett.* A **51**, 141.
Liang, E. P. 1977 *Astrophys. J.* **211**, 361.
Liang, E. P. 1978*a* Preprint.
Liang, E. P. 1978*b* *Astrophys. J.* **230**, 325.
Liang, E. P. & Baker, K. 1977 *Phys. Rev. Lett.* **39**, 191.
Lifshitz, E. M. 1946 *Soviet Phys. JETP* **16**, 587.
Lukash, V. N., Novikov, I. D. & Starobinskii, A. A. 1975 *Soviet Phys. JETP* **42**, 757.
Meszaros, P. 1974 *Astron. Astrophys.* **37**, 225.
Meszaros, P. 1975 *Astron. Astrophys.* **38**, 5.
Mitchell, R. J., Culhane, J. L., Davison, F. J. N. & Ives, J. C. 1976 *Mon. Not. R. astr. Soc.* **176**, 29.
Nariai, H. 1956 *Sci. Rep. Tôhoku Univ.* **39**, 213.
Novikov, I. D. & Zeldovich, Y. B. 1970 *Astrophysics* **6**, 379.
Olson, D. W. & Sachs, R. K. 1973 *Astrophys. J.* **185**, 91.
Ozernoi, L. M. & Chernin, A. D. 1968 *Soviet Astr.* **11**, 907.
Parker, L. 1968 *Phys. Rev. Lett.* **21**, 562.
Peebles, P. J. E. 1965 *Astrophys. J.* **142**, 1317.

Peebles, P. J. E. 1968 *Nature, Lond.* **220**, 237.
Peebles, P. J. E. 1971 *Astrophys. Space Sci.* **11**, 443.
Peebles, P. J. E. 1970 *Phys. Rev.* D **1**, 397.
Peebles, P. J. E. & Dicke, R. H. 1968 *Astrophys. J.* **154**, 891.
Peebles, P. J. E. & Groth, E. J. 1975 *Astrophys. J.* **196**, 1.
Peebles, P. J. E. & Yu, J. T. 1970 *Astrophys. J.* **162**, 815.
Penrose, R. 1979 In *General Relativity: an Einstein centenary survey* (ed. S. W. Hawking & W. Israel), pp. 581–638. Cambridge University Press.
Press, W. H. & Schecter, P. 1974 *Astrophys. J.* **187**, 425.
Rees, M. J. 1971 In *General Relativity and gravitation* (ed. R. K. Sachs), pp. 315–346. New York: Academic Press.
Rees, M. J. 1977 In *The evolution of galaxies and stellar populations* (ed. B. M. Tinsley & R. B. Larson), pp. 339–368. New Haven: Yale University Observatory.
Rees, M. J. & Ostriker, J. P. 1977 *Mon. Not. R. astr. Soc.* **179**, 541.
Ryan, M. P. 1972 *Astrophys. J.* **177**, L79.
Saslaw, W. 1968 *Mon. Not. R. astr. Soc.* **141**, 1.
Schrödinger, E. 1939 *Physics* **6**, 899.
Sexl, R. & Urbantke, H. K. 1967 *Acta phys. austriaca* **26**, 339.
Silk, J. 1968 *Astrophys. J.* **151**, 459.
Silk, J. 1977 *Astrophys. J.* **211**, 638.
Sunayev, R. A. & Zeldovich, Y. B. 1970 *Astrophys. Space Sci.* **7**, 1.
Taub, A. H. 1948 *Phys. Rev.* **74**, 328.
Terletskii, Y. P. 1955 *Soviet Phys. JETP* **2**, 262.
Tomita, K. 1972 *Prog. theor. Phys.* **48**, 1503.
von Weizsacker, C. F. 1951 *Astrophys. J.* **114**, 165.
Wheeler, J. A. 1957 *Ann. Phys.* **2**, 604.
White, S. D. M. & Rees, M. J. 1978 *Mon. Not. R. astr. Soc.* **183**, 341.
Zeldovich, Y. B. 1970 *Soviet Phys. JETP Lett.* **12**, 307.
Zeldovich, Y. B. 1973 *Soviet Phys. JETP* **37**, 33.
Zeldovich, Y. B. & Novikov, I. D. 1966 *Soviet Astr.* **10**, 602.
Zeldovich, Y. B. & Novikov, I. D. 1970 *Soviet Astr.* **13**, 754.
Zeldovich, Y. B. & Starobinskii, A. A. 1978 *Zh. éksp. teor. Fiz. Pis'ma Redakt.* **26**, 373.

Discussion

C. J. Hogan (*Institute of Astronomy, Cambridge, U.K.*). If primordial fluctuations had a power-law spectrum down to *ca.* $10^6\ M_\odot$, matter in the early stages of the hot big bang ($10^3 < z < 10^{12}$) would have been extremely inhomogeneous. Its dynamics are dominated by the effect of the radiation background; heat conduction by photons (at $z \sim 10^3$) or neutrinos (at $z \sim 10^{12}$) tends to inflate clouds of matter with a baryon mass less than

$$M_{\mathrm{Jb}} \approx 2\pi^{\frac{3}{2}}(\delta/S)^{-\frac{1}{2}}\, M_\odot,$$

where δ is the density contrast of the cloud and S is the specific entropy of the Universe. Clouds more massive than M_{Jb} tend to collapse by a similar diffusion process.

Two interesting results emerge from a detailed analysis (Hogan 1978) of the behaviour of matter during this epoch. First, since neutrino diffusion operates before nucleosynthesis, matter that does not collapse into black holes must have $\delta \lesssim 2\times 10^5$ when nucleosynthesis occurs. This means that no gas can emerge from the big bang with a heavy-element abundance greater than about 3×10^{-5}, by mass, no matter what the initial conditions are. The observed low abundances in old stars therefore come as no surprise even in exotic big bangs. Secondly, if any compact objects (other than black holes) form in the early Universe, they must be stars. A wide range of initial conditions at $z \approx 10^{12}$ leads to stars emerging at recombination already burning on the main sequence. A 'primordial' population of stars might provide metals for visible stars and make a contribution towards the 'missing mass'.