The origin of galaxies: A review of recent theoretical developments and their confrontation with observation

Bernard J. T. Jones

Department of Astronomy, University of California, Berkeley, California 94720 and Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA, England

The subject of Galaxy Formation has advanced considerably during the past decade. On the theoretical side two theories in particular have been developed to the point where confrontation with observation will be possible; these are the "Gravitational Instability Picture" and the "Cosmic Turbulence Theory." These theories are discussed at some length here, with particular attention to the question of the origin of cosmic angular momentum and the nature of the initial conditions. There is now a considerable body of data on galaxies; the problem is in deciding which kind of observation is most relevant to understanding the origin of galaxies. Throughout the review an attempt is made both to put the present research in its historical perspective and to stress the possibilities for future advances towards the goal of understanding the origin of cosmic structure.

CONTENTS

Preface 107
I. Introduction 107
II. Historical Perspective 108
III. The Gravitational Instability Theory 111
A. Basic ideas 111
1. The horizon scale 111
2. The Jeans length scale 112
3. The damping scale 112
4. Time evolution of adiabatic density perturbations 113
B. The damping of adiabatic density perturbations 113
C. Adiabatic density perturbations: Nonlinear effects 114
D. Isothermal perturbations 115
IV. Cosmic Angular Momentum 115
V. The Formation of Clusters of Galaxies 118
VI. The Cosmic Turbulence Theory 119
A. The Theory of Ozernoi *et al.* 119
1. The scale of cosmic turbulence 119
2. The radiation-dominated era 120
3. The matter-dominated era 120
4. Characteristic mass scales 121
5. After recombination: the condensation of galaxies 121
6. Galactic angular momentum 121
B. The Pre-recombination Era 122
1. The Olsen–Sachs analysis 122
2. Does turbulence decay? 122
3. Analytic models of turbulence 122
C. The "Supersonic" regime 123
D. The Recombination era 123
E. The generation of density fluctuations by turbulence 124
VII. The Earliest Epochs 124
A. The initial conditions 125
B. The general homogeneity and isotropy of the universe 126
C. Chaotic cosmologies: General speculations 126
VIII. Confrontation with Observation 127
A. The masses of the galaxies 128
1. The internal dynamics of the galaxy 128
2. Binary galaxies 128
3. Dynamical studies of small groups of galaxies 129
4. Large aggregates of galaxies 129
5. The luminosity function of galaxies 129
B. The angular momenta and binding energies of galaxies 130
C. The morphology of galaxies 132
D. The clustering of galaxies 133
E. Young galaxies 135
IX. Other Theories of Galaxy Formation 136
X. Concluding Remarks 138
Acknowledgments 139
Appendix A. The "Big Bang" Theory 139
Appendix B. Hydrodynamics 142
References 145

PREFACE

It is hoped that this review will give the general reader an overview of current thinking regarding the problem of the origin of galaxies, while at the same time providing the specialist reader with a fairly unified and critical picture. For the general reader who might not be familiar with the details of the cosmological models underlying the present discussion, Appendix A has been included to give an over-all picture with definitions of common technical terms ("jargon") and units that are used throughout the review. Appendix B contains a summary of concepts from hydrodynamics that are needed to follow some of the more detailed discussions that appear in Secs. IV and VI. The review covers a selection from the literature on galaxy formation published prior to 1975. Articles appearing since early 1975 have been referred to only where they might add something really new to the pre-1975 state of affairs, or where clarification of some subtle point might result.

I. INTRODUCTION

In recent years considerable advances have been made towards an understanding of the origin and evolution of galaxies in the universe. The establishment in 1965 (Penzias and Wilson, 1965; Dicke *et al.*, 1965) of a definite over-all picture of the universe (the "hot big bang theory")[1] has been a major factor, but not the only one. Improvements in the techniques of optical observational astronomy have provided a greater understanding of the

[1]Although most cosmologists would feel that, at the present time, the evidence for the "hot big bang" theory of the universe is very compelling, there are still questions raised concerning the validity of the now orthodox interpretation of the cosmic microwave background radiation field (see, for example, Burbidge, 1971). Recently, the strength of such criticism has been somewhat eroded by observations of the radiation spectrum at submillimeter wavelengths (Robson *et al.*, 1974; Woody

material content of the universe, so it is no longer quite so easy to formulate cosmogonic theories without due regard to the observational data. The aims of this review are therefore twofold: firstly to appraise the current theoretical situation regarding the principle schools of thought on the subject of galaxy formation, and secondly to elucidate the role played by the observational data in evaluating theories and in providing further stimulus for investigation. The review is to be regarded as complementary to Field's review article for *Stars and Stellar Systems,* Vol. IX (Field, 1967), which discusses many topics at length that have been omitted here.

There are a number of other reviews of Galaxy Formation theories. The reviews of Layzer (1964) and Zel'dovich (1965) predate the "renaissance" in cosmology,[2] but have not been rendered "incorrect" in any way. Indeed, Layzer's discussion of von Weizsacker's turbulence cosmogony and Zel'dovich's discussion of the spectrum of random density fluctuations are still unequalled in their clarity. Subsequent to 1965 a number of review articles on specialized aspects of galaxy formation theory appeared by Zel'dovich and Novikov (1967), Harrison (1967), and the Japanese groups.[3] More recently, Rees (1971) and Rees and Silk (1970) have provided particularly readable articles reviewing the gravitational instability theory. The textbooks of Peebles (*Physical Cosmology*) (Peebles, 1971a) and Weinberg (*Gravitation and Cosmology*) (Weinberg, 1972) contain chapters on galaxy formation, but again there is little or no discussion of the cosmic turbulence theory. The recent *I. A. U. Symposium No. 58* on the Formation and Evolution of Galaxies contains a comprehensive selection of current papers and discussions at a greater level of detail, together with a general review by Peebles.

The choice of the material for this review is based on developments since Field's review article was written. The principle "new" theoretical ideas are perhaps the revival of the cosmic turbulence theory by Ozernoi and his co-workers, the angular momentum problem in the gravitational instability theory, and the "primeval globular clusters" hypothesis of Dicke and Peebles. On the observational side there are a number of developments which have led to a deeper understanding of the nature of the galaxies, though in some cases the observations have served to confuse what was previously thought to be a clear situation. Rather than reemphasize or enlarge upon points made by other reviewers of the field, I have chosen to discuss these more modern and often controversial observations partly in the hope of stimulating further interest on the part of observers.

There remains a serious gap in the scheme of things described here. It is not clear yet how the theories of galaxy formation discussed here link up with the structures seen through telescopes; a considerable amount of evolution must take place for the insignificant protostructures presumed to survive the primordial fireball to evolve into the symmetric structures that are observed. While considerable understanding of the processes involved can be obtained from the simple evolutionary models of Larson and Tinsley, they are a long way yet from presenting a complete picture. It is perhaps here that most of the theoretical research remains to be done. It is perhaps useful to keep in mind the goal expressed by Lemaître in his book *The Primeval Atom: The purpose of any Cosmogonic Theory is to seek out ideally simple conditions which could have initiated the world and from which, by the play of recognized forces, that world, in all its complexity, may have resulted.* (Lemaître, 1950).

II. HISTORICAL PERSPECTIVE

The problem of the origin of structure in the universe—the science of Cosmogony—has been a subject of debate among scientists, philosophers, and theologians for at least two and a half thousand years. While the physicists of today can hardly be influenced by the works of the most ancient philosophers, there can be little doubt that there is a chain of interactions among successive generations of cosmogonists which influence the general train of thought. Kepler, Copernicus, and Galileo were undoubtedly influenced by the philosophies of Aristotle and Epicurus.[4] Descartes and Newton (1644) were in turn motivated by the ideas expressed earlier by Kepler and Copernicus, and themselves influenced the later work of Laplace (1884), whose "nebular hypothesis" had repercussions extending right into this century.[5] There are certain very general trends of thought that are common to cosmogonic theories throughout the past. The Epicureans, who maintained that the primordial state was one of chaos out of which order arose by some means, find their modern parallel in the protagonists of the cosmic turbulence theory for the origin of

et al., 1975). These observations lend support to the contention that the radiation spectrum is Planckian. The conventional interpretation of the cosmological redshift of spectral lines of galaxies as a Doppler shift has also been questioned (see, for example, Field *et al*., 1975). Our presently accepted view of the universe would need serious revision if indeed there were such an ambiguity in the conventional redshift interpretation. An example of a non-Friedmann cosmological model which provides an alternative explanation for the origin of the cosmic microwave background radiation, and which also includes the possibility of non-Doppler redshifts, has been discussed by Hoyle and Narlikar (1974) and Hoyle (1975).

[2]A term used by D. W. Sciama in his lectures on cosmology (Sciama, 1971).

[3]"Evolution of the Universe and Formation of Galaxies," a collection of review articles in Prog. Theor. Phys. Suppl., 1971, No. 49 (Aizu *et al.*, 1971; Sato *et al.*, 1971; Nariai and Tomita, 1971).

[4]"The Milky Way" by Jaki (1973) is a very readable account of theories of the nature of the Milky Way system from early Greek times to the present. The section there on the ideas of the early Greeks outlines mainly the Aristotelian point of view and discusses the "atomistic" point of view of Demokritos briefly. (Unfortunately, only a few fragments of Demokritos' writings have survived and his ideas must be learned from the commentary of others.) Demokritos' views formed a foundation for the philosophy of Epicurus, though again, only fragments of his poem *De Natura* describing his atomistic viewpoint survive. The later works of Lucretius (*De Rerum Natura*) and Ovid (*Metamorphoses*) are important sources for the ideas of Demokritos and Epicurus.

[5]See, for example, Poincaré (1894), Aitken (1906), Jeans (1918; 1928), and Lemaître (1950), especially Chap. IV.

galaxies, in which theory the universe in its initial state is assumed to be extremely chaotic. On the other hand the theory of Aristotle held that the initial state was perfectly ordered and simple, and finds a parallel in modern cosmogony in the gravitational instability picture, wherein the degree of irregularity of the universe increases with time. Aristotle had argued against the Epicureans on the grounds that order could not be created out of disorder, and in parallel with this, Newton (1687) argued against the vortex cosmogony of Descartes that it would fail to reproduce the cosmic order displayed by Kepler's laws discovered eighty years earlier (Kepler, 1609) (Descartes having been unaware of Kepler's laws). Despite several attempts, Newton failed to prove this point; the equations of hydrodynamics were not written down until 1755 (by Euler), and it took another two centuries for the theory of turbulence to emerge at a level where such a discussion was possible.

The discovery of the spiral structure of galaxies by Lord Rosse (1850a, 1850b) over the period 1842–1850 might have inspired further speculation as to the origin of the solar system and the galaxy, for here was direct observational evidence for the kind of scheme envisaged fifty years earlier by Laplace. However, the discovery stimulated few papers on the subject, that of Alexander in 1852 being particularly noteworthy. It is true that there had been earlier speculation about the *spiral* structure of the Galaxy. The somewhat remarkable works of Swedenborg (1734), Kant (1755), Wright (1750), and Lambert (1761) are often cited as predictions of this; however, they can hardly be attributed the same importance as the later Laplace nebular hypothesis which dominated cosmogonic thought for almost 150 years after its formulation. Alexander clearly saw Rosse's discovery as evidence in favor of Laplace's hypothesis; his underlying scheme is the collapse of an initially oblate rotating gas cloud. An interesting feature of his work is his idea that the discriminating feature between "spiral nebulae" and "elliptical nebulae" was the amount of angular momentum possessed by the initial system. It is perhaps surprising that Alexander's work never stimulated any further discussions.

The earliest photography of galaxies (Roberts, 1889; Keeler, 1900) produced a greater reaction and around the turn of the century there appeared a large number of papers on cosmogony by Darwin (1908), Faye (1885), See (1910), Wilczyinski (1908), Jeans (1902), Chamberlain (1900, 1901), Moulton (1900, 1905), Sutherland (1911), and others. Laplace's hypothesis was in all cases a basic theoretical source of motivation. Faye introduced some kind of turbulence into the primeval nebula, whereas Darwin introduced the idea that the original nebula was not gaseous but consisted of a swarm of meteorites. Darwin's hypothesis was taken up by Wilczyinski, Moulton, and Sutherland. See, and later Chamberlain, introduced the idea that the nebulae were the result of the collision between two stars or gas clouds. Jeans' work, which has dominated modern cosmogonic theory, was motivated by the problem of the fragmentation of the Laplace nebula into planets, though Jeans himself later applied his theory of gravitational stability to the problem of the origin of the nebulae themselves. The works of Poincaré (1894) and Jeans (1918, 1928) provide excellent critical reviews of the state of cosmogony up to this time.[6]

The establishment of an extragalactic distance scale[7] and the accompanying realization that the nebulae had dimensions to be measured in thousands of parsecs caused a serious revision of ideas: the spiral nebulae were no longer to be considered merely as a part in the train of evolution of a solar system, but as entities consisting of individual stars and possibly solar systems, not unlike our own Galaxy. The problem of the origin of spiral galaxies was therefore put on a different footing from the problem of the origin of the solar system, and Jeans (1918, 1928) suggested that the mechanism by which they originated was the gravitational instability of a uniform universe. The impact of Hubble's work in this respect can be judged by comparing Jeans' two works on cosmogony: *Cosmogony and Stellar Dynamics* published in 1918 and *Astronomy and Cosmogony* published in 1928.[8] Jeans did comment on the mystery associated with the spiral structure, but it is a little surprising that he did not wonder about the origin of the rotation of the galaxies, especially since he actually discusses tidal angular momentum transfer in the same volumes.

The expansion of the universe was not a factor taken into account by Jeans and Poincaré: it was Lemaître who took the first steps towards reconciling the observed cosmic expansion with Einstein's theory of gravity, and then considered the question of the origin of galaxies as a problem of the gravitational instability of a uniform expanding system. It is certainly indicative of the impact of the Laplace hypothesis that right until 1945 Lemaître[9] thought of the universe as being one huge primeval nebula whose inherent instability led to both the condensation of galaxies and the over-all cosmic expansion. The problem was taken up independently in 1939 by Gamow and Teller (1939), who also applied the "Jeans instability" to the expanding universe. Their paper is a very lucid description of the gravitational instability theory and, qualitatively, little has changed since that paper was written 35 years ago.

Photographs of spiral galaxies convey the immediate impression that galaxies could well be the remnant "whirlpools" of some primordial turbulence (indeed, the galaxy M51 is commonly known as the "Whirlpool Nebula"). Some twenty years ago, von Weiszacker (1951) and Gamow (1952) put forward just such a hypothesis, and these early ideas have subsequently been exploited within the framework of the "hot big bang" cosmology by Ozernoi and his co-workers (Ozernoi and Chernin, 1968,

[6]There is also an interesting review by Aitken (1906).

[7]Before Hubble's work (Hubble 1925 a, b, c; 1926; 1929) the question of the distances to the nebulae was a major point of controversy. The principle protagonist of the view that nebulae were external galaxies rather like our own was Curtis (see, for example, Curtis, 1919). Curtis was strongly opposed by Shapley (see, for example, Shapley, 1919). An interesting account of this conflict is given by Sandage (1961).

[8]The relevant chapters to compare are Chap. IX of *Problems of Cosmogony and Stellar Dynamics* (Jeans, 1918) and Chap. XIII of *Astronomy and Cosmogony* (Jeans, 1928).

[9]See text of the lecture given by Lemaître in 1945 in Chap. IV of his book *The Primeval Atom* (Lemaître, 1950).

1969; Ozernoi and Chibisov, 1971a, 1971b, 1972) in the Soviet Union and by Nariai and his associates (Nariai, 1956a, 1956b; Tomita *et al.*, 1970) in Japan. The development during the 1930s and 1940s of the theory of hydrodynamic turbulence[10] and particularly von Weizsacker's involvement with the theory of turbulence set the scene for the application of the new theory to problems of cosmogony: the origin of the solar system (von Weizsacker, 1944), the internal dynamics of galaxies (Heisenberg and von Weizsacker, 1948), and the origin of galaxies (von Weizsacker, 1948). Indeed it has been remarked (Chandrasekhar, 1949) that it was von Weizsacker's emphasis on the role of turbulence in cosmogony that led Heisenberg to examine the physical basis for the nature of turbulence.

The climax of the gradual development of the understanding of turbulence and its possible application to astrophysics was surely reached in 1949. In that year, Chandrasekhar (Chandrasekhar, 1949) delivered his important Henry Norris Russell lecture on the future role of turbulence in astrophysical theories, and a joint IAU–IUTAM conference was held in Paris on the subject of "Cosmical Aerodynamics: the Motion of Gaseous Masses of Cosmical Dimension." It must have been about that time also that Gamow changed his mind about the way in which galaxies originated. In 1948 (Gamow, 1949) he had reinforced his earlier (Gamow and Teller, 1939) suggestions that galaxies formed as a result of gravitational instability in the cosmic medium, while four years later in 1952 (Gamow, 1952) he published his paper supporting von Weizsacker's view (von Weizsacker, 1951) that galaxies originated from cosmic turbulence. The reason for the change of opinion on Gamow's part is clearly stated in the paper: he felt that the Jeans instability does not cause small density irregularities to grow fast enough to explain the origin of galaxies from merely statistical fluctuations in the cosmic matter density. This result had been derived in 1946 by Lifschitz (1946), and was confirmed independently by Gamow, Ulam, and Metropolis (1948).

At this stage one is confronted with one of the historical paradoxes of cosmogony. Apart from the papers of Nariai published in a not widely read journal in 1956 (Nariai, 1956a, b), there is a complete lack of papers on the subject of cosmic turbulence and the origin of galaxies from 1952 until the time of the revival of the concept by Ozernoi and his co-workers in the late 1960s. In 1952, one might have felt that the question was close to being resolved in favor of the cosmic turbulence theory. Yet, for no obvious reason, the subject was dropped, and even Gamow's subsequent publications (Gamow, 1956; Alpher, Gamow, and Herman, 1967) repeatedly emphasized the importance of the galaxy formation problem and the possibility of resolving the problem by gravitational instability! No refutation of the turbulence theory had been published, though doubts had been expressed by Gamow (Gamow, 1951, 1952) and later by Bonnor (1956) about the need to invoke the existence of turbulence *ab initio* in order to explain galaxies: One had only replaced the question of the origin of galaxies with the question of the origin of turbulence. Bonnor (1956) even expressed some doubt as to whether primordial turbulence could even survive dissipation during the early phases of expansion.

When, in 1964, Layzer (1964) reviewed the status of the subject of galaxy formation, he commented both on the turbulence and the gravitational instability theories, as well as other theories, and reaffirmed the conclusions reached by Bonnor that gravitational instability was too slow to form galaxies out of statistical fluctuations, and that von Weizsacker's theory was faced with the problems both of the origin and the decay of the turbulence.

This somewhat stagnant situation changed rapidly with the discovery in 1965 of the 3 °K cosmic microwave background radiation field. The existence of such a radiation field at approximately this temperature had been postulated by Gamow (1953) as long ago as 1953, so this discovery provided powerful evidence in favor of the hypothesis that the universe expanded from a *hot*, dense, singular state, a finite time in our past. The establishment of a cosmological theory provided a concrete framework within which to discuss the physical processes taking place in the universe, and so stimulated a re-evaluation of problems such as cosmic nucleosynthesis (Peebles, 1966; Wagoner, Fowler, and Hoyle, 1967) and the question of the origin of galaxies. The presence of the radiation field means that gravitation is not the only agent affecting the evolution of irregularities in the universe: the radiation pressure and dissipative processes act to counter the enhancing effect of the gravitational field. Moreover, the accurately Planckian character of the cosmic radiation spectrum, together with its high degree of isotropy, provides a valuable tool for investigating the structure of the universe at very early epochs.

The first paper to appear on the subject of galaxy formation in the newly established cosmology was that of Peebles in 1965 (Peebles, 1965), where he showed that the effect of the Thomson drag force of the cosmic radiation field on the electrons would severely limit the growth of optically thin perturbations in density during the period when the matter and radiation fields were coupled. The behavior of adiabatic perturbations in an expanding universe had been considered as long ago as 1946 by Lifschitz (1946), who had recovered both the growing long-wave modes and the oscillating short-wave modes discussed by Jeans for perturbations in a stationary gas cloud. Since then there have been numerous reappraisals of this problem. Field's review article (Field, 1967) provides a good account of some of these. The existence of the cosmic radiation field added a new phenomenon, namely the dissipation of the acoustic modes by radiative diffusion; this was discussed independently by Michie (1967), Peebles (1967a), and Silk (1968). By 1968, the elementary aspects of the evolution of primordial density irregularities had been well understood. It was left only to refine the analysis, as for example in the discussion by Peebles and Yu (1970) of the damping of acoustic modes through the recombination era. The

[10]Among the classical papers on the subject of turbulence during this period, the reader may find it of interest to consult those by Taylor (1935; 1936; 1938), Karman and Howarth (1938), Kolmogorov (1941), Obhukoff (1941), Onsager (1945), Lin (1947), and Heisenberg (1948a, b).

situation at this stage has been well reviewed by Harrison (1967), Field (1967), and Rees (1971).

In contrast with the situation only 15 years earlier, in 1968 one might have felt that significant progress had been made towards understanding the origin of galaxies in terms of small initial irregularities in density. Two coincidences played a role in strengthening this impression. Firstly, the damping mass associated with adiabatic perturbations at the onset of recombination could perhaps be identified with large galaxies or with clusters of galaxies. Secondly, the Jeans mass immediately after recombination is of the same order as the mass of a globular cluster.

The problem of vorticity perturbations had been discussed in 1946 by Lifschitz, and more recently by Lifschitz and Khalatnikov (1963), Hawking (1966), and Sachs and Wolfe (1967). In these analyses, it was shown that isolated vorticity perturbations would evolve conserving their angular momentum. It was therefore unsatisfactory to invoke such perturbations as an explanation for the presently observed spin of galaxies; one was still left with the problem of why the initial angular momentum distribution had a particular form. It must be remembered that the problem of the initial conditions in the turbulence theory of von Weizsacker and Gamow had been a source of concern even for Gamow. The successful reintroduction of the cosmic turbulence theory by Ozernoi and his co-workers must therefore be considered a remarkable achievement. In one coup, the new version of the theory seemed to deal with the problems of the nature of the initial conditions, of supporting the turbulence against viscous decay, of the origin of galactic spin, and even provided an independent estimate for the epoch of galaxy formation—all in terms of one parameter characterizing the initial strength of the turbulence. Added to the fact that there has been considerable debate as to whether the gravitational instability picture can indeed account satisfactorily for the spin of galaxies, the cosmic turbulence theory seems once more to have gained a large number of supporters.

III. THE GRAVITATIONAL INSTABILITY THEORY

A. Basic ideas

As remarked in the previous section, the idea that galaxies have grown as a result of gravitational instability can be traced back to Jeans (1902). However, the problem of the stability, or otherwise, of an expanding universe is more complicated than the situation studied by Jeans, and there now exists a considerable body of literature on the subject.[11] The essential ideas underlying the gravitational instability picture have been reviewed in detail by Zel'dovich (1965), Field (1967), Peebles (1971), Rees (1971), and Weinberg (1972). There is also a readily readable review by Rees and Silk (1970) in *Scientific American*. A brief resumé providing a background for discussion of the more recent developments is sufficient here.

It is not possible to obtain exact solutions of the Einstein field equations (or even of Newton's equations) for a nonuniform cosmological model having no special symmetries. Thus most approaches to galaxy formation consider the evolution of small deviations from a prescribed cosmological model by using linear perturbation theory. In the absence of pressure gradients, a small-amplitude enhancement of density will expand more slowly than the universe as a whole because the gravitational field will be slightly stronger than average where the density is a little greater than average. Thus the density contrast of an inhomogeneity may be expected to grow with time (unless pressure forces should at some stage be able to balance the perturbed gravitational field) and one arrives at a picture wherein the universe becomes more lumpy as it evolves. The hope in the early days of galaxy formation theory had been that initial statistical fluctuations in the matter distribution might have time to grow by this process into galaxy-like lumps before the present epoch. The growth rate of the density contrast is too small, however, unless one is prepared to specify the statistical initial conditions at such early epochs where we cannot say with any confidence that the physics of the universe is understood. (See Sec. VII.A.)

The evolution of a density inhomogeneity as a function of time can be understood in simple terms by comparing its size with three important characteristic scales. These scales (which will be defined subsequently) are the distance to the horizon, the Jeans length scale, and the dissipation length scale. The cosmic fluid is a matter–radiation mixture and it is necessary to distinguish between two kinds of density perturbation. On the one hand, both the radiation and matter can be perturbed together in such a way that the ratio of the photon number density to the baryon density within the perturbation is the same as in the ambient medium. For such a perturbation, the relative fluctuation $\delta\rho/\rho$ in the matter density and the relative temperature fluctuation $\delta T_r/T_r$ are related by $\delta\rho/\rho = 3\delta T_r/T_r$ and are accordingly referred to as *adiabatic modes*. On the other hand, the matter distribution can be perturbed, leaving the photon density unaltered. The temperature within such a perturbation is the same as the temperature of the ambient medium, and these are referred to as *isothermal modes*. Since, in such a perturbation, the ratio of the photon number density to the baryon number density is different from that in the ambient medium, these are also referred to as *entropy fluctuations*. During the pre-recombination period, $\delta\rho/\rho$ for entropy fluctuations remains constant since any attempt to increase it is opposed by the Thomson drag force. The behavior of the adiabatic modes during this regime is more complicated.

1. The horizon scale

At a time t after the origin of the universe, observers A and B separated by a distance greater than ct (where

---

[11] There is now a considerable body of literature discussing the evolution of density perturbations in various cosmological models, using a wide variety of techniques and approximations. Perturbations to Friedmann–Lemaître universes have been discussed by Lifschitz (1946), Bonnor (1957), Lifschitz and Khalatnikov (1963), Peebles (1965), Hawking (1966), Harrison (1967), Sachs and Wolfe (1967), Michie (1967), Peebles (1967a, b), Nariai *et al.* (1967), Field and Shapley (1968), Silk (1968), Arons and Silk (1968), Layzer (1968), Tomita (1969a, b, c), Rees and Sciama (1969a, b), Zel'dovich (1970), and Sunyaev (1971). Perturbations to non-Friedmann universes have been discussed by Doroshkevich (1966), Perko *et al.* (1972), Hu and Regge (1972), and Chitre (1972).

c is the speed of light) will not have had time to see one another or communicate in any way. In more formal language, the past light cone of A at time t does not contain B's world-line. (See Appendix A.) A and B are said to be outside one another's *horizon*. When the spatial separation of A and B is ct, communication between A and B is established, and thereafter A and B are said to lie within one another's light cone. From the point of view of galaxy formation, the horizon is of importance, since at cosmic time t the edges of a spherical density inhomogeneity whose radius is greater than ct cannot causally affect one another, and the center of such a perturbation cannot "know" yet about the surrounding background universe. It will be readily appreciated that in General Relativity the use of perturbation theory (where the choice of background cosmology plays an important role) in discussing the evolution of such an inhomogeneity might lead to conceptual difficulties. This important problem will be discussed later.

The spatial distance to the horizon will be denoted by $\lambda_H = ct$, and this may be associated with a characteristic mass scale

$$M_H = \tfrac{1}{6}\pi\rho_m(ct)^3 .$$

This is the mass of baryons within the horizon at time t. (ρ_m here is the baryon mass density at epoch t.) It is a matter of convenience to talk in terms of the baryon mass rather than the total mass: the mass of baryons in a volume that partakes in the cosmic expansion remains constant. In a matter-dominated universe where $\rho_m \propto t^{-2}$, this mass scale increases as $M_H \propto t$. The number of baryons an observer can see within his horizon increases with time.

(The distance to the horizon is not to be confused with the *cosmic scale factor*, which is the factor by which the universe has linearly expanded between epochs t_1 and t_2. If a number of baryons occupies volume V_1 at t_1, and the same baryons occupy volume V_2 at t_2, the cosmic scale factors R_1 and R_2 at times t_1 and t_2 are related by $R_1/R_2 = (V_1/V_2)^{1/3}$. The dependence of $R(t)$ on cosmic time is found by solving the Einstein field equations for a homogeneous, isotropic fluid with a given equation of state. For an Einstein–de Sitter universe, $R \propto t^{2/3}$ for a $p=0$ equation of state, and $R \propto t^{1/2}$ for a $p = \frac{1}{3}\rho c^2$ equation of state.)

2. The Jeans length scale

The Jeans length scale λ_J is of fundamental importance in the gravitational instability picture. This scale is the minimum scale on which pressure gradients in a sphere of material can balance gravitational forces. For scales $\lambda > \lambda_J$, gravitation dominates the dynamical motions, whereas for $\lambda < \lambda_J$ pressure dominates and the inhomogeneity behaves like an acoustic wave. In a uniform sphere of fluid where in the absence of pressure forces the free-fall time scale for gravitational collapse would be t_f, the Jeans length is roughly $c_s t_f$, where c_s is the adiabatic sound speed in the fluid. Physically, a volume of fluid is stabilized by pressure against gravitational collapse if a sound wave can cross the volume on a time scale shorter than the collapse timescale. Modes $\lambda \ll \lambda_J$ behave as sound waves, while small-amplitude perturbations on scales $\lambda > \lambda_J$ will be gravitationally enhanced. Jeans (1902) showed that the density contrast of small-amplitude perturbations in a stationary self-gravitating gas cloud grows exponentially with time for $\lambda \gg \lambda_J$. In the expanding universe, however, the density contrast grows only as a fractional power of the cosmic time.

At time t in the expanding universe, the free-fall time scale t_f for a small-scale density perturbation is approximately t. The Jeans length is thus given by

$$\lambda_J \simeq c_s t ,$$

and since $c_s < c$, this is always less than the scale of the horizon. Prior to recombination, when the matter and radiation are closely coupled, the adiabatic sound speed c_s is given by

$$c_s^2 = \frac{c^2}{3}\,\frac{1}{(1+3\rho_m/4\rho_r)} , \tag{1}$$

where ρ_m and ρ_r are, respectively, the mass densities of the baryons and the radiation field ($\rho_r = aT_r^4/c^2$ at temperature T_r). During the radiation-dominated era ($\rho_r \gg \rho_m$), $c_s \simeq c/\sqrt{3}$ and the Jeans length is only slightly less than the horizon distance. During the matter-dominated era, but before recombination, $c_s \propto R^{-1/2}$, and so the Jeans mass $M_J \simeq \frac{1}{6}\pi(c_s t)^3\rho_m$ (defined for convenience as the mass of baryons in a sphere whose radius is the Jeans length) remains constant. After the matter has become neutral, Eq. (1) for the sound speed is no longer valid; the sound speed is determined by the gas pressure $P_m = nkT_m$. Thus during recombination the sound speed drops from a significant fraction of the speed of light to a few kilometers per second. The Jeans mass immediately before and after recombination is approximately

$$M_{J(\text{before})} \simeq 1.4\times10^{18}\,\frac{(\Omega h^2)^{-1/2}}{1+30(\Omega h^2)^{3/2}}M_\odot ,$$
$$M_{J(\text{after})} \simeq 7.3\times10^4(\Omega h^2)^{-1/2}M_\odot . \tag{2}$$

(These values are approximate, as are all mass scales which are quoted in this paper. The mass associated with a particular length scale varies as the cube of the length scale, and so a factor of two error in the length scale results in an error amounting to almost an order of magnitude in the corresponding mass).

3. The damping scale

Although adiabatic density perturbations on scales $\lambda \ll \lambda_J$ behave as acoustic modes, it is not possible to propagate disturbances of arbitrarily small wavelength through the universe. Limitations on the wavelength are imposed by the viscosity and thermal conductivity of the cosmic fluid, since both these processes can remove energy from sound waves of sufficiently high frequency. If the damping time scale is shorter than the cosmic expansion time scale, the wave will be efficiently damped before the universe has had time to expand by an appreciable factor. The shortest wave length whose damping time scale due to either of these dissipative processes is longer than the cosmic expansion time scale is referred to as the *damping scale* λ_D.

In the pre-recombination universe, both viscosity and heat conduction are governed by the Thomson scatter-

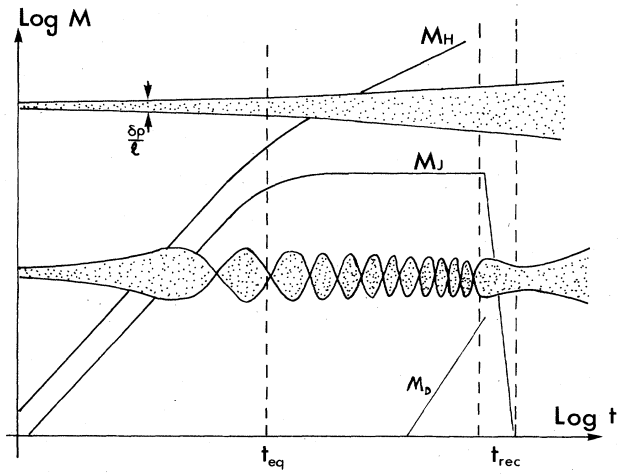

FIG. 1. The evolution of adiabatic density perturbations prior to recombination. M_H represents the mass of baryons within the horizon, M_J represents the mass of baryons within a sphere whose radius is the Jeans length. The figure displays the evolution of perturbations on two scales: the width of the shaded region represents the value of $\delta\rho/\rho$. The larger-scale perturbation has $M > M_J$ at all times, and $\delta\rho/\rho$ grows at all times. The smaller goes through a phase where $M < M_J$, during which it executes acoustic oscillations (with a slight decrease in amplitude), but after recombination $\delta\rho/\rho$ grows as a result of gravitational instability.

ing process. The photon mean free path to Thomson scattering is $l_\gamma \simeq (\sigma_T n_e)^{-1}$, where σ_T is the Thomson cross section and n_e is the free electron density. The time taken for a photon to diffuse a distance λ is therefore $t_{\rm diff} \sim \lambda^2/cl_\gamma$, and during the pre-recombination matter-dominated regime a sound wave will be damped when at time t its wavelength $\lambda < (l_\gamma ct)^{1/2} = (l_\gamma l_H)^{1/2}$. The damping scale in this regime is therefore

$$\lambda_D \simeq (l_\gamma l_H)^{1/2}, \quad \rho_m > \rho_r$$

and the corresponding mass (subject to the usual uncertainties) is

$$M_D \simeq 3\times10^{12}\,(\Omega h^2)^{-5/4}\,M_\odot \ .$$

Owing to the rapid increase in the photon mean free path during the recombination period, there will be some damping of motions on larger scales still; this will be the subject of discussion in Sec. III.B.

4. Time evolution of adiabatic density perturbations

Consider now a small-amplitude density perturbation occupying a volume in which the mass of baryons is M. The history of the perturbation can be qualitatively described by reference to Fig. 1, where the mass scales M_H, M_J, and M_D at various epochs are schematically depicted.

At sufficiently early times, $M>M_H$ and the perturbation extends beyond the horizon. Since $M>M_J$, one naturally expects that there will be an enhancement of the density contrast $\delta\rho/\rho$, relative to the background, as time proceeds. In Newtonian cosmology this expectation is confirmed unambiguously: in Newtonian cosmology a universal cosmic time is defined and the perturbation "knows" about the detailed structure of the universe on arbitrarily large scales (there is no horizon). In general relativistic cosmology, as mentioned earlier, there is an ambiguity arising from the fact that a perturbation with scale $M>M_H$ cannot "know" about the universe at large. It is not surprising, therefore, that there is a variety of opinions concerning the power of cosmic time t with which $\delta\rho/\rho$ grows under such circumstances. Various powers have been obtained by using different coordinate systems and taking different views as to which terms in the perturbed equations may justifiably be neglected. Technically, the difficulty stems from the lack of a gauge invariant description of a density perturbation [see, for example, Sachs and Wolfe (1971) or Sachs (1973) for a discussion of this]; it is not possible to distinguish a "true" density perturbation from something which looks like a density perturbation simply because of a curious choice of coordinate system. Some mathematical theorems which may shed light on this matter have been proved by Hawking (1966) and Shepley and Taub (1967). These theorems tell us that if the shear and vorticity (see Appendix B) of a perfect fluid motion are zero, then the metric of the space time is of the Robertson–Walker type. There is little to be gained, therefore, in considering "pure" density fluctuations in a Robertson–Walker universe: in the absence of any shear or vorticity, such an apparently inhomogeneous universe is still a Robertson–Walker universe and the shear-free density fluctuations must be an artifact of the choice of coordinate system.

Of course, though this is an interesting technicality, it need not inhibit further progress with the gravitational instability picture. It is not necessary to stipulate the amplitude of the irregularities at $t=0$; we can simply state the amplitude of the perturbation when it first comes within the horizon as an "initial" condition.

Once the perturbation comes within the horizon it will grow in amplitude for as long as $M>M_J$. Adiabatic density perturbations which are smaller than the Jeans mass immediately prior to recombination [see Eq. (2)] will start to oscillate when $M<M_J$ and will continue to do so until the universe recombines. If they survive until recombination, they will begin to grow in amplitude since they once again satisfy $M>M_J$. The isothermal perturbation modes are prevented from growing in amplitude prior to recombination by the Thomson drag force (Peebles, 1965). After recombination, they will grow in amplitude if $M>M_J$. Note that isothermal perturbations survive the pre-recombination fireball unscathed; the conditions emerging at recombination for such modes thus directly reflect the initial condition of the universe.

B. The damping of adiabatic density perturbations

Adiabatic density fluctuations on sufficiently small scales are damped by both the action of viscosity and radiative diffusion. The former process acts to smooth out velocity gradients, while the latter smooths out temperature gradients; however, because the relevant transport process is Thomson scattering in both cases, and because of the special relationship between the density and temperature in an adiabatic perturbation, such a distinction is not often made. [This distinction was made most clearly by Weinberg (1971).] The damping of adiabatic density perturbations was first calculated independently by Michie (1967), Peebles (1967), and Silk (1968), though only the latter work was published. At the onset of recombination, perturbations on scales smaller than

$$M_{D,\max} = 3\times10^{12}\,(\Omega h^2)^{-5/4}M_\odot \tag{3}$$

will have been severely attenuated. Thus the physical processes of the cosmic fireball serve to distinguish a scale which is about the size attributed to a large galaxy or cluster of galaxies. Michie (1967) followed the evolution of perturbations numerically and was able to continue the computations beyond the onset of recombination until the universe was almost neutral. His computations indicate further that a significant amount of damping takes place during the actual recombination process and the mass scale $M_{D,\max}$ given by Eq. (3) might be a severe underestimate of the size of the smallest surviving scale.

Since that early work there have been studies of the damping phenomenon by Bardeen (1968), Field (1971), Peebles and Yu (1970), Sato (1971), Weinberg (1971), Chibisov (1972), and by Silk (1974). The discussion of Peebles and Yu involves direct numerical integration of the Boltzmann equation for a coupled matter–radiation plasma right through the recombination era, whereas the discussions of Chibisov and Silk involve analytic approximations to the radiative transfer problem. There is only qualitative agreement between the results of these authors. A critical damping mass of

$$M_D \simeq 3\times10^{12}\,(\Omega h^2)^{-1}M_\odot \tag{4}$$

fits the results of Peebles and Yu. In comparison, Chibisov's analysis seems to overestimate M_D by a factor 30 for values of $\Omega h^2 \sim 1$, whereas Silk's analysis overestimates M_D only at the low-Ωh^2 range, and then by several orders of magnitude. In view of the importance of this scale there is clearly a need for further analytic discussion of the problem; a refined numerical analysis would not be without value either. The similarity of the values (3) and (4) for the damping mass indicates that much of the damping takes place before the onset of recombination; this is contrary to Silk's (1974) result, where most of the damping takes place towards the end of the recombination era.

Sunyaev and Zel'dovich (1970) remarked that perturbations which have an appropriate phase at the onset of recombination might retain a sufficiently large compressional component of velocity to partially restore the damped fluctuation after the completion of the recombination. This phenomenon can be seen in the numerical integrations of Michie and of Peebles and Yu (1970). (In the latter paper, all long-wavelength perturbations are assumed to come within the particle horizon with the same phase; this establishes a correlation between the phase and the perturbation mass at recombination so that at long wavelengths the power spectrum of irregularities peaks strongly at discrete mass values after recombination. This is an interesting way of creating a discrete hierarchy of mass scales from an initially featureless spectrum.) If the velocity field of the perturbation were unaffected by recombination, the amplitude would increase by a factor

$$\frac{\delta\rho/\rho\,|_{\text{before}}}{\delta\rho/\rho\,|_{\text{after}}} = \begin{cases} 14(M/10^{11}M_\odot)^{-1/3}, & \Omega h^2 = 1 \\ 380(M/10^{11}M_\odot)^{-1/3}, & \Omega h^2 = 0.02 \end{cases}$$

due to this effect. If there were any damping during the recombination, this would overestimate the amplification factor. As remarked earlier, according to Peebles and Yu, much of the damping takes place prior to the onset of recombination and hence some amplification should be observed due to this effect. A study of their results shows this. On the other hand, the analytic studies of Chibisov and Silk indicate that most damping takes place towards the end of recombination, so the amplification would not be significant. This would seem to be an important point when considering the role played by primeval adiabatic perturbations in the formation of cosmic structure.

It is tempting to compare the mass scale M_D with the scales attributed to galaxies and clusters of galaxies. At first sight it might seem that if primeval adiabatic density perturbations existed, they would give rise to objects on the scale of clusters of galaxies, rather than galaxies, since large galaxies are typically ascribed masses of only $10^{11}M_\odot$. If this argument were correct, it would be necessary to find a secondary cause for the origin of galaxies. However, there is now a small body of opinion that direct determinations of the masses of galaxies, by rotation curves, for example, tend to severely underestimate the actual mass of the galaxy (see Sec. VIII.A). It would seem premature at this stage to attempt to relate objects of mass $M_{D,\max}$ to observed systems. It is curious to note that the gravitational instability picture seems to overestimate the mass of a galaxy and to underestimate the angular momentum by roughly the same factor. If galaxies do indeed possess massive haloes, both these discrepancies would disappear together (Sec. IV).

C. Adiabatic density perturbations: Nonlinear effects

It was Peebles (1970) who first pointed out the possible importance of nonlinear effects on the propagation of sound waves in the early universe. A perturbation of mass $10^{12}M_\odot$ in an Einstein–de Sitter universe would just survive dissipation by the fireball and will have performed about ten acoustic oscillations since coming within the particle horizon. At the epoch t_{eq} when $\rho_r = \rho_m$, the linear theory for the propagation of a wave amplitude $\delta\rho/\rho$ is only strictly applicable for $(\delta\rho/\rho)^{-1}$ oscillations of the wave. The fact that the sound speed in the peak of the wave differs from the sound speed in the ambient medium means that the peak will catch up the trough, causing a tipping over or breaking of the wave. Thus, if the hypothetical perturbation of $10^{12}M_\odot$ had an amplitude somewhat in excess of 10% when it came within the horizon, it would break before the epoch of recombination. The breaking of such a wave propagating in the universe (viewed as the successive generation of higher-order harmonics of the original frequency) was first computed by Peebles (1970). The essential result of Peebles' calculation is that there is an upper limit on the amplitude of an adiabatic density perturbation of mass M_λ of

$$(\delta\rho/\rho)_\lambda \simeq (M_\lambda/M_J)^{1/3}\,,$$

where M_J is the Jeans mass at the epoch of recombination. A nonlinear upper limit on the amplitude of the smallest surviving scale M_D is therefore

$$(\delta\rho/\rho)_D < (M_D/M_J)^{1/3} \simeq 0.05\,(\Omega h^2)^{1/3} \ .$$

These scales would not collapse any earlier than a red shift of $z_{\rm coll} \leqslant (\delta\rho/\rho)_D^{-1} \simeq 20\,(\Omega h^2)^{-1/3}$. The problem has been investigated in more detail by Jones (1973a), who shows that any wave that will shock does so before the epoch t_{eq} when the energy densities of the matter and radiation are equal. When the shock forms, the wave amplitude decays until the decay time scale is comparable with the expansion time scale. This "freezing in" of the amplitude occurs at the epoch t_{eq} and the amplitude is then given by $\delta\rho/\rho = (M_\lambda/M_J)^{1/3}$, where M_λ is the mass associated with the perturbation and M_J is the Jeans mass at t_{eq}.

This kind of nonlinear process is of interest since it provides a natural limitation on the amplitude of density perturbations in the universe. Cosmic shock waves may also be an interesting source of entropy for the universe, but only if they arise out of large-amplitude acoustic modes: a small-amplitude sound wave will result only in a weak shock wave generating very little entropy.

D. Isothermal perturbations

Although isothermal perturbations do not evolve during the fireball phase of the universe (Peebles, 1965; 1969a), they are not dissipated either, and so the conditions at recombination precisely reflect the initial conditions of the universe. Just after recombination has been completed, perturbations on scales greater than the Jeans mass (which is then of the order of $10^5 M_\odot$) can grow as $\delta\rho/\rho \propto t^{2/3}$ until they reach $\delta\rho/\rho > 1$, and then collapse to form bound stellar systems. The Jeans mass just after recombination is of the order of the masses attributed to the globular clusters; this led Dicke and Peebles (1969) to call these objects "primeval globular clusters" (see also Ruzmaikina, 1972). Attempting to identify these $10^5 M_\odot$ isothermal fluctuations with protoglobular clusters opens up some exciting possibilities. If in the initial burst of star formation a significant amount of uncondensed gas in the protoglobular cluster was blown out of the cluster, the cluster would be left in an unbound state and the stars would evaporate to fill intergalactic space. This may be one of the few viable possibilities of explaining the existence of stars in the outer haloes of galaxies hypothesized by Peebles and Ostriker (1973), or of explaining the existence of stars in intergalactic space. One of the difficulties with this hypothesis is that the globular clusters associated with different galaxies seem to differ, whereas on the basis of this theory all globular clusters everywhere should be more or less the same. Thus one wonders why some galaxies like M87 have an enormous number of globular clusters, whereas others have comparatively few (Jaschek, 1957; Vorontsov-Vel'yaminov, 1966; de Vaucouleurs, 1970). One might also wonder why the globular clusters of M31 should be systematically underluminous compared with the globular clusters in our Galaxy (van den Bergh, 1969).

Some time ago, Doroshkevich, Zel'dovich, and Novikov (1967) put forward the idea that these $10^5 M_\odot$ objects would not first fragment into a lot of stars but would form one supermassive star: an "uhrstar". These uhrstars heat up the uncondensed gas to temperatures of the order of 10^5–10^6 °K. The high temperature prevents the formation of any further supermassive objects. Galaxies then form by a secondary process of thermal instability in the hot plasma.

Whether or not one chooses to make such an identification, it is clear that the mass is small enough that it can be regarded as a basic building block for larger structures. In his first paper on galaxy formation, Peebles (1965) considered the details of the agglomeration of such objects into larger systems. He concluded that one could certainly explain the origin of galaxies this way and, depending on the power spectrum of the original density distribution, one might even explain the origin of clusters of galaxies. What distinguishes a galaxy from a cluster of galaxies in this picture is that a galaxy is the largest entity within which a significant amount of mixing of the material has been taking place. Thus one does not expect, on the basis of such a theory, to find any preferred scale of clustering from the scale of a galaxy ($\sim 10^{11} M_\odot$) upwards to the largest aggregates of matter in the universe ($\sim 10^{13\,\rm or\,14} M_\odot$). This expectation has received striking support in the statistical analysis of the distribution of galaxies on the sky by Peebles and Hauser, who show that indeed there are no preferred scales of clumping of galaxies. This important result will be discussed further in Sec. VIII.D.

An interesting possibility in this theory is of computing the luminosity function or the mass spectrum of objects that condense out of the universe. This has been attempted by Balko (1971) and by Press and Schecter (1974), but although the results seem fairly impressive, the problem is fraught with difficulties. In both these papers an attempt is made to compute the probability that an object of mass M will condense from the universe. However, in both papers it is considered sufficient that $\delta\rho/\rho$ should reach the value unity for an inhomogeneity to be distinguished as a distinct object. The trouble with this simple view is that some condensations of mass M may end up buried in larger condensations and therefore not be distinguished at the present day as individual objects. Thus instead of simply computing the probability that a given volume V contains a mass whose fractional deviation from the ensemble average is $\delta M/M$, one ought to compute this probability subject to the condition that a slightly larger volume contains a mass whose fractional deviation from the ensemble average is less than $\delta M/M$. This of course complicates the problem enormously.

IV. COSMIC ANGULAR MOMENTUM

One of the central problems associated with the gravitational instability theory for the origin of galaxies is the question of the origin of galactic spin. Intuitively, one might feel that the appearance of spin in a fluid which is originally irrotational somehow violates locally the principle of conservation of angular momentum. This feeling finds its mathematical expression in Kelvin's Circulation Theorem [Helmoltz, 1858; Kelvin, 1869; a good, modern discussion of this theorem is contained in Batchelor (1970b) Sec. 5.3]. Roughly speaking, the circulation theorem states that, in the absence of dissipative processes, an initially irrotational velocity field remains irrotational. It might seem therefore that it is necessary to postulate the existence of vorticity *ab initio* as is done in the cosmic turbulence theory. Such a conclusion rests on the belief that the circulation theorem is indeed applicable. It suffices to remark at this juncture that galaxies have undoubtedly suffered a great deal of dissipation before arriving at their present states. During the collapse of a galaxy therefore the circulation theorem is clearly inapplicable.

Two suggestions have been put forward to explain the origin of galactic rotation in the gravitational instability picture. One suggestion, due independently to Hoyle (1951) and Peebles (1969b), is that galaxies acquired spin as a result of the influence of tidal stresses. The other idea, due to Doroshkevich (1973) and Zel'dovich and Sunyaev (1972), is that galaxies formed during the collapse of a protocluster of galaxies at the time of maximum compression and dissipation. We shall discuss these two ideas and the arguments against them below.

The principle of the tidal torques theory can be understood by inspection of Fig. 2, where motions are produced in an extended body of mass M as a result of the tidal stresses induced by a neighboring body of mass M_0. Since the parts A of M nearest to M_0 experience a greater force towards M_0 than those parts B of M away from M_0, the body M suffers a torque about its center of mass. The tidally induced motions therefore possess a net angular momentum, but as can easily be verified by considering the accelerations of A and B, the motions have zero vorticity. *The induced velocity field is a shear flow.*[12] The circulation theorem is not violated, and the total angular momentum is conserved since the two bodies gain orbital angular momentum (M_0 experiences a net component of force perpendicular to the line joining the centers of the bodies). The magnitude of the induced angular momentum is easily shown to be on the order of

$$|H| \simeq 3(GMM_0/d^3)\,\epsilon k^2 T$$

where ϵ is a factor accounting for the orientation and shape of the protogalaxy, and k is its radius of gyration. T is the time scale over which the torque acts, and d is the separation of the bodies. The values ascribed to these parameters depend on the particular situation. Given that the torque is strong enough to account for the angular momentum of a galaxy (this question is discussed below), the problem one faces is that of converting the induced shear flow into a vortical flow associated with the internal motions of galaxies.

Hoyle originally applied the theory to the problem of galaxy formation in the steady-state theory. There, he identified the torque-induced body M_0 as the cluster of galaxies in which the newly formed galaxy was situated. The reason for this choice was that the angular momentum is maximized by maximizing the ratio M_0/d^3. The

[12]Batchelor (1970b) discusses the subtle differences between *shearing* and *vortex* motions. A comprehensive discussion of vortex flows is given in Chap. 5 of Boyantovich and Radok (1964), where there is also a summary of the conditions under which vortices may form.

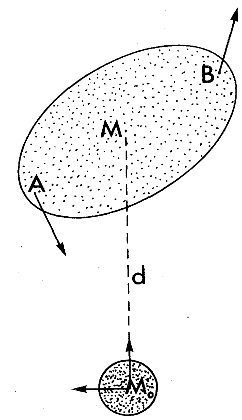

FIG. 2. The tidal action of a mass M_0 on an extended body of mass M. The arrows represent the accelerations *relative to the mass center of M*. There is a net torque about the mass center of M, and M_0 acquires a component of orbital angular momentum as a result of the component of force acting on M_0 perpendicular to the line joining the mass centers of M and M_0.

time T was chosen to be the collapse time scale of the protogalaxy, this being the time scale on which the radius of gyration k changed. By inserting for M_0/d^3 the typical density of a cluster of galaxies, and for k the initial radius of the galaxy, and T, the collapse time scale of a galaxy, a value for the angular momentum typical for that found for large galaxies, was found. The idea is of course subject to the usual criticism that galaxies in the field do not appear to be deficient in angular momentum.

The problem is not quite so simple as regards the gravitational instability picture for galaxy formation in an expanding Friedmann universe. Consider the simple situation where the two masses M and M_0 represent similar neighboring protogalaxies. At times when both protogalaxies are merely small perturbations to the background, it is the *fluctuation* in mass associated with the perturbation M_0 that is the source of the strain acting on the protogalaxy M. The rate at which M acquires angular momentum is in this regime given by

$$\frac{d}{dt}|H| \simeq \frac{3GMM_0}{d^3}\left(\frac{\delta\rho}{\rho}\right)k^2 \ , \tag{5}$$

where d is the separation of the protogalaxies. The rate of angular momentum transfer therefore attains a maximum when $\delta\rho/\rho$ reaches the value unity. At this epoch, both protogalaxies start to collapse and angular momentum transfer becomes less efficient thereafter owing to the decrease in the radius of gyration. If the separation of the protogalaxies is comparable with their diameters, the induced angular momentum is given in order of magnitude by

$$|H| \sim (\epsilon GM^2/R_{\max})\, t_0 \sim \epsilon MR^2_{\max}/t_0 \ , \tag{6}$$

where t_0 is the epoch at which the protogalaxies begin to collapse, that is, the epoch of galaxy formation; $R_{\max}$ is the maximum radius of the protogalaxy at this epoch, and ϵ is the efficiency factor accounting for the orientation and shape of the protogalaxies and the fact that the separations are in fact different from the diameters. Taking the radius of the protogalaxy to be 50 kparsec, in accordance with Eggen, Lynden-Bell, and Sandage

(1962), and the epoch of galaxy formation corresponding to a redshift of 30, yields a value for the angular momentum of the galaxy of

$$|H| \simeq 10^{75}\, \mathrm{cm^2 g\, s^{-1}}\,.$$

Considering the crudeness of the arguments, this value is agreeably close to the value commonly ascribed to the angular momenta of galaxies. Indeed, the more rigorous calculation of Peebles (1969) yields a value that is a factor 5 less than Innanen's value (Innanen, 1966) for the angular momentum of the galaxy. This factor 5 deficiency has been found independently by Jones (1975) and by Silk (1974). The numerical simulation of galaxy formation by Peebles (1971b) yielded a similar deficiency. Whether this deficiency is to be regarded as a serious point of objection to the tidal torques theory is a matter for debate. As can be seen from Eq. (6), the predicted angular momentum is fairly sensitive to the epoch of galaxy formation, though it is unlikely that one can account for the whole factor 5 in this way. A more important point, perhaps, is that the angular momentum of a galaxy is not really well determined. There is now some evidence that even spiral galaxies possess extensive haloes, which could contain a large amount of mass (see Sec. VIII.A).

Let us now turn our attention to the criticisms that have been levelled against the tidal torques theory. The earliest criticisms (Harrison, 1971; Oort, 1970) against the theory were that for various reasons Peebles' estimate for the angular momentum was excessive. The subsequent numerical simulation of Peebles (1971b) was in essential agreement with his earlier theoretical calculation, and would seem to resolve any doubts as to the adequacy of the theory to account for the angular momentum in order of magnitude. However, the numerical calculation falls somewhat short of explaining the discrepancy between the analytic results of Peebles and those of Harrison. Harrison's argument was that the torque experienced by a protogalaxy depends in detail on the distribution of matter about the galaxy. Since density perturbations grow at the expense of the surrounding material, he argued that the mass fluctuation $\delta M/M$ that comes into evaluating the tidal torque [cf. Eq. (5)] is not given by the mean square deviation of the density about its mean value. To illustrate the point, Harrison evaluated the angular momentum transfer due to the tidal action of a local region of density enhancement surrounded by a shell wherein the density was below average. The resultant angular momentum transfer was several orders of magnitude lower than that estimated by Peebles. It seems, however, that Peebles' analytic calculation (Peebles, 1969b), which is a statistical calculation, ought automatically to take account of this effect, and that the realization evaluated by Harrison was simply a statistically unlikely one. A detailed rebuttal of Oort's calculation is contained in Peebles' second paper (Peebles, 1971b) on the tidal torques theory.

More recently, Hunter (1970), Tomita (1973) and Binney (1974), have argued against the theory on the grounds that it fails to produce any *vorticity*; rather, the induced motions are shearing motions.[12] Peebles had only computed the expectation value of the magnitude of the angular momentum without enquiring in detail as to the precise nature of the induced flow. It is possible to perform an analogous calculation for the magnitude of the vorticity and show that this remains zero in the absence of dissipation. Peebles (1973a) published a rebuttal of Tomita's arguments, and in the absence of further analytical calculations one can only restate and reinforce his arguments. One should remark firstly that in order for a protogalaxy to settle down to the kind of disklike configuration presently observed, a considerable amount of dissipation must take place. Whatever the details of this dissipation may be, it is certain that the Kelvin circulation theorem is violated during the relaxation of the galaxy. Consider first the case of a gaseous protogalaxy. During the collapse, the density will increase, the Reynold's number of the shear flow increases, and shock waves will inevitably form owing to the lack of symmetry of the initial state. It is difficult to imagine that vorticity is not generated by such a process. If the initial protogalaxy is stellar, the final state will be achieved by violent relaxation. Violent relaxation is a collisionless process, so the total angular momentum of the stellar system is conserved. However, contrary to the assertion made by Hunter (1970), the fact that violent relaxation is collisionless does *not* imply that vorticity is not generated. Although Lynden-Bell proves a circulation theorem for a stellar system undergoing violent relaxation (Lynden-Bell, 1967b), the conservation of vorticity referred to in that theorem does not refer to a conservation along stream lines of the mean flow, but rather conservation of vorticity along a path in phase space identified with a particular set of particles. The problem of the generation of vorticity during the collapse of a protogalaxy endowed with shearing motion will undoubtedly be an active area of future research.

Assuming that these problems are resolved in favor of the tidal torques theory, the following points are of relevance. Firstly, in binary galaxies one would expect the orbital angular momentum to be equal to the spin angular momentum of the components, and one would expect also that the planes of rotation of the galaxies and of the orbit should coincide. This would only be true insofar as the binary system had not interacted with a third galaxy since its formation. However, since such triple encounters would be relatively rare, it should be possible to make a meaningful test of these elementary predictions of the theory. There seems to be no reason to expect any particular set of orientations if galaxies formed according to the precepts of the cosmic turbulence theory.[13] Secondly, one might expect isolated field galaxies, if such exist, to have on average less angular momentum than cluster galaxies, since there is a lower *a priori* probability that such a galaxy had a close neighbor. Owing to the difficulty in determining the angular momentum of distant galaxies, it is unlikely that this could be

[13] Holmberg (1969) studied the distribution of satellites of Shapley–Ames galaxies relative to the parent galaxy. The observation that these satellites tend to lie away from the plane of the parent galaxy should not be construed as evidence against the tidal torques theory, since most of these satellites are far too small to have been responsible for tidally inducing the spin of the parent galaxy.

tested unless some other indicator of angular momentum, for example, morphological type, is found. If morphological type were indeed correlated with angular momentum, as has been suggested by Brosche (1973), one might expect to find a systematic difference between the morphological types of field and cluster galaxies (Neymann, Scott, and Zonn, 1962). The trouble here is that any such differences could easily be attributed to other causes.

According to the gravitational instability picture, clusters of galaxies would possess angular momentum only if they were formed in the vicinity of another cluster of galaxies. Thus isolated clusters of galaxies, like the Coma cluster, would not be rotating, whereas clusters like the Hercules cluster or Abell 2197, which condensed in the vicinity of another cluster, would possess rotation (Jones, 1976). On the other hand, the cosmic turbulence theory requires angular momentum wherever there is an excess in density, and so even isolated clusters of galaxies should possess some rotation. The only cluster of galaxies for which there is at the moment sufficient radial velocity data is the Coma cluster (Rood *et al.*, 1972; Forêts and Schneider, 1973; Gunn and Sargent, 1974). There is no evidence for rotation, though the data nonetheless allows a significant amount of angular momentum. In the absence of sufficient radial velocity data, one might construe any systematic alignment of the galaxies as evidence for systematic rotation of the cluster. Such an effect has been sought by Brown (1964, 1968), though his findings have not been confirmed by the subsequent more detailed analysis of Hawley and Peebles (1976).

The question of the existence of a rotating supercluster is a still current point of debate. Sandage, Tammann, and Hardy (1972) find no evidence for any significant systematic rotation, whereas on the other hand, Stewart and Sciama (1967) and de Vaucouleurs (de Vaucouleurs, 1953; de Vaucouleurs and Peters, 1968) find evidence for considerable rotation. The evidence put forward by these latter authors is rather unconvincing, though if such a rotation as claimed should exist it would be hard to explain by any theory of galaxy formation.

V. THE FORMATION OF CLUSTERS OF GALAXIES

The problem of the origin of clusters of galaxies is of special interest because in both the gravitational instability picture and in the cosmic turbulence theory clusters of galaxies are thought to have formed by the collapse of density fluctuations (Peebles and Yu, 1970; Ozernoi and Chibisov, 1972). In the gravitational instability picture density fluctuations on the appropriate mass scale are thought to have existed *ab initio*. On the other hand, in the turbulence theory, the relevant density perturbations are generated by the turbulence.

Although there exist numerous versions of what might happen during the collapse of a cluster, it is necessary at the outset to reconcile the basic theoretical view that clusters of galaxies are gravitationally bound with the observations. The luminous mass in clusters of galaxies appears inadequate to bind a cluster gravitationally by almost a factor 10. The nature of the "missing mass" required to bind the clusters of galaxies is certainly one of the outstanding problems of modern cosmology. If one denies the existence of any missing mass it becomes necessary to find a cause for the apparent expansion of the clusters of galaxies. Two explanations have been put forward which are of interest in the present context. It has been suggested by Noedlinger (1970) that a violent event taking place in a cluster of galaxies in the relatively recent past might drive out sufficient gas to force the cluster to become unbound. This idea is open to two objections. Firstly, Hills (1973) has pointed out that the free-free emission from the heated gas of all the clusters would be visible at the present time in the x-ray band. Secondly, the phenomenon is not confined to large clusters of galaxies, but seems to exist on virtually all scales down to small groups of three or four galaxies. The possible detection of gravitational waves by Weber (1970) in 1970 led to the suggestion (Field and Saslaw, 1971; Dearborn, 1973) that clusters of galaxies were more massive in the past and may indeed have been bound at the time of their formation. However, in view of the difficulty in explaining Weber's detected flux in terms of known mechanisms of gravitational radiation, and in view of the failure of other groups to confirm Weber's results, such a suggestion must be considered somewhat implausible. In what follows, therefore, it will be expedient to assume that the clusters of galaxies are bound gravitationally by some unseen component [see, for example, Turnrose and Rood (1970), or the review by Tarter and Silk (1974)].

Indirect evidence that galaxy clusters are gravitationally bound comes from the "infall" models for the x-ray emission of some rich clusters of galaxies. These models were proposed by Gunn and Gott (1972) and examined in considerable detail by Lea (1974). In these models, gas that has not condensed to form galaxies falls into the cluster and is heated to high enough temperatures to make it observable in x rays. Of course, these models are not without their difficulties, and there are alternative theories such as that proposed by Yahil and Ostriker (1973), where the x-ray emission is attributed to an outflow of hot gas from galaxies. The main difficulty with the infall model is that, unless there is some source of heat to the gas, the temperature distribution of the gas is more strongly peaked than the observations would indicate. One possible source of heat which may resolve this difficulty is the heat input due to the stirring of the gas by galaxy motions.

It has been pointed out by Doroshkevich (1971) that clusters of galaxies will not in general collapse isotropically, since any statistical variation in the shape would be amplified during the collapse. The problem of collapse through a series of oblate configurations has been considered by Sunyaev and Zel'dovich (1972), and the converse problem of the collapse to a prolate configuration has been considered by Icke (1973). It is interesting to note in passing that when Rood and Sastry (1971) classified the shapes of clusters of galaxies, they found a rather greater number of spherical systems than this kind of argument would lead us to expect. Icke (1973), Stein (1973), Binney (1974), and Sunyaev and Zel'dovich (1972) all have considered the problem of the nonsymmetric collapse of a cluster of galaxies. All argue that galaxies form as a result of the collapse of the cluster with which they are associated. Angular momentum of the galaxies is acquired at the time of their formation, and there is no need to resort to tidal torques. With this view of galaxy formation one might expect a correlation between the orientations of the galaxies in the clusters and the shape of the cluster. Although some such correlation has been claimed by Brown (1964, 1968), the more recent analysis of Hawley and Peebles (1976) fails to confirm his conclusions. On the other hand, Sastry (1968) found a correlation between the shape of a cluster of galaxies and the orientation of the central galaxy, and Rood and Sastry's (1972) investigation of the cluster Abell 2199 seems to reveal a systematic orientation of the galaxies in that cluster. Perhaps there is some evidence that galaxies formed after the condensation of clusters of galaxies, in that it is difficult to understand how apparently flattened systems like the Virgo supercluster or the Perseus cluster could have formed if the galaxies had formed first.[14] There is no way for a system of galaxies to lose binding energy.

Two arguments can be levied against the idea that galaxies formed subsequent to the collapse of clusters. Firstly, Peebles has remarked on the basis of statistical analysis of catalogues of galaxies that it is difficult to understand galaxies forming in a way different from clusters. The cosmic power spectrum for the distribution of galaxies shows no breaks nor discontinuities indicative of different formation modes for systems on different scales. This is discussed in more detail in Sec. VIII.D. Secondly, there is evidence that not all galaxies have passed through the centers of clusters of galaxies (Jones, 1976a). Judging from the counts of galaxies summarized by Sandage, Tammann, and Hardy (1972), the local density fluctuation $\delta\rho/\rho$ is less than unity and so cannot yet have collapsed; the local group at least may not have formed during the collapse of the Virgo cluster.

VI. THE COSMIC TURBULENCE THEORY

Perhaps the most significant development in galaxy formation theory of the last five years is the reintroduction of the cosmic turbulence theory by Ozernoi and his co-workers (Ozernoi and Chernin, 1968a, 1968b; Ozernoi and Chibisov, 1971a, 1972). To do justice to the Ozernoi picture, I have presented a review of his theory in Sec. VI.A) without criticism. The description differs slightly from that of Ozernoi and co-workers; the hope is that the different point of view will add to the understanding of the theory rather than confuse. The following sections review other contributions to and criticisms of the theory. Here the discussion is divided into separate considerations of the pre-recombination and post-recombination eras.

A. The theory of Ozernoi *et al.*

There are two important ways in which the modern theory of cosmic turbulence proposed by Ozernoi and his group differs from its predecessors over fifteen years earlier. Firstly, there is a realization that prior to recombination during the matter-dominated era the scale of the turbulence *decreases* because the large-scale motions become "frozen out" as the universe expands. The freezing out process saves the turbulence from catastrophic decay. Secondly, the sound speed in the cosmic medium prior to the epoch of recombination is a significant fraction of the speed of light, but after recombination it has fallen to only $\sim 3\ \mathrm{km\,sec^{-1}}$. Thus random motions which were subsonic prior to recombination are thrown into a state of hypersonic chaos as a result of the recombination process, and large-density fluctuations can be generated after a redshift of 1000 without being postulated to exist *ab initio*.

To understand the concept of the "freezing out" of eddy motions it is necessary to appreciate that energy is transferred in turbulence from large-scale dynamical motions to smaller scales as a result of collisions between randomly moving streams of fluid. This transfer of energy to smaller-scale eddies is often referred to as "the turbulent energy cascade." The time scale for the energy of a large eddy to be transferred to such small scales that it is turned into heat by the action of viscosity is roughly the large eddy rotation (or "turnover") time scale. (A turbulent eddy persists for about one rotation before dissolving in the background fluid motions.) If in the universe the cosmic expansion time scale is shorter than the eddy turnover time scale, the eddy is said to be "frozen in." This terminology stems from the assumption that such a "frozen eddy" will not take part in the turbulent energy cascade, but will nonetheless provide a sufficiently large straining field to support turbulence on smaller scales.

It is convenient (if possible) to resolve the velocity field of the cosmic fluid $\vec{V}$ into two components: one part due to the general expansion of the universe and the other representing deviations from uniform expansion:

$$\vec{V} = \vec{v}_H + \vec{v}_T.$$

The Hubble expansion component relative to an observer O is isotropic and proportional to the distance: $\vec{v}_H = H\vec{r}$. In the situation where $\vec{v}_T$ represents cosmic turbulence, $\vec{v}_T$ is a time-varying random function of position. A first-order description of the velocity field $\vec{v}_T$ is contained in the velocity autocorrelation function

$$v^2(r) = \langle v_T(\vec{\xi})v_T(\vec{\xi}+\vec{r})\rangle_\xi.$$

The angular brackets denote an ensemble average over all pairs of points separated by a vector distance $\vec{r}$. For homogeneous and isotropic turbulence, $v(\vec{r})$ depends only on the magnitude of $\vec{r}$, and it is one of the fundamental tenets of the theory of homogeneous and isotropic turbulence that

$$v(r) \propto r^{1/3}$$

over a range of scales known as the "inertial range". (See Appendix B.)

1. The scale of cosmic turbulence

The evolution of cosmic turbulence can be understood by reference to Fig. 3. The figure is drawn to repre-

[14] It should be remarked that the Virgo supercluster is so large that it cannot have undergone any relaxation. Moreover, the Perseus cluster as a whole is not as highly flattened as the line of galaxies in the central region of the cluster would lead one to believe (Bahcall, 1974).

sent some arbitrary initial epoch t_0 where a primordial velocity spectrum $v(r)$ is shown. Shown also on the diagram is the relative velocity of two points separated by a distance r due to the general background cosmic expansion. This latter velocity attains the speed of light at the horizon, that is, for a scale $r_H = ct$. In general, there will be a maximum radius r_* for which

$$r_* = t\,v(r_*). \tag{7}$$

The scale r_* if it exists plays a vital role in turbulence theory. There are two situations in which the scale r_* is not defined. On the one hand, the turbulence velocity [by which we shall mean $v(r)$] may be less than the cosmic expansion velocity on all scales r. This is the situation of weak cosmic turbulence. On the other hand, the turbulence velocity may reach the speed of light at a radius smaller than the scale r_H. This is the situation of a chaotic cosmology, which will be discussed in Section VII.C. For the purposes of the present section, it will be assumed that the scale r_* defined by Eq. (7) exists.

The scale r_* is the largest one whose dynamical time scale is equal to the cosmic expansion time scale. On scales greater than r_* the hydrodynamic interaction time scale is in excess of the cosmic expansion time scale. These scales are said to be *frozen*. However, on scales r less than r_* hydrodynamic interactions [through the $(\vec{u}\cdot\vec{\nabla})\vec{u}$ term of the Navier–Stokes equation] proceed rapidly and there is a transfer of energy from larger to smaller scales until the viscosity becomes of importance. The hypothesis of Ozernoi and co-workers is that a Kolmogorov spectrum

$$v(r) \propto r^{1/3}, \quad r_\nu < r < r_* \tag{8}$$

is set up on scales r less than r_*. The smallest scale r_ν is the length scale on which the fluid viscosity becomes effective in dissipating hydrodynamic energy. An important condition for the validity of Eq. (8) is that the motion on the largest scale r_* should be subsonic. The fluid motion is then essentially incompressible. The turbulent eddies of scale r_* will generally be referred to as the largest turbulent eddies, the implication being that eddy motions on larger scales are frozen and not a part of the turbulence.

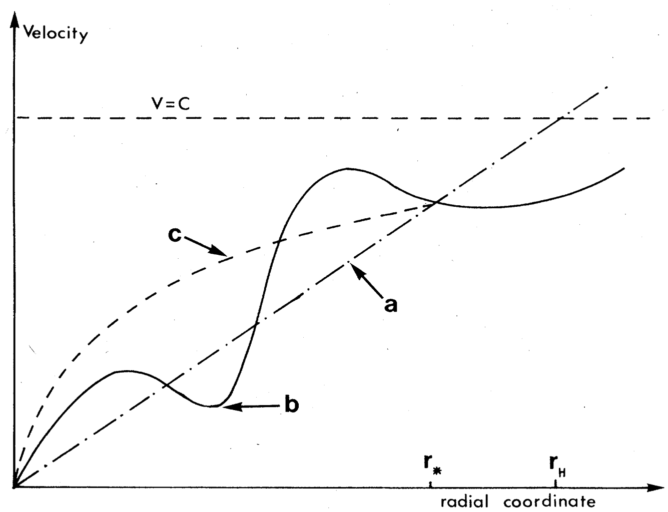

FIG. 3. The spectrum of cosmic turbulence at some early epoch t. The line a represents the Hubble expansion $v_H = Hr$. The radial coordinate r_H where $v_H = C$ is the "horizon." Line b is a schematic representation of an initial spectrum of random velocities superposed on the Hubble flow. Points on this curve lying above the Hubble expansion line a correspond to scales of motion whose hydrodynamic timescale is shorter than the Hubble expansion time H^{-1}. The largest scale on which this happens (if it happens) is denoted by r_*. The motion on scales $r > r_*$ is said to be *frozen*. Dynamical interaction between motions on scales $r < r_*$ results in a transfer of kinetic energy between various scales and perhaps the establishment of a *Kolmogorov spectrum*, depicted by line c. In that case, the motions on scales $r < r_*$ are independent of the initial conditions there (line b). Note that if line b were such that $r_* < r_H$, it would make little sense to talk of a background Friedmann universe. Also, if line b is such that r_* does not exist, the whole spectrum is frozen and the universe is only weakly turbulent.

2. The radiation-dominated era

Consider the evolution of the turbulence spectrum during the radiation-dominated era $t < t_{eq}$. The frozen scales do not transfer any energy to smaller scales and evolve conserving their angular momentum. The angular momentum of a co-expanding eddy in this regime $H \sim \rho_r r^4 v$ remains constant. Thus, since $\rho_r \propto (1+z)^4$ and $r \propto (1+z)^{-1}$,

$$v(\tilde{\omega}) = \text{const}, \quad t < t_{eq} \tag{9}$$

where $\tilde{\omega} = r(1+z)$ is the comoving scale for a region of radius r at redshift z. (A co-expanding eddy therefore is associated with a fixed value of $\tilde{\omega}$ and a particular baryon mass M.) During the radiation-dominated era, therefore, the frozen part of the turbulence spectrum will remain fixed for a particular co-expanding scale. The cosmic expansion velocity on such a scale, however, decreases as $(1+z)$. In other words, the slope of the line representing the cosmic expansion velocity in Fig. 3 decreases, as a consequence of which the largest turbulence scale changes so as to encompass a greater mass (a greater value of $\tilde{\omega}$). Thus *during the radiation-dominated era the scale of the turbulence is increasing*. The way in which the scale increases, however, depends on the precise form of the initial velocity spectrum.

3. The matter-dominated era

We now consider the situation during the matter-dominated era. In the absence of turbulent dissipation the conservation of angular momentum for a co-expanding eddy would lead to a decrease in the rotation velocity in proportion to the redshift,

$$v(\tilde{\omega}) \propto (1+z), \quad t > t_{eq}. \tag{10}$$

If turbulence dissipation is important on the scale r the decline in the velocity will be at least as fast. The cosmic expansion velocity on this scale, however, decreases only as $(1+z)^{1/2}$. Thus *during the matter-dominated era, the comoving scale of the largest turbulent eddy decreases* (because the turbulent velocity on this scale decreases faster than the cosmic expansion velocity). The decline of the scale of the largest turbulent eddy in the matter-dominated era is described by Ozernoi as the *freezing out* of the turbulence.

4. Characteristic mass scales

It is clear from the preceding arguments that the turbulence reaches its maximum comoving extent at epoch t_{eq}. If the turbulent velocity at this epoch is v_{eq}, the mass associated with the largest turbulent eddy is given by

$$M_{eq} \simeq 3\times 10^{15}\,(v_{eq}/c)^3\,(\Omega h^2)^{-2} M_\odot. \quad (11)$$

Motions on mass scales larger than this never become turbulent and retain their primordial form. Since the spectrum of motions is in principle determined by statistical turbulence theory on smaller mass scales, the conditions on mass scales $M < M_{eq}$ at later epochs are independent of the initial conditions. Since the mass M_{eq} is (for reasonably large values of v_{eq}) comparable with masses attributed to galaxies or even clusters of galaxies, in the cosmic turbulence theory the parameters of the galaxies that form is entirely independent of the initial conditions. The one arbitrary parameter of the theory is v_{eq}, though if it is required that the turbulence be subsonic at t_{eq} then $v_{eq} < 0.4c$.

One of the assumptions made by Ozernoi and co-workers is that the turbulence on scales $M < M_*$ can be supported against viscous dissipation by the straining motions of the frozen eddies. With this, the scale of the turbulence at the epoch of recombination is given by

$$M_{rec} = 1.5\times 10^{12}\,(v_{eq}/c)^3(\Omega h^2)^{-17/4} M_\odot. \quad (12)$$

5. After recombination: The condensation of galaxies

The assumption generally made at this point is that the recombination takes place instantaneously or has little effect on the turbulent motions in general. The sound speed after recombination is so low (~5 km sec^{-1}) that all motions are supersonic in the sense that the characteristic velocities are much greater than the sound speed. The motions on scales $M < M_{rec}$ whose hydrodynamic time scales are less than the cosmic expansion time scale generate large density fluctuations. The motions on scales $M > M_{rec}$, being slower than the cosmic expansion, generate density fluctuations only slowly. Thus the scales $M < M_{rec}$ are associated with the formation of galaxies at relatively early epochs, while the scales $M > M_{rec}$ are associated with the later formation of clusters of galaxies. The possibility of choosing the parameters Ωh^2 and v in such a way that M_{rec} is reminiscent of the largest galaxies is an attractive feature of the theory.

If the characteristic radius and velocity associated with motion on scale M are denoted by r_M and v_M we can write down the typical kinetic energy $\langle T\rangle$ and typical gravitational potential energy $\langle U\rangle$ associated with that scale so that the total energy associated with the motions of currents of mass M is

$$\langle T\rangle + \langle U\rangle = \tfrac{1}{2}Mv_M^2 - \tfrac{3}{5}GM^2/r_M. \quad (13)$$

This is positive immediately after recombination. Ozernoi and co-workers now neglect the dissipation of turbulence through the formation of shocks and argue that because of the cosmic expansion, the kinetic energy decreases faster than the magnitude of the potential energy and eventually the inequality

$$\langle T\rangle + \langle U\rangle \leq -\langle T\rangle \quad (14)$$

becomes satisfied. The epoch at which this happens, z_{vir}, is identified with the epoch at which the mass scale M could condense to form a galaxy and corresponds to a redshift (Ozernoi and Chibisov, 1971; Jones, 1973),

$$z_{vir} \simeq \tfrac{2}{15} z_{rec} \simeq 200. \quad (15)$$

Note that all scales condense simultaneously, and this redshift is independent of the details of the hypersonic state appearing just after recombination. This epoch of condensation of galaxies is independent of the parameters of the turbulence and so represents an important test of the hypothesis as put forward by Ozernoi and co-workers.

6. Galactic angular momentum

Another important aspect of the cosmic turbulence theory concerns the angular momentum associated with the mass M_{rec} of the largest galaxies. The mass M_{rec} [cf. Eq. (12)] has radius given by Eq. (7) and its rotation velocity is $v = r/t_{rec}$. The angular momentum of the scale M_{rec} is therefore

$$H = \tfrac{2}{5}Mrv = 2\times 10^{75}(\Omega h^2)^{-1/2}\left(\frac{M_{rec}}{10^{12}M_\odot}\right)^{5/3} \text{g cm}^2\text{sec}^{-1}. \quad (16)$$

If M is chosen to be the mass of a large galaxy, like our Galaxy, the resultant value of the angular momentum is remarkably close to the value deduced by Innanen (1966). This is not, strictly speaking, a prediction of the cosmic turbulence theory, but it is an important cosmological coincidence that allows the cosmic turbulence theory to give *both* the mass and the angular momentum of a galaxy. To determine the angular momentum associated with a galaxy of arbitrary mass requires some assumption about the spectrum of the objects that condense out of the supersonic chaos. To this end Ozernoi assumes that a spectrum of the form

$$v_r \propto r^n \quad (17)$$

will be established during the supersonic era, where the index n is in principle to be determined by some theory of compressible turbulence.[15] $n = 0$ would correspond to a fluid of uncorrelated shock waves, whereas $n = 1$ would correspond to a fluid of fully correlated shock waves, in other words, an N wave. Knowing the velocity spectrum, one can then calculate the mass–angular momentum and the mass–radius relationships of the galaxies that condense out. These relations are found to be

[15]von Hoerner (1958) is often cited as a reference to compressible turbulence. This reference is a summary of results achieved by C. F. von Weizsacker's group in this field. The first mention of von Hoerner's work in this field appears in the IAU–IUTAM conference on Cosmical Aerodynamics in a paper delivered by von Weizsacker (1951) (see p. 201 of that reference). Compressibility in turbulence and supersonic turbulence were also discussed in an interesting, but often overlooked, article by Lighthill (1953) in the Cambridge Symposium on Cosmical Gas Dynamics.

$$H \propto M^{(n+4)/3}, \quad r \propto M^{(2n-1)/3}. \tag{18}$$

A preference is stated for the value of $n = 1$ although the reasons for this choice are not really clear.

B. The pre-recombination era

There are two aspects of the pre-recombination era upon which one can focus attention. There is the question of whether in fact the frozen eddies can support the turbulence against its natural tendency to decay. It is an unproven assumption of the cosmic turbulence theory that the mean straining rate due to these frozen eddies is in fact sufficient to maintain the turbulence on smaller scales. Then one can consider whether or not a Kolmogorov spectrum is in fact established. From the point of view of the predictions of the theory, the fact that the turbulence spectrum is assumed to be Kolmogorov fixes the turbulence scale at the epoch of recombination. Both of these problems are associated with the difficult question of the mode of energy transport by turbulence, and the lack of any simple analytic theory of turbulence makes it very difficult to discuss these problems in anything other than a rather qualitative fashion.

1. The Olsen–Sachs analysis

The rigorous discussion by Olsen and Sachs (1973) of the evolution of the mean square vorticity of turbulence in an expanding perfect fluid is therefore of considerable importance. They find a critical value of the mean square vorticity[16] at epoch t, given by

$$\langle \vec{\omega}^2(t) \rangle_{\rm crit} \simeq 3/t^2. \tag{19}$$

If at this epoch the mean square vorticity exceeds this critical value, the mean square vorticity increases and becomes infinite within a finite time (in the absence of viscosity). If the mean square vorticity is less than this critical value, after a finite time the vorticity decays as $\langle \vec{\omega}^2 \rangle \propto (1+z)^4 \propto t^{-8/3}$. This is the natural decay law for the decrease of vorticity in an expanding universe in the absence of transport processes. Thus the critical value of the mean square vorticity divides the possible turbulence spectra into two classes, frozen and decaying.[17]

If the initial spectrum (at t_{eq}) is such that the mean square vorticity is supercritical, the turbulence will decay through nonlinear inertial range energy transfer. If the intial mean square vorticity is subcritical, the whole spectrum will freeze out. So although the Olsen–Sachs analysis contains no *detailed* spectral information it does provide a criterion for deciding whether or not inertial transfer can win out over the cosmic expansion. It will obviously be of great value to extend the Olsen–Sachs type analysis yet further.

2. Does turbulence decay?

The question of the decay of the turbulence was first brought up by Tomita *et al.* (1970) and later discussed more extensively by Jones (1973b). Jones pointed out that although prior to t_{eq}, when successively larger scales become turbulent, there is always a source of energy for the already decaying turbulence, after t_{eq} the scale of the turbulence decreases, and the support from larger scales is considerably reduced. The turnover time scale of a frozen eddy increases as $(1+z)^{-2}$ whereas the cosmic expansion time scale increases only as $(1+z)^{-3/2}$. There is, therefore, some doubt as to whether the frozen scales can transfer their energy sufficiently fast (and yet remain frozen) to support the turbulence against viscous decay. The consequences of allowing for the free decay of cosmic turbulence during the matter-dominated regime were investigated by Jones (1973b) and by Dallaporta and Lucchin (1972, 1973), who concluded that all the turbulence would decay prior to recombination unless the cosmological model was such that the epoch t_{eq} when the turbulence spectrum is set up and $t_{\rm rec}$ are almost coincident (that is, low Ωh^2).

From the point of view of the cosmic turbulence theory, therefore, the issue of the decay of the turbulence prior to recombination is a crucial one. The Olsen–Sachs result referred to earlier has some bearing on this point. The main contribution to the mean square vorticity comes from the smaller scales; in fact, the mean square vorticity defines a characteristic length scale of the turbulence known as the Taylor microscale, which is associated with the inertial range of the turbulence (Lin and Reid, 1959). So if the turnover time scale of the eddies in the inertial range is short compared with the cosmic expansion time scale, according to the Olsen–Sachs result, inertial transfer takes place efficiently so as to increase the mean square vorticity until a balance with the viscosity is achieved. Inertial range transfer continues while the Olsen–Sachs criterion is satisfied until such a time as the viscosity beats down the vorticity to a subcritical level, in other words, until the turbulence has decayed and is frozen out. The time scale for this process will be the turnover time scale of the eddies in the inertial range.[17]

3. Analytic models of turbulence

Another mode of attack presents itself in the generalization of the so-called "inertial transfer theories" of turbulence to an expanding universe. The equation for the evolution of the energy spectrum of the turbulence in an expanding universe has been written down by Tomita *et al.* (1970):

---

[16]The vorticity of a fluid velocity field $\vec{u}$ is defined as $\vec{\omega} = {\rm curl}\,\vec{u}$. See Appendix B and Batchelor (1970b) for further discussion of the physical interpretation of this quantity.

[17]It would be nice if we could say that $\langle \vec{\omega}^2 \rangle_{\rm crit}$ distinguishes "strong" turbulence from "weak" turbulence; for this would amount to saying that turbulence decays if, for the largest eddies $v/l > t$, or is frozen if $v/l < t$. A complication arises, however, from the fact that the mean square velocity of the turbulence, $\overline{u^2}$, taken with the mean square vorticity $\langle \vec{\omega}^2 \rangle$, defines a lengthscale λ:

$$\lambda^2 = \overline{u^2}/\langle \omega^2 \rangle.$$

λ is called the *Taylor microscale* and is a characteristic of the inertial range. At high Reynolds number, λ is very much smaller than the scale of the largest eddies, l. So, even if the largest eddies turn over on the cosmic expansion timescale, those eddies that are characteristic of the inertial range will rotate much faster and, according to the Olsen–Sachs result, *the turbulence decays until eddies of the Taylor microscale turn over on the cosmic expansion timescale*. It may be interesting to rework the cosmic turbulence theory from this point of view.

$$-\frac{\partial E(k,t)}{\partial t}=T(k,t)+2\left\{\frac{\nu k^2}{R^2}+\frac{(\rho_T R^4)^{\cdot}}{(\rho_T R^4)}\right\}E(k,t) \tag{20}$$

Here $E(k)$ denotes the energy spectrum of the turbulence and $T(k)$ is the inertial transfer term; in this equation $T(k)$ arises out of taking the Fourier transform of the third-order velocity correlation. $T(k)$ is determined in terms of higher-order velocity correlations and so the equation as it stands cannot be solved for the energy spectrum. However, it is possible to replace the actual expression for $T(k)$ with a phenomenological model of the energy transfer. This was first achieved by Obhukoff (1941) and by Heisenberg (1948). Using the Obhukoff variant we obtain an integro-differential equation for the energy spectrum as a function of time:

$$-\int_0^k \frac{\partial E}{\partial t}dp=\frac{\gamma_0}{R}\left(\int_0^k E(p)dp\right)^{1/2}\int_k^\infty p^2E(p)dp$$

$$+2\int_0^k\left\{\frac{\nu p^2}{R^2}+\frac{(\rho_T R^4)^{\cdot}}{\rho_T R^4}\right\}E(p)dp, \tag{21}$$

the solution of which is hopefully the answer to our problem (γ_0 is a coupling constant of order unity). By analogy with the situation in the nonexpanding case (and this can be established rigorously, see Chan and Jones, 1976), the inertial range solution of this equation is the Kolmogorov energy spectrum. In principle one could proceed by solving Eq. (21) numerically for arbitrary initial conditions and in that way we hope to describe the evolution of the spectrum of turbulence in the expanding universe. As yet, this has not been attempted, but the results of any such attempt should clearly be viewed with caution. Firstly, such a scheme is only as good as the inertial transfer theories themselves. The inertial transfer theories are certainly an expression of our intuitive understanding of the energy transfer processes in turbulence and the reasonableness of the theories as a model manifests itself in the ability of the theory to reproduce the Kolmogorov spectrum. However, it is not clear that such a theory has much meaning when the initial spectrum deviates significantly from Kolmogorov. One has only to think of an energy spectrum which is initially zero everywhere but at one frequency. In that case the physical transfer theory certainly fails to produce any spread in the energy spectrum due to the turbulent energy cascade. In particular, one can violate the theorems proved by Olsen and Sachs with regard to the increase in vorticity using such models.

In view of these comments it is certain that the problem of the establishment and maintenance of a Kolmogorov spectrum is an even more difficult one than the problem of the decay of the turbulence. It is not even clear that the *spectrum* of the turbulence is the relevant thing to discuss with regard to the problem of the formation of galaxies. The spectrum merely represents some kind of an ensemble average of one physical quantity and any one realization of the turbulence, that is, any finite region of the universe where galaxies will form, would in fact differ considerably from the average state of affairs described by the spectrum.

C. The "supersonic" regime

Peebles (1971c) pointed out that just after recombination, on scales such that the hydrodynamic time scale was much shorter than the cosmic expansion time scale, there will be a tendency to form gravitationally bound, dense lumps of material. The collisions of supersonic matter currents with one another would lead to strongly compressive isothermal shocks which would result in enormous compressions. A large fraction of the matter in the universe would therefore end up in small very dense bound lumps of material which cannot in any simple straightforward way be identified with galaxies as we see them today. It is rather difficult to see how the cosmic turbulence theory could be altered so as to avoid this problem. An essential feature of the cosmic turbulence theory is its ability to predict both the masses and the rotations of galaxies, and to do this it requires that the protogalaxies correspond to eddies at the epoch of recombination whose hydrodynamic time scale is rather shorter than the cosmic expansion time scale. If the theory were altered so that all eddies turned over slowly compared with the cosmic expansion time scale, one would be faced once again with the problem of explaining the origin of galaxies and the theory would have no more to commend it than the simpler gravitational instability theory.

The highly dissipative nature of the post-recombination era was also remarked upon by Jones (1973b), who pointed out that the velocity spectrum immediately after recombination would evolve through the formation of shocks to the state where the hydrodynamic time scale on all scales was equal to the cosmic expansion time scale. This implies a velocity spectrum

$$v_r \propto r \tag{22}$$

for the residual matter that had not ended up in dense lumps. As described by Ozernoi and Chibisov, this matter would condense to form galaxies at a redshift of 200. Since the velocity spectrum is $v\propto r$, the mass–angular momentum relation of the protogalaxies formed in this way is [cf. Eq. (18)]

$$H\propto M^{5/3}.$$

This in some sense provides some *a posteriori* justification for the choice $n=1$ made by Ozernoi and Chibisov (1971a) for the spectral index of supersonic compressible turbulence [cf. Eq. (17)]. However, it should be remarked that a considerable dissipation of energy must accompany the establishment of a spectrum of the form $v\propto r$. Thus the angular momentum of the protogalaxies estimated from pre-recombination velocities will be an overestimate [cf. Eq. (16)]. Detailed calculations are given in Jones (1973b). An interesting model for the sort of regime that might be envisaged after recombination has been discussed by Silk and Lea (1973).

D. The recombination era

The recombination of the cosmic plasma at $z=1000$ is not an instantaneous process, but follows the Saha law until the ionization drops to about 10^{-5}, taking roughly 20% of the cosmic expansion time (Zel'dovich *et al.*, 1969; Peebles, 1968a). During this period, the photon mean free path increases rapidly and the Thomson drag force acting on the free electrons exerts a significant damping influence on motions on scales up to about 10^{12} $M_{\odot}$. This aspect of the problem was first discussed by Chibisov (1972), who solved a simple radiative transfer problem on the assumption that, even with cosmic turbulence, the ionization history follows the Saha law. It should be noted that Chernin (1970) and Matsuda *et al.* (1971) have argued that the ionization history is affected by the dissipation of the turbulence, which is becoming supersonic through recombination. Their model is, however, somewhat overidealized since they do not consider in any detail the mechanism by which the turbulence could transfer energy to the electrons without also transferring any to the photons. Chan and Jones (Chan and Jones, 1975; Jones and Chan, 1976) have considered this in some detail and find that the electron and radiation temperatures remain equal throughout most of the recombination.

One of the important criticisms of the turbulence theory which has yet to be met is that there may be a tendency to form bound systems of large mass at the recombination epoch (see previous section). The calculations of Peebles (1971c) did not take account of dissipation processes taking place during the recombination, yet these may be of importance. One must consider whether an eddy will be damped by Thomson drag before a large density irregularity is generated. (Of course, this would not weaken the criticism of the theory, since if this happened and no dense lump formed it would only be because there was no motion left on that scale.) The problem has been investigated by Jones and Chan (1976), who found that in universes of low present density ($\Omega h^2 \ll 1$), no eddies of galactic scale survive the recombination damping, or given rise to any dense lumps. This is because of the severity of the radiation damping in these models. In universes of high density, the radiation drag cannot prevent the rapid formation of bound objects at the completion of recombination. Since these objects do not condense until after the recombination is completed, they cannot affect the recombination history.

The recombination history in the presence of turbulence has not so far been considered in any great detail. So while the possibility of collisional ionization causing departures from Saha recombination seems to have been ruled out, there remain several as yet unconsidered processes. Among these it may be remarked that whereas in a uniform medium the recombination to the ground state is inhibited by the great opacity to Lyman-α radiation, in a turbulent medium this may not be so.

E. The generation of density fluctuations by turbulence

The cosmic medium is not perfectly incompressible, and at early epochs there will be a generation of small-amplitude density fluctuations by the turbulence. The *spontaneous* density fluctuation[18] will have amplitude of the order $\delta\rho/\rho \sim m^2$, where m is the turbulence Mach number. As pointed out by Ozernoi and Chibisov (1972) and Silk and Ames (1972), this might be large enough to generate protogalaxies which would evolve in the manner described by the gravitational instability picture (Sec. III). The central theoretical problem is to determine the characteristic scale and amplitude of these turbulence-generated density modes.

The attempt by Silk and Ames (1972) to compute these quantities indeed found a spontaneous density fluctuation of order m^2 on the scale of the largest turbulent eddies. However, as pointed out by Jones (1973b), it is not the spontaneous density fluctuation that is important, but rather the density variation at a point due to the summed contributions from all acoustic waves emitted from other points at earlier times (on the past acoustic cone).[18] The appropriate tool for this problem is Lighthill's theory of noise generation by turbulence (Lighthill, 1952), modified so as to account for the fact that waves can only have propagated a finite distance even though the medium is infinite (Crighton, 1969). Jones (1973b) argued that the acoustic path length was determined by the viscosity. However, as remarked by Stein (1973) and by Matsuda *et al.* (1973), this path length is greater than ct and hence Jones' estimate for the amplitude is excessive. The amplitude given by these authors is

$$\delta\rho/\rho \simeq (v_T/c_s)^{7/2}, \qquad (24)$$

on the scale of the largest eddies.

The characteristic frequency of the acoustic mode is just the turbulence frequency, that is $t_{\rm exp}^{-1}$. The characteristic wavelength is thus the Jeans length. The amplitude of the density fluctuations are not enhanced gravitationally until after recombination, so to produce density fluctuations that collapse fairly soon after recombination requires a fairly large Mach number. Stein, in his model for the formation of galaxy clusters (1973), takes $V/c \sim 0.7$ initially, and considers a flat (V = const on all scales) initial velocity spectrum.

There are still several subtle points to deal with in the problem of the density fluctuation amplitude. Firstly, the characteristic scale of the density variations is the Jeans length: such modes do not in fact propagate until the Jeans mass increases further. Secondly, the turbulent viscosity may be the most important factor limiting the path length traversed by the sound waves; this does not come out of classical Lighthill theory, which is only strictly applicable to sources of finite extent, nor is reabsorption included in Crighton's (1969) extension of Lighthill theory to infinite media. Thirdly, it is not clear that the relevant eddies are necessarily the largest ones: pressure gradient fluctuations are correlated over the much shorter Taylor microscale (Batchelor, 1951), which tends to be characteristic of the inertial range of the turbulence (rather than either of its extremes). [See, for example, Lin and Reid (1959).]

VII. THE EARLIEST EPOCHS

All that has been said in the previous sections relates to dynamical processes occurring at epochs $t > 1$ sec. At

[18]Here the term "spontaneous" is used to describe density variations associated with local variations in pressure caused by turbulent motions. These local pressure variations are correlated over large scales and act as sources of the acoustic noise commonly associated with turbulence. It is important to distinguish the spontaneous variations in density at a point (the "sources") from the acoustic noise field that is radiated to that point from the global distribution of sources.

such times there are no complications due to particle production processes such as are characteristic of the earlier epochs. The universe is for such times a simple expanding hydrogen–helium plasma, complicated only for a short time around a redshift of 1000 when the hydrogen recombines. Indeed, since the casual horizon at $t=1$ sec barely contains $1M_{\odot}$, the physical processes taking place at such epochs cannot be directly related to the formation of galaxies. There are, however, some interesting issues that may be relevant to galaxy formation that are associated with the era $t<1$ sec, for this is essentially the era of the initial conditions for the formation of galaxies.

A. The initial conditions

Let us consider the general nature of the initial conditions. The underlying problem is that, if galaxies are postulated to be the end product of some particular set of initial conditions, the problem of the origin of galaxies has been merely replaced by the problem of the origin of those initial conditions. One would prefer to imagine that the initial conditions were in some sense random and that the physical processes of the primeval fireball have somehow selected and modified just those features that give rise to galaxies as we see them today. Indeed, the dissipative processes of the fireball eliminate much of the small-scale structure and define a characteristic scale at each epoch. *The recombination epoch is a special event which leaves its imprint on the initial conditions by selecting out particular scales.* There are two schools of thought concerning the general nature of the initial conditions [Peebles (1972) has an interesting discussion of these points of view]. On the one hand, there is the view that near $t=0$ the universe was truly homogeneous and isotropic except for small "statistical" fluctuations in the matter distribution (Peebles, 1967a, 1968b; Zel'dovich and Novikov, 1969). These fluctuations correspond classically to the $\sqrt{N}$ fluctuations associated with the distribution of molecules in a gas, and in some sense represent the maximum amount of order that can be possessed by a fluid whose particles are not confined to particular sites. This is to be contrasted with the view that statistical fluctuations exist at arbitrarily early epochs with amplitude in excess of the "thermal" value. According to this latter view, the universe does not tend locally to the Friedmann–Lemaître universe as the singularity is approached. [Such a point of view may be referred to as "Initially Chaotic Cosmology" and is often associated with Misner (1968).] (See also Stewart, 1969; Collins and Stewart, 1971; Matzner and Misner, 1972a, 1972b; Collins and Hawking, 1973.)

Small-amplitude density fluctuations are amplified only slowly by gravitational instability; it is therefore necessary to postulate the existence of statistical fluctuations at very early epochs ($t<10^{-23}$ sec). A considerable amount of argument has centered around the question of whether or not this epoch is *too* early to postulate the existence of statistical fluctuations, since at such times electrons and protons are not contained within their own casual horizons. We have no firm reason to believe that the concept of "particle" has any meaning under such circumstances. Moreover, for $t<10^{-24}$ sec the strong interaction time scale is longer than the cosmic expansion time scale, and the meaning of "equilibrium" has little meaning. This is indeed serious grounds for concern, though Peebles (1968b) has pointed out that if the spectrum of elementary particle mass increases rapidly enough at higher energies, the dominant constituent of the universe may always have a de Broglie wavelength lying within the casual horizon at least as far back as $t=10^{-23}$ sec. Hagedorn (1970) takes a more extreme view of the spectrum of elementary particles and assumes an *exponential* mass spectrum. The result of this is that the de Broglie wavelength is the size of the horizon only at very early epochs, $t\ll 10^{-23}$ sec, and thermal fluctuations have larger amplitude than in the Peebles (1968b) model at the same epoch (Kundt, 1971). In recent years the Hagedorn model has received some support from the success of the Veneziano model for strong interactions (Veneziano, 1968), which also predicts an exponential particle spectrum. As promising as this view might appear, there is an underlying difficulty which will not be easily overcome. When talking about "$\sqrt{N}$ fluctuations" of the number of particles in a volume V, one is in principle comparing the occupation number of differently located identical volumes "at the same time," while the volumes extend beyond the casual horizon. Whereas such a point of view presents little conceptual difficulty in terms of a Newtonian cosmology, it is not at all clear how such a comparison of volumes could be affected in a relativistic cosmology without reference to some *ad hoc* coordinate system. We are once again faced with the question of "what is a density perturbation?" (cf. Sec. III.A.4).

The alternative assumption that the deviations from homogeneity have always been in excess of the "thermal" value is not without difficulty either. One of the attractive features of the Friedmann–Lemaître cosmologies which describe the large-scale structure of the universe so well at the present epoch is their inherent simplicity. To concede the existence of an initially irregular state would be to lose ground on this strong point of the theory, and at the same time demand an explanation for the over-all homogeneity and isotropy of the universe. (Indeed, this was one of the strongest objections put to the cosmic turbulence theory in the early 1950s and may have been responsible for Gamow's reverting back to his early original idea that galaxies resulted from gravitational instability.) Just how far can we deviate from the canonical Friedmann–Lemaître universe and yet maintain that it is still the same simple theory? This is more than a philosophical point because different assumptions about the initial state may occasionally be confronted with observation. Thus one of the attractive features of the simple Hot Big Bang theory is the ready explanation for the observed helium abundance. Abandoning the simple model in favor of some primordial chaos means that the helium abundance is no longer a simply derived parameter.

There is a remark that can be made about initial conditions in both the gravitational instability picture (assuming initial density fluctuations in excess of "thermal") and the cosmic turbulence theory. If the primordial variations in density were large, we might expect to see now a number of "black holes" that resulted from the early collapse of overdense regions (Hawking, 1971) or regions of the universe that are now just emerging from their local cosmic singularity (Novikov, 1965; Ne'eman, 1965; Bahcall and Joss, 1972). The lack of any convincing evidence for such phenomena encourages one to exercise caution before allowing *ad hoc* initial density variations. This seems to constrain the concept of "super-thermal" variations in the initial matter distribution. In the cosmic turbulence theory, although the variation in the peculiar velocity of matter remains finite at all epochs in the past, the associated perturbation to the metric becomes singular as the cosmic singularity in approached and the cosmic expansion is locally non-Friedmann even for $t \sim 10^5$ sec (Zel'dovich and Novikov, 1970). Associated with this is an irregular spatial variation in the local expansion rate, and so to postulate vorticity perturbations at t_{eq} without density fluctuations would seem to demand a rather specially restricted class of initial conditions.

B. The general homogeneity and isotropy of the universe

The problem of the global homogeneity and isotropy of the universe is intimately connected with the problem of the origin of galaxies, (unless it is argued that galaxies grew out of statistical $\sqrt{N}$ fluctuations in the distribution of matter in an otherwise homogeneous and isotropic universe). If one is to postulate an initially chaotic universe containing anisotropies and inhomogeneities of all kinds, the problem of the global homogeneity and isotropy of the universe becomes a serious one. It was first suggested by Misner (1968) that neutrino viscosity acting at times when the cosmic neutrinos decoupled from the other matter in the universe might isotropize an otherwise anisotropic but homogeneous universe. However, subsequent investigations (Stewart, 1969; Collins and Stewart, 1971; Matzner and Misner, 1972a, 1972b) have shown that isotropy is only achieved provided the initial anisotropy is not too great. Collins and Hawking (1973) further showed that the class of spatially homogeneous anisotropic cosmological models which tend towards isotropy of their own accord is of measure zero in the set of all spatially homogeneous cosmological models. It therefore seems unlikely that the initial conditions, if in some sense randomly prescribed, were just such that the universe would be as isotropic as it is presently.

There has recently been considerable interest in the production of particles from the gravitational field. The effect was first discussed by Parker (1969; 1971a, b; 1972a, b) and Sexl and Urbankte (1969), who showed that at sufficiently early times (around the Planck epoch, $t_p \sim 10^{-43}$ sec) the virtual particles making up the quantum vacuum state would gain energy from the gravitational field with the resultant creation of real particle–antiparticle pairs. The pair-production process is particularly efficient in anisotropic universes (unlike the Friedmann–Lemaître universes, they are not conformally flat), and proceeds at the cost of the "anisotropy energy" in the gravitational field. The process, it was thought, might be capable of damping anisotropies in the cosmic expansion at extremely early epochs, as is strongly suggested by the discussions of Zel'dovich (1971) and Zel'dovich and Starobinsky (1972). Of course there are a number of difficulties associated with such a scheme. It is a considerable extrapolation of our known physical laws to go back to such epochs which are on the verge of the domain of quantum gravity. There is, for example, the difficult problem of properly defining the vacuum state in a strong gravitational field. Collins and Hawking (1973) have further remarked that even this process cannot damp out sufficiently great initial anisotropies and we are not much better off than with neutrino viscosity in this respect.

As unsatisfactory as the situation might appear from the point of view of the global homogeneity and isotropy of the universe, from the point of view of galaxy formation it is perhaps just as well that some structure can survive these epochs of damping.

C. Chaotic cosmologies: General speculations

If one is to invoke initial conditions where the deviations from homogeneity and isotropy are in excess of the "thermal" $\sqrt{N}$ fluctuations that might otherwise be postulated, it seems unduly restrictive to argue that the initial conditions were dominated by one particular kind of irregularity rather than another. Indeed, it has already been remarked that to postulate purely vortical motions at some very early epoch, without invoking density variations at still earlier epochs, may not be possible. If it could be established that completely arbitrary initial conditions somehow led to the present state of affairs, that would indeed be an impressive theory, though the present state of knowledge unfortunately leaves us far short of achieving this. Such a broad proposal has been put forward by Rees (1972), who suggests that this scheme may also have some useful by-products.[19] In this picture, the dissipation of the energy of random motions heats up the universe until equipartition between the thermal and dynamic motions is reached. By suitable (but arbitrary) choice for the maximum scale of the chaos ($\sim 10^{17} M$) the amount of entropy thus generated would accord with the present value or 10^{7-9} for the entropy per baryon of the universe, thereby providing a "prediction" of the present temperature of the cosmic microwave background radiation field. Although the existence in the theory of an *ad hoc* maximum mass scale for the chaos detracts somewhat from the picture as a whole, such a limit is nonetheless required by the present constraints on the spectrum and isotropy of the background radiation field.

At this point one might speculate further about the global homogeneity and isotropy of the universe. Part of the reason for the ineffectiveness of neutrinos or gravitational particle production in reducing large anisotropies in the cosmic expansion is that these processes act for only a short period of time in the life of the universe. This is not true, however, of turbulent viscosity (or any other effective viscosity arising out of nonlinear transport processes) which acts as long as there is any turbulence. Moreover, if the turbulence velocity is always close to the speed of light (in some sense this is "maximal chaos"), momentum can be transported a dis-

[19]A similar idea has been discussed by Chernin (1971).

tance $\sim ct$ within time t and the turbulent eddies are in some sense just collisionless. One can easily imagine that under such circumstances anisotropies on scales $\sim ct$ will be damped out. (The anisotropies on scales $< ct$ are determined by the dynamics on scales ct at early epochs.) Of course, large anisotropies will not be damped out at any one instant, but while any motions on scale ct survive there is continual isotropization. The same process might operate for a vacuum cosmology comprising a random ensemble of gravitational waves which had expanded from a singularity, since energy could be transported over a distance ct in a time t by such a system. Indeed, Gowdy (1971a, b) has found a solution of the Einstein field equations which seem to describe such an ensemble of waves, and as the expansion proceeds, the universe becomes globally isotropic. The physical reason for the isotropization remains unclear at this stage.[20]

Some progress has been made towards a detailed, systematic understanding of "cosmic chaos" through the investigation of nonlinear hydrodynamic processes such as the generation of density fluctuations by turbulence (Eidel'man, 1969; Chernin, 1969; Nariai, 1970, 1971; Tomita, 1971, 1972; Silk and Ames, 1972; Jones, 1973b; Stein, 1973; Matsuda *et al.*, 1973), the generation of vorticity from large-scale shearing motions (Ruban and Chernin, 1972; Silk, 1973), the generation of entropy from the dissipation of large-amplitude density fluctuations (Zel'dovich, 1972), and the interaction of acoustic modes with turbulence (Jones, 1970). The eventual goal might, of course, be a theory of supersonic turbulence, towards which a considerable amount of effort was devoted in the 1950s by Von Weizsacker's group, who concentrated on the physics of systems of shock waves.[15] However, at the moment any pretense of understanding anything more than low Mach number subsonic turbulence would be overenthusiastic. It is perhaps here that numerical simulations of comples flows will eventually prove fruitful.

VIII. CONFRONTATION WITH OBSERVATION

Any theory must finally be confronted by a comparison with the observational data, and it is on the basis of such comparisons that the value of a theory ought, in the first instance, to be decided. Of course, since the theories are often motivated by particular observations, some correspondence between observation and theory exists *a fortiori*, so essentially the power of a theory is to be gauged in terms of its predictive capability. At present, however, galaxy formation "theories" are merely idealizations wherein the possibility of *prediction* is not a central issue. They are not "theories" as much as "explanations" or "models" that provide a basis for deeper understanding of the physics of the universe. Accordingly, the best that can be hoped for in confronting cosmogonic theories with observation is to compile a list of relevant observations, and point at those entries which are, or are not, compatible with the theory. Any obvious incompatibility clearly detracts from the value of a theory. Should a theory prove satisfactory at this level, one hopes that it will provide further understanding as to the nature of galaxies, and it is here in some sense that there is any "predictive power" in a cosmogonic theory.

The program is confused by the difficulty in deciding which observations are relevant (some characteristics of galaxies may have developed subsequently to their formation), and even then there is the question as to whether a particular observation has been interpreted correctly. Added to this is the fact that the protostructures discussed by cosmogonic theory are very far removed from the highly evolved objects seen today.

What then might be some relevant observations? In the universe we readily recognize structures such as galaxies and clusters of galaxies and we might enquire which physical properties define these systems. Then we should consult the theories and see if any of the objects of those theories can correspond in a simple way with observation by virtue of possessing similar physical characteristics. As a particular example, the mass and angular momentum of a galaxy are reasonable candidates in this respect since, for some objects at least, we can imagine that these quantities have remained unchanged during much of the evolution of the system. Unfortunately, frustration is encountered even at this early stage of the confrontation: It is extremely difficult to evaluate these parameters for a given galaxy, and even if this is attempted, no obvious characteristic scales emerge. The next step is then to determine an empirical mass function for galaxies in the hope of deriving a mean mass that in some way typifies the system under consideration. Having done that, and bearing in mind the great uncertainties, should we be disappointed if this "typical mass" differs from the characteristic mass derived for the theory by an order of magnitude? This particular example is by no means atypical of the situations frequently encountered, yet it is an important one since theories of galaxy formation have generally been motivated principally by the wish to "explain" the masses and angular momenta of galaxies.

When one considers the available data on galaxies, one is struck by the following basic impression.[21] (1) There are two quite distinct morphological types of galaxy: disk galaxies (S systems) and elliptical galaxies (E systems). (2) Disk galaxies rotate, and we can therefore obtain dynamical models of their central regions. (3) Galaxies are clustered on several scales, from groups containing a few members to extensive structures containing over 1000 objects; we are particularly impressed by the rich clouds of galaxies. (4) There is a large variety of clusters, but generally speaking, the more compact rich clusters are regular and dominated by elliptical galaxies. (5) The luminosity functions of rich clusters of galaxies are remarkably similar. (6) Many rich clusters of galaxies are dominated by a central "cD" galaxy. (7) There is a larger number of comparatively smaller systems. (8) Most galaxies have an underlying

[20] The Gowdy (1972) universe is locally Kasner-like, whereas the chaotic cosmologies of Belinskii *et al.* (1970) are locally mixmaster-like.

[21] This list is based on discussions with Prof. P. J. E. Peebles at Princeton during 1972.

Population II stellar content and so presumably are in the vicinity of 10^{10} years old, and there are no manifestly "young" systems. (9) The spiral galaxies form a regular sequence $S0, Sa, \ldots, Sd, Sm$ of morphological types; neutral hydrogen content, size of nucleus, and general aspect of the spiral arms is closely related to the morphological type. (10) It is difficult to understand the dynamics of clusters of galaxies without appeal to some as yet unobserved entity.

Beyond the observations in this list, which may all bear some direct relation to the problem of galaxy formation, there is a further set of observations that could be construed as providing further evidence for theories of galaxy formation. (a) Galaxy clusters may themselves be clustered to form superclusters of galaxies. (b) Globular clusters are associated with most nearby galaxies, and similar systems are also found at great distances from galaxies. (c) QSOs are extremely numerous at redshifts $z \sim 2-3$, and may be associated with clusters of galaxies. (d) Galaxies that are not in groups or clusters are rare.

Added to this list of observational impressions is a list of questions which, when answered, may provide an important clue to the origin of galaxies. For example, just how clear a distinction is there between rich clusters of galaxies on the one hand, and loose aggregates with only a few members on the other? Again, one might wonder what the primordial cosmic helium abundance was. Is the apparent uniformity an indication of cosmic nucleosynthesis, or just a comment on the gross similarities between extragalactic systems? The plan of the rest of this section will be to discuss these aspects of observational cosmology individually. It should be stressed, however, that the order of discussion is not intended to convey any ordering of importance or of observational status; the discussion has been organized so as to maintain a fairly continuous train of thought.

A. The masses of the galaxies

Since the principle distinguishing feature of a galaxy (over and above recognizing it as a coherent structure) is its size and brightness, it is appropriate to consider first, and in some detail, the masses of galaxies. The subject of the dynamical determination of galaxy masses has been reviewed by Burbidge (Burbidge, 1976), where listings of mass determinations can be found. Recent articles by Roberts (1969, 1976) summarize the 21 cm rotation curve mass data, and Page (1970) has given a list of mass determination of binary galaxies. Mass determinations of groups and clusters of galaxies have been summarized in the Santa Barbara Conference Proceedings (Neyman *et al.*, 1961), and more recently by Rood, Rothman, and Turnrose (1970) and by Oemler (1973).

The methods by which the mass of a galaxy or system of galaxies may be determined dynamically are as follows:

1. The internal dynamics of the galaxy

The internal dynamics of the galaxy can be found from detailed observations of the velocity field over the face of the object. This is easiest for systems containing neutral hydrogen which show good optical emission lines, or which can be observed in 21 cm emission. A rotation curve is constructed and a mass distribution fitted; by extrapolating the mass distribution to infinite radius the "total" mass of the galaxy is estimated (Burbidge, 1976). The mass within the observed region is thus fairly well determined, but beyond that it is model dependent. The agreement between the 21 cm and optical rotation curves (in those cases where the resolution makes such a comparison meaningful) is good in the central regions, but deteriorates at larger radii as the limit of observation is reached. For a sufficiently centrally condensed galaxy, the rotation velocity $V(r)$ at radius r would fall as $r^{-1/2}$ for large r. While there is some evidence for this from some of the optical rotation curves, Roberts and Rots (1973) have shown curves where $V(r) \sim$ constant at large radii, with no sign of falling off. Even though this result has recently been questioned (Emerson and Baldwin, 1973), the lesson to be learned is that extrapolation of the mass distribution beyond the domain of observation must be made with caution. The situation is still less satisfactory as regards galaxies that have no gaseous component; rotation curves must then be based on broad optical absorption lines. Determining the mass is further frustrated by the difficulty of observing the lines very far from the central regions of the galaxy. Recently, King and Minkowski (1972) have improved the situation by combining detailed surface photometry with high-resolution spectroscopy. The photometry determines an appropriate model with which to extrapolate the mass distribution beyond the region where the velocity data is available. This procedure should yield good results for elliptical galaxies where good photometry can be fitted by a dynamical model [see, for example, King (1966)]; however, the method should be applied with caution to disk galaxies. Nordseick (1973a, b) has fitted exponential disks to the surface photometry of several disk galaxies and thereby deduced a mass for the disk; however, the photometry rarely extends far enough for the disk to be unambiguously defined (I shall return to this point in connection with the angular momentum determinations).

2. Binary galaxies

Binary galaxies provide a means of estimating masses by using Kepler's law together with statistical assumptions about the nature and orientation of the orbits.[22] Because of the possibility of line-of-sight coincidences it has been the practice to apply the method to pairs of galaxies that are almost touching, or that show evidence of mutual tidal interaction. Therefore in using close binaries one only monitors the mass distribution within a volume of scale $R \lesssim 50$ kparsec. Masses determined in this way generally agree with masses based on the internal dynamics where the observed rotation curve is fairly extensive. To determine the distribution of matter around galaxies on a scale of hundreds of kiloparsecs it would be necessary to consider widely separated binary systems

[22]This idea seems to have originated with Lundmark (1920; 1926a, b; 1927a, b). Recent work has been summarized by Page (1970) and Dickens and Peach (1972).

(Jones, 1972). One is then up against the serious problem of contamination of the sample by optical pairs, which would lead to anomalously high mass estimates. Widely separated galaxies that display evidence of mutual tidal interaction in the form of a connecting bridge would be useful objects of study in this respect.

3. Dynamical studies of small groups of galaxies

Dynamical studies of small groups of galaxies yield mass estimates that are generally high compared with the previous methods. Here the problem of identifying group members by criteria other than similarity of recession velocity is acute: For example, Gott *et al.* (1973) have shown that two of the nearby groups listed by de Vaucouleurs (1976) are only an apparent association of galaxies; assigning a dynamical mass to these systems therefore has little meaning. Geller and Peebles (1973) have presented a method of analyzing the dynamics of groups statistically without having to identify the individual groups. The method is therefore free from the influence of selection effects and one important aspect of future research on this problem will be to update their analysis as more and better radial velocities become available. Their result can be expressed by saying that a mean mass per galaxy some ten times the mass commonly attributed to galaxies on the basis of rotation curves is indicated by the dynamics of relatively nearby galaxies ($D < 15$ Mparsec). The result is supported by the detailed analysis of the individual groups by Rood *et al.* (1970). The method gives no indication where the mass would lie; however, it has been remarked by Jones (1972) that the more extensive the system used to determine masses, the greater the indicated mass. This may suggest the existence of extended and massive haloes around the galaxies; however, one ought to be careful that the apparent correlation between mass and scale is not merely an artifact of the way the data is selected and plotted.

4. Large aggregates of galaxies

Large aggregates of galaxies have been used to determine masses of galaxies, and it is here that the problem of the "missing mass" makes itself felt most strongly. Not only are the deduced masses per galaxy some thirty times the rotation curve masses, but there are also strong constraints on the nature and distribution of any hypothetical missing mass from observations of the x-ray flux, $H\beta$ flux, and direct attempts to observe neutral hydrogen or the light from diffuse stellar components (Turnrose and Rood, 1970; Tarter and Silk, 1974). It seems unlikely that the intergalactic matter that has been detected so far could account for the mass discrepancy, and the question must remain open. An excellent review of the situation has been given by Tarter and Silk (1974).

On the theoretical side there are several indications that there is more to spiral galaxies than the disks that are seen optically. Ostriker and Peebles (1973) have considered the stability of cold, uniformly rotating stellar disks and concluded that they are secularly unstable to the development of bar-shaped instabilities. One way of stabilizing such a disk would be to embed it in an extended spherical halo of greater mass than the disk; Ostriker and Peebles considered the effects of halo:disk mass ratios up to $2\frac{1}{2}$:1. Belton and Brandt (1963) and Vandervoort (1970) have mentioned the possibility of large haloes on the basis of the dynamics of stellar motions within the Galaxy: our present ignorance of stellar motions at heights $z > 1$ kparsec above the galactic plane allows a fair degree of latitude in superposing a massive halo, so the most serious constraint would come from the Oort limit on the local density (Oort, 1965).

Bearing in mind the uncertainties just mentioned, it is noticed that the masses determined from rotation curve studies fall within a rather narrow range. The lower end of this range is, of course, a limitation imposed by observational selection; however, there does seem to be a quite definite upper limit to the mass of a galaxy determined in this way. The upper limit for ellipticals is about $2\times 10^{12}\ M_\odot$ and for spirals about $2\times 10^{11}\ M_\odot$, though with the suggestion of Geller and Peebles (1973) and of Jones (1972) that spiral galaxy masses may have been underestimated by a factor ten or so, these two limits would be about the same.

5. The luminosity function of galaxies

The luminosity function of galaxies[23] provides information about the distribution of galaxy masses provided assumptions are made concerning the mass to light ratios of galaxies. It is interesting that the *shapes* of the luminosity functions of several rich clusters of galaxies should be so similar (Abell, 1965; Oemler, 1973), and also similar to the luminosity function deduced for field galaxies (Shapiro, 1971). The luminosity functions expressed as counts of galaxies down to some limiting magnitude show (a) a cutoff at the bright end and (b) a change in slope around 3^m fainter than the bright end (the so-called "knee"). The apparent universality of the shape makes it tempting to argue for the universality of either of these characteristic features; thus several workers argue that the "knee" occurs at the same absolute magnitude everywhere, while others argue that the bright end cutoff occurs at the same absolute magnitude everywhere. The problem is complicated by the fact that the knee is not generally very pronounced, nor is there any consensus of opinion regarding the shape of the bright end of the luminosity function. An outstanding problem associated with the cluster luminosity function is a detailed evaluation of the selection effects involved in constructing it; systems of low surface brightness seem to be more common among intrinsically faint galaxies, yet there has been little discussion of the contribution of such systems to the luminosity function. An important point that does not seem to have been taken into account in the statistical analyses of Neymann, Scott, and co-workers (Neymann and Scott, 1959, 1961)

[23]The *luminosity function* can be defined as the number of galaxies per unit volume, brighter than absolute magnitude M. The *differential luminosity function* is the number of galaxies per unit volume with absolute magnitudes in the range M, $M+dM$, and from a statistical point of view is more useful than the luminosity function itself. The normalizations of these functions adopted by various authors vary.

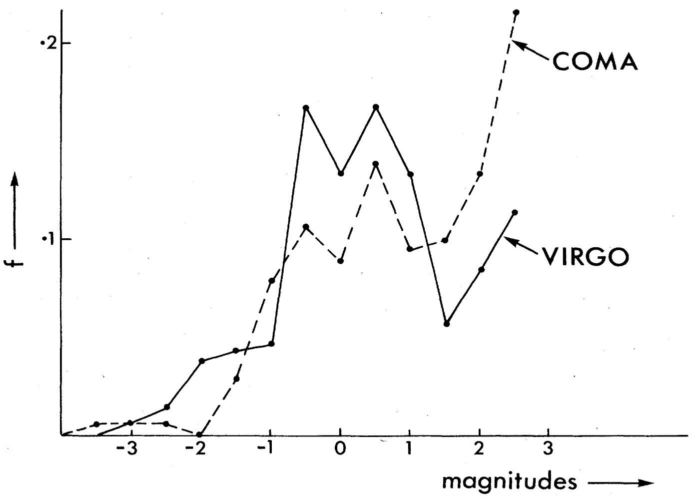

FIG. 4. The differential luminosity functions of the Coma (Rood, 1970) and Virgo (Holmberg, 1969) clusters of galaxies. Plotted are the brightest 210 galaxies in the Virgo cluster, and the brightest 232 galaxies of Coma; no correction for foreground or background galaxies has been made (this is small over most of the range shown). f is the fraction of the galaxies of each sample falling in the magnitude interval m to $m+0.5$. The magnitude scale of each cluster has been shifted so as to emphasize the similarity between the curves. This magnitude shift provides an estimate for the ratio of the distances of the clusters, which is in good agreement with the ratio of the distances inferred from the clusters' recession velocities (Jones, 1976).

is that galaxies are not only selected according to magnitude, but also in regard to surface brightness.

At present, it seems that the differential luminosity function (number of galaxies per unit interval of magnitude) is ever increasing towards fainter absolute magnitudes, and has a feature, or hump, giving rise to the knee of the integrated luminosity function. Figure 4 shows the differential luminosity functions deduced for field galaxies and the Virgo cluster by Holmberg (1969) and for the Coma cluster by Rood (1969); this is perhaps the best data available. Holmberg comments that for field galaxies the "hump" can be attributed entirely to *Sa*, *Sb*, and *Sc* galaxies; the luminosity function for the *E*, *S*0, Irr galaxies is then an exponential of the type first suggested by Zwicky (1957), and the luminosity function for the spirals is a Gaussian, as suggested originally by Hubble (1936; see also Neymann and Scott, 1959). In the Virgo cluster, Holmberg notes a residual hump in the *E*, *S*0, Irr luminosity function that could be indicative of a transmutation of spirals into *S*0s by some means. Rood comments with regard to the Coma cluster that the hump is much reduced in the luminosity function for galaxies not near the center of the cluster and so may be partly attributed to an excess of intrinsically bright galaxies near the cluster center.

The knee in the luminosity function for the Coma cluster occurs around absolute visual magnitude $M_V^* = -19.5$ (for a Hubble constant of 100 km s^{-1} Mparsec^{-1}), and for objects with mass to light ratio of 30 (King and Minkowski, 1972) this translates into about $2\times10^{11}\,M_\odot$. Abell has commented (see Peebles and Yu, 1970) that according to his luminosity function, the total luminosity in galaxies brighter than M_V is about equal to the total contribution from fainter galaxies; so in this sense also M_V is a characteristic of the luminosity function.

B. The angular momenta and binding energies of galaxies

The angular momentum and binding energy of a galaxy are characteristic parameters that have played a central role in discussions of theories of galaxy formation. The rotation curves of galaxies yield the angular momentum distribution within the observed region of the galaxy, and by fitting suitable models it is possible to infer the total angular momentum of the system. This latter quantity is, however, subject to the same criticisms as the extrapolated total masses. Nordseick's article (Nordseick, 1973a, b) on the determination of angular momentum provides a clear discussion of the ambiguities.

The angular momentum of the Galaxy has been determined by Innanen (1966) using a Schmidt-type model:

$$H_{\rm gal} \simeq 2\times10^{74}\ {\rm g\,cm\,s^{-1}}. \qquad (26)$$

The angular momenta deduced from rotation curve studies of galaxies are of the same order (see, for example, Burbidge, 1976). Nordseick's value for the angular momentum of the Galaxy (Nordseick, 1973b) exceeds Innanen's by a factor of ten, even though it is based on the same data. The reason for this higher value lies in the way that Nordseick fitted an exponential disk to the data, which only extends out to 10 kparsec from the galactic center. The scale of Nordseick's disk for the Galaxy is considerably in excess of the value deduced by Freeman (1970) (again from the same data!), which gives an angular momentum similar to Innanen's. Furthermore, if Nordseick's extrapolation were valid, not only would the Galaxy have the largest angular momentum of any galaxy yet studied, but it would also have the lowest central density.

It has frequently been remarked that galaxies of low mass have low angular momenta, and there have been numerous attempts to establish general mass:angular momentum relationships (for example, Heidmann, 1968; Takase and Kinoshita, 1967). However, such relationships should be viewed with caution, as has recently been emphasised by Nordseick (1973b). As an example, if all the galaxies of a sample have the same mean density $\langle\rho\rangle$ (or, equivalently, if they all have the same central density) the masses would vary as R^3, where R is some characteristic radius of the galaxy. Determining the masses from a relationship of the form $M\propto V^2R$, where V is a characteristic velocity associated with the rotation curve, and the angular momenta from $H\propto MVR$ leads *a fortiori* to a mass:angular momentum relationship $H\propto M^{5/3}$. The interpretation of this relationship found by Heidmann (1968) and others is therefore open to some doubt; it may only reflect the way in which the sample of galaxies observed was selected and the observations reduced. Takase and Kinoshita (1967) found $H\propto M^{7/4}$ from optical rotation curves, and this has been strongly criticized by Freeman (1970) as being merely a consequence of the data sample and its analysis.

Interestingly, Freeman (1970) has deduced a relationship of the form $H \propto M^{7/4}$ for disk galaxies from the photometric properties (a) that the disks of spiral galaxies are characterized by one length scale, and (b), that the (extrapolated) central surface brightness of disks varies by only a fraction of a magnitude per square arc second over a wide range of galaxy luminosities. The result would be changed, for example, if there were any systematic variation of central surface brightness with luminosity; this could easily happen and might not be detected on the basis of the data used by Freeman.[24] The work does emphasize the value of good surface photometry.

Thus while some agreement in order of magnitude between the prediction of a theory and the observational value for the angular momentum of a large galaxy is desirable, the observational data do not as yet warrant the demand of "good agreement". Arguments "predicting" a mass:angular momentum relationship can hardly be compared with observation except to say that large systems have greater angular momenta than smaller systems. The method discussed by Nordseick (1973) and Freeman (1970) of combining rotation curve data with extensive photometry will be a powerful tool, particularly when surface brightness distribution at very low levels becomes available.[25]

An outstanding question is whether elliptical galaxies rotate at all. Few attempts have been made as yet to fit models of rotating stellar systems to the observed brightness distribution so it is necessary to rely on indirect indicators of rotation.[26] An obvious indicator might be the degree of flattening of the system, though there are both observational difficulties (like trying to define properly the degree of flattening) and theoretical difficulties (like a lack of suitable theoretical models with which to compare the data) to face. Fish (1964) noted a *lack* of correlation between the flattenings and luminosities of a sample of elliptical galaxies from the HMS catalogue (Humason *et al.*, 1956), and this is confirmed by an analysis of the more extensive catalogue of de Vaucoulers and de Vaucouleurs (1964). Gott (1973) has followed the formation and relaxation of an incipient rotating elliptical galaxy, and found that the general flattening increased with the angular momentum. However, the ellipticity of successive isophotes in his models decreases with increasing scale (because of insignificant relaxation in the outer regions) and is contrary to the generally observed pattern. The same is true of the rotating models of Prendergast and Tomer (1970), and it is difficult to say just how serious a defect this is. There are numerous models of stellar systems that display the core-truncated halo type structure, and these have two parameters with which to fit the observations. Usually a good fit can be obtained to the run of surface brightness with radius and so the isophote shape distribution might be an important discriminant between various models.

Of course, the outer regions may still reflect the varied nature of the initial conditions. Gott's models failed to reproduce Hubble's surface brightness law at large radii, and Gott suggested that the outer regions might have been accreted onto the galaxy once it had formed. The first suggestion that elliptical galaxies have formed by accretion onto some initial condensation was that of Hoyle and Narlikar (1966), and Oemler's recent photometry of the central cD galaxy of the cluster A2670 (Oemler, 1973), might be construed as evidence in support of such a picture.

In the case of elliptical galaxies it might be easier to estimate the binding energy than the angular momentum, since this can be done by fitting a stellar system model to the photometric luminosity profile. The binding energy is an important parameter: a galaxy must be more tightly bound now than at the time of its formation, hence knowledge of the binding energy of a system puts an upper limit on the redshift at which the galaxy could have formed. In 1964, Fish (1964) produced estimates of the binding energies of 29 elliptical galaxies based on the photometric work of himself, Holmberg, and Liller. By fitting a de Vaucouleurs' surface brightness distribution (de Vaucouleurs, 1959) to the data to find the radius $R_{1/2}$ enclosing half the mass of the galaxy, he could use an expression due to Poveda to obtain the binding energy: $|B| = \text{const. } M^2/R_{1/2}$. The masses were estimated from the total luminosity (which itself was found from the fitted model) by using an average mass to light ratio for all the galaxies. Thus for NGC 3379 he obtained a value $|B| \sim 10^{59}$ ergs for the magnitude of the binding energy. ($H_0 = 100$ km s^{-1} Mparsec^{-1}). The number is of course uncertain, but could doubtless be improved with a better estimate for the mass, and a better dynamical model.[27]

Considering the data on all 29 of the galaxies of his sample, Fish deduced a mass–potential energy relationship,

$$|B| \propto M^{3/2}, \tag{27}$$

which commonly bears the name of Fish's Law. As is usual with such relationships, considerable caution is required before one can say that this is indeed the correct relationship (even for the sample under consideration); without proper consideration of the selection

---

[24]There is the implicit assumption in Freeman's analysis that the mass of the galaxy lies mainly in the disk.

[25]It is interesting to note that Nordseick's list of disk galaxies for which good photometry is available contains 17 galaxies. Freeman's list contains 35 galaxies for which he considers the photometry to be good enough to define an exponential disk component. Yet only three galaxies are common to the two lists: the Galaxy, M31, and NGC 5005. This may indicate an urgent need for more extensive surface photometry of disk systems.

[26]Wilson's models of rotating elliptical galaxies can be made to fit the observed surface brightness distribution rather well. This might open up a possible way of investigating the rotation of elliptical galaxies, were it not for the fact that the fitted model is not unique. Thus Wilson's model for NGC 3379 fits no better than a King model (which has no rotation) with an appropriate tidal cutoff (see King, 1966).

[27]The de Vaucouleurs (1959) surface brightness distribution is only an empirical law and does not fit the observed surface brightness distribution so well as a King model (see King, 1966). For the case of NGC 3379, a King model yields a binding energy of

$$|B| \simeq 2.5\times 10^{59} (M/10^{11} M_0)^2 h \text{ ergs},$$

where M is the total mass of the galaxy.

effects and of the way in which the data was analyzed we should not even assert the existence of any relationship! In view of the possible importance of Fish's law it would be of value to compute the binding energies of low surface brightness systems (like Fornax and Draco) and of some compact objects to see if they fit the relationship.

C. The morphology of galaxies

The existence of two distinct morphological types of galaxies ("spiral" or "disklike" and "elliptical") is a striking observation,[28] and since the relaxation time scales for stellar systems are so long, there can hardly be any question of one type having evolved from the other. This division thus indicates that the initial conditions for the formation of these systems were different. It is also striking that the members of each type can be arranged in a linear sequence: The ellipticals can be arranged in a sequence of increasing flattening ($E0 \rightarrow E7$), and the spirals can be arranged in a sequence of increased prominence of spiral arms ($S0 \rightarrow Sd \rightarrow$ Irr). Whether or not these orderings themselves are consequences of different initial conditions will be the subject of some discussion later. It should be remembered that although $E7$ and $S0$ galaxies look similar (in that they are rather flattened stellar systems), they are photometrically quite different (Sandage, 1961; Freeman, 1970).

The spiral galaxies are highly flattened with intrinsic axial ratio $q \sim 0.25$,[29] whereas the intrinsic flattenings of elliptical galaxies lie in the range $q = 0.3-1.0$, with a peak around $q \sim 0.6$ (Sandage *et al.*, 1970). (The exact location of the peak is somewhat uncertain, but the effect was noted by de Vaucouleurs (1961), who in 1961 remarked on an apparent excess of $E4$–$E5$ systems.) Gott (1973) has argued that the dissipationless collapse of a stellar system cannot lead to a configuration flatter than $E4$–$E5$, and so the classification of the few $E6$ and $E7$ systems should be checked carefully. (Systems classified as $E6$ are NGCs 670, 1209, 4386, 4564, 4660, 4697, 4863, 5028, 6875, 6877, and 6909. Systems classified as $E5/6$ are NGC 3377, NGC 7785, IC 4797, and IC 4889).

The different distributions of intrinsic flattenings may provide an important clue to the broad distinction between spirals and ellipticals: the collapse of a protogalaxy to form a disk must be accompanied by a considerable amount of dissipation. One suggestion is that when the bulk of the star formation takes place prior to the collapse of the galaxy, the resulting system will be an elliptical, whereas star formation during the collapse of the protogalaxy would lead to a disk system (King, 1971). Although such a scheme involves a number of assumptions, it seems difficult to obtain highly flattened systems in any other way. The problem of forming stars prior to the collapse of a protoelliptical galaxy is a serious one. One possibility would be that the stars formed even earlier in protoglobular clusters after the manner suggested by Dicke and Peebles (1969). A no less plausible theory is that the initial mass function for star formation might have differed from one galaxy to another (J. Jones, 1976); those galaxies where the stellar mass function favored a large fraction of high-mass stars would lead to disk systems. Such a model would have the advantage of explaining the large range in disk: bulge ratios in spiral galaxies. (This ratio seems to be almost independent of morphological type: Freeman, 1970.)

There have been attempts to identify the whole, or at least a part, of the Hubble sequence as angular momentum sequence. While it is certainly plausible that the sequence $E0$ to $E7$ represents an angular momentum sequence, the considerable structural differences between the ellipticals and disktype systems [emphasized strongly by Sandage *et al.* (1970)] makes it unlikely that the sequence extends into the spirals with $S0$s forming some kind of bridge between the E and S sequences. The sequence $S0 \rightarrow Sd \rightarrow$ Irr could also be an angular momentum sequence: systems of lower intrinsic angular momentum would collapse to greater density and so might consume a greater part of their gas in star formation. Brosche (1970) has developed such a scheme in some detail,[30] and the work of Holmberg (1964) on the mean densities of galaxies lends some support to this. Brosche (1973) has also done some principle components analysis to determine which parameters are needed to define a given galaxy. One of his basic parameters looks rather like the angular momentum, though such conclusions should be viewed with caution since it is not clear what selection effects are operating in choosing particular galaxies for study. The principle arguments against this kind of picture have been given by Sandage, Freeman, and Stokes (1970). They make two important points: (a) Holmberg's densities are based on masses inferred from the integrated photographic magnitude and absorption-corrected color index; these masses for $Sc \rightarrow$ Irr galaxies are systematically lower than the dynamically determined masses of Roberts (1969, 1976) and the difference is sufficient to account for Holmberg's result. (b) There exist pairs of galaxies which have photometrically similar structures but differ in morphological type; the similarity in structure is taken to indicate similarity of mass and angular momentum. Further examination of these two points would be of interest.

Another striking feature of galaxies is their morphological relationship to the kind of cluster in which they are found. Thus spiral galaxies of types $Sc \rightarrow$ Irr seem to be generally absent from rich, *dense* clusters, and are generally found among the loose irregular clusters. The immediate impression is that galactic morphology may be related to the type of cluster a galaxy lies in, and from there one may be tempted to associate the origin of galaxies in some direct way with the formation of clusters. The crucial thing, of course, is to measure for each cluster of galaxies the fraction of *disk* galaxies,

[28] Even Alexander (1852) felt that there was an important *dynamical* difference between the green and white nebulae.

[29] Here q is the ratio of the minor axis to the major axis of an oblate spheroid. Seen edge on, such a system would be classified as "En" with $n = 10(1-q)$.

[30] There are a number of errors in the paper by Brosche (1970), though qualitatively one can hardly doubt the conclusion. It would be interesting to repeat the calculation.

and relate this to, say, the richness of the cluster or the central velocity dispersion. The emphasis on disk systems is important since it may be possible to transform *Sc* galaxies into *S*0s by direct collision with the center of the cluster or another galaxy (Spitzer and Baade, 1951). (In view of the long relaxation times involved, it seems unlikely that disk systems could be changed into ellipticals, or *vice versa*.) Unfortunately, careful classification of the membership of galaxy clusters has not been systematically undertaken; there is always the danger of confusing *S*0s with ellipticals. It has been suggested by Neyman, Scott, and Zonn (1962) that this dependence of galaxy type on the gross properties of the cluster is not real, but due merely to the difficulty of recognizing disk systems at large distances. They constructed a morphological type-luminosity functuon for field galaxies and then estimated the apparent abundance by type of clusters of galaxies at various distances, on the assumption that the membership of the clusters is the same as the field. The results are surprisingly good.

D. The clustering of galaxies

It has long been recognized that galaxies are not randomly distributed on the sky, but often lie in clusters.[31] Early in the days of extragalactic astronomy, Hubble felt that clusters contained but a few percent of all galaxies (Hubble, 1936). However, subsequent surveys have changed that picture. In 1962, van den Bergh (1962) found that only 24% of ellipticals and 52% of spirals and irregulars could not be readily assigned to even a small group ($n \leq 2$) of galaxies. More recently, de Vaucouleurs (1971, 1976) has argued that the number of "true" field galaxies is extremely small and perhaps even zero. Of course, the problem ultimately hinges on the definition of a cluster of galaxies.

When one asks "What is the characteristic scale of clustering of galaxies?" there are several problems to be faced. There is the question of what is a cluster of galaxies, and then there is the problem of obtaining a suitable statistical sample of such objects. Such information might best be sought in counts of galaxies which are complete down to some limiting magnitude, a program initiated by Hubble in the 1930s and continued in greater depth by Shapley and his co-workers,[32] by Zwicky and his co-workers (Zwicky *et al.*, 1961–8), and more recently by Shane and Wirtanen (1967). The last of these contains a considerable amount of data, so much so that it is only with the advent of high-speed computers that it can all be handled effectively.

The pioneering analysis was that of Bok (1934) and Mowbray (1938), who sought simply to characterize the deviations from randomness of the distribution of galaxies on the sky; their main conclusion was that galaxies are not randomly distributed on the sky. The Shane–Wirtanen (1950, 1953) survey of Lick plates provided further impetus for the statistical investigation of galaxy clustering. Here there were two opposing schools of thought about how the clustering of galaxies might be described. On the one hand Neyman, Scott, and coworkers sought to construct statistical models from which they could deduce the size and population distributions of clusters of galaxies (Neyman, Scott, and Shane, 1953, 1956), while on the other hand, Limber (1953, 1954, 1957) and Rubin (1954) sought less detailed information and concentrated on simply characterizing the nonrandomness of the galaxy distribution in terms of the autocorrelation function for the distribution of pairs of galaxies. There are arguments for and against either point of view. The method of Neyman and Scott depends on an assumed model for clusters, and is sensitive to variations in the data from one photographic plate to the next. The method of Limber and Rubin is perhaps not so sensitive to irregularities in the data, but it tells us nothing about the nature of the clustering. (In fact, it only tells us how galaxies are distributed pairwise.) On the basis of their analysis, Neyman *et al.* (1953) concluded that clusters exist, having on average 150–250 member galaxies within a sphere of radius 2 Mparsec ($H_0 = 100$ km/sec/Mparsec).

Abell (1958) later compiled a catalogue of "Rich Clusters of Galaxies." Clearly, at this stage, nobody doubted the existence of galaxy clusters, but Abell (1961) was able to identify a few "superclusters". The existence, or otherwise, of superclusters became a point of controversy. At this level, the problem is more difficult, since the proposed superclusters covered large areas of sky and questions of the homogeneity of the data become important. Kiang and Saslaw (1969) analyzed the distributions of Abell clusters in three dimensions (assigning each cluster a distance on the basis of the assumption that the tenth brightest cluster member is a standard candle). They concluded that superclustering was a real phenomenon, extending over scales possibly as large as 100 Mparsec. However, more recently, Fullerton and Hoover (1972) have questioned the statistical significance of this result (the magnitudes of the tenth brightest cluster galaxies are not randomly distributed, but are clumped so as to give the impression that Abell clusters lie on concentric shells of about 50 Mparsec thickness). Yu and Peebles (1969) performed a power spectrum analysis of the distribution of Abell clusters on the sky. This technique is analogous to the technique used by Limber and Rubin to examine the significance of the apparent clustering of galaxies, and likewise has the advantage that large-scale surface density gradients (due to galactic absorption, for example) should easily be picked out. Yu and Peebles saw no compelling evidence that superclustering is a general phenomenon (of course, their technique cannot rule out the existence of a few superclusters), and this conclusion was supported by Fullerton and Hoover (1972). The situation at that time has been fully and carefully reviewed by de Vaucouleurs (1971).

The first compelling evidence that superclustering is a real phenomenon came with the elegant analysis of Bogart and Wagoner (1973), who considered the nearest

[31]Herschel (1811) undoubtedly noticed the clustering of galaxies in Virgo. The Coma and Perseus clusters were first noted by Wolf (1901, 1905). Hubble and Humason (1931) can be credited with the discovery of the Pisces cluster. Superclusters were first noted and listed by Shapley (1933).

[32]Among the early surveys one can note the following: Fath (1914), Sears (1925), Shapley and Ames (1932), Mayall (1934, 1936), Reiz (1941).

neighbor distribution among Abell clusters in various distance classes. Within a particular distance class, they concluded, a random Abell cluster is more likely to have a close neighbor than would be expected if the clusters were randomly distributed on the sky. Since the nearest neighbors lie within a few degrees of one another, this method too avoids the problems of getting homogeneous data over large areas of the sky. There is, of course, an outstanding problem in resolving the difference between the results of Bogart and Wagoner (1973) and of Yu and Peebles (1969) and Fullerton and Hoover (1972).

There remains the question of the scale of clusters of galaxies and possible superclusters. The complexities and underlying assumptions of the method of Neyman *et al.* (1953, 1956) means that the analysis of the covariance function for the distribution of galaxies may provide the best clues. The complete Shane–Wirtanen (1967) counts were not available to Limber, and it was Totsuji and Kihara (1969) who first deduced the covariance function for the distribution of galaxies observed by Shane and Wirtanen. The probability of finding a galaxy in an elemental volume δV situated at a distance $\vec{r}$ from a randomly chosen galaxy can be written as

$$\delta P = n[1+\xi(|\vec{r}|)]\,\delta V,$$

where $\xi(\vec{r})$, the two-point correlation function for the spatial distribution of galaxies, represents the excess probability (over the random chance) of finding a galaxy in δV.[33] Totsuji and Kihara found that the data were well represented by

$$\xi(r) = (4.7h^{-1}/r_{\mathrm{Mparsec}})^{1.8}, \quad 2\ \mathrm{Mparsec} < r < 25\ \mathrm{Mparsec}.$$

These authors did not comment on this result; however, it is immediately clear that this provides strong evidence for a nonrandom distribution of galaxies with correlations extended over scales far greater than clusters of galaxies. The lack of features in the $\xi(r)$ curve indicates an absence of a preferred scale of large-scale clustering.

Independently, Peebles (1973b) initiated a long-term program to investigate the distribution of galaxies by covariance analysis of catalogues and surveys of galaxies. The data analyzed so far consists of the Zwicky Catalogue (Zwicky, 1961–8) and Shane–Wirtanen (1967) counts by Peebles and Hauser (1974). Peebles (1974 a, b) discusses this data and finds that the combined data, obtained by taking account of the different distances covered by the catalogues, is well fitted by a power law

$$\xi(r) = (5.4h^{-1}/r_{\mathrm{Mparsec}})^{1.77}, \quad 100\ \mathrm{kparsec} < r < 30\ \mathrm{Mparsec}.$$

[The difference between the normalization factors of Totsuji and Kihara (1969) and Peebles (1974 a, b) is probably due to their different assumptions regarding the luminosity function of galaxies.] In the range covered by Peebles' (1974 a) data, one might have expected to pick up features in the covariance function corresponding to a characteristic scale for rich clusters of galaxies. However, the two-point correlation function tells us only about the pairing aspect of galaxies, and says nothing about the distribution of galaxies taken many at a time.

Peebles has also performed a cross-correlation analysis of the Abell and Shane–Wirtanen catalogues (Peebles, 1974 c). Unlike the simple two-point covariance analysis, this does indeed contain information about the clustering of galaxies. The conclusion to be drawn from this analysis is that 12% of bright galaxies are correlated with rich clusters in Abell's distance classes 3, 4, and 5. The correlation between Shane–Wirtanen galaxies and Abell clusters extends over scales of some $35h^{-1}$ Mparsec, though this does not indicate that the Abell clusters are this big. Hauser and Peebles (1973) have shown that the Abell clusters themselves are not randomly distributed, but can be grouped into superclusters (or "clouds") containing on average about two clusters. Thus one might say that "clouds" contain 25% of all galaxies.

Interpretation of this statistical analysis is not straightforward; our naive intuition leads us to think in terms of clustering of galaxies rather than in the relative distribution of galaxies taken two at a time. Thus one must build models for the distribution of galaxies and check the model by comparing its two-point correlation function with that observed by Peebles. The enormous range of possibilities can be narrowed a little by computing higher-order correlation functions, but the amount of computing rapidly becomes prohibitive. Peebles and Groth (1975) have computed the three-point correlation function for the Zwicky catalogue and find that, given two randomly chosen galaxies in the catalogue, there is a significant probability over and above what might be expected on the basis of the two-point correlation function, of finding a third galaxy at a given place. The precise form of the three-point function enables Peebles and Groth (1975) to rule out, for example, a model of clustering wherein (a) all galaxies lie in clusters, (b) cluster centers are randomly distributed, and (c) the density of galaxies in the outer regions of clusters falls off with some universal power of the radius.

Perhaps this is an appropriate point to stress some of the limitations of the covariance analysis of catalogues of galaxies. It has already been pointed out that the two-point correlation function says nothing about high-order clustering, so what can be concluded from $w(\theta)$? That $w(\theta)$ looks smooth can be misleading; the points that make up the $w(\theta)$ curve are not independent. The Fourier transform of $w(\theta)$, the "power spectrum," does not throw any more light on this since the Fourier components are not independent. The function $w(\theta)$ must become negative somewhere (this is required by the definition of the mean density) so $w(\theta)$ cannot behave like a power law on all scales, and there must be a scale where $w(\theta_0)=0$. This latter scale may be $30h^{-1}$ Mparsec if one takes the anlaysis of the Shane–Wirtanen counts at face value. Finally, the matching of the covariance

[33] From a catalog or table of counts of galaxies we deduce $w(\theta)$, the projected *angular covariance function.* $w(\theta)$ is the excess probability δP of finding a galaxy in a solid angle $\delta\Omega$ at a distance θ away from a randomly chosen galaxy. Getting from $w(\theta)$ to the spatial covariance function $\xi(r)$ is not straightforward (see Peebles, 1973). However, a spatial covariance function $\xi(r) \propto r^{-n}$ yields a projected angular covariance function $w(\theta) \propto \theta^{-n+1}$ (the converse is not necessarily true).

functions deduced from the Zwicky catalogue and the Shane–Wirtanen counts depends on the relative distances which the surveys cover; there could be a feature (like a break or discontinuity) where these surveys overlap.

It may well be that to better understand the relevance of clusters of galaxies to the galaxy formation problem it is necessary to understand individual clusters. Acquiring data on galaxy clusters is a long, arduous task and at present relatively few clusters of galaxies have been studied in great detail. Of course, galaxy clusters have undergone a significant amount of dynamical relaxation since their formation and consequently much information about the initial conditions has been wiped out. We must therefore rely on trying to understand their masses, angular momenta, and binding energies (the situation in analagous with galaxies). Unfortunately, the mass determinations are clouded by the issue of the "virial mass discrepancy": if clusters of galaxies are dynamically bound, they must contain almost a factor of 10 more masses than would be attributed to their constituent galaxies. We have no idea at present of what form this hidden mass may take (Tarter and Silk, 1974). The determination of cluster angular momenta requires a large sample of redshifts for member galaxies. The only cluster where a significant upper limit can be placed on the rotation is the Coma cluster (Rood *et al.*, 1972), and there it can only be said that the data are consistent with no rotation. It has been pointed out, however (cf. Sec. V), that if cosmic angular momentum has been generated by tidal torques, large clusters would not be expected to have as high rotation velocities as their smaller, neighboring clusters. Thus it may be profitable to see whether the cluster A1367 is rotating, since in view of its present position relative to Coma these two clusters may have interacted tidally in the past (Jones, 1976 b). There is also evidence for such tides in the Shane–Wirtanen (1967) counts for the region. Similarly, A2197, the smaller companion of A2199, may have detectable rotation.

On the basis of their present mean densities (a highly uncertain quantity in view of the uncertainties in the mass and distance), clusters of galaxies may have formed in the relatively recent past (Noerdlinger, 1970), and some clusters may still be forming. Again, a systematic investigation of the properties of galaxy clusters, like that initiated by Oemler (1973), will undoubtedly yield further clues as to the origins of these systems.

E. Young galaxies

To better understand the process of galaxy formation it would be useful to observe a galaxy in the process of forming. This involves a search for galaxies that may be young in comparison with most others. As we have seen, within the framework of the hot big bang theory, the epoch of galaxy formation may have been a long time in the past. In that case all galaxies observed within a redshift of $z \sim 1$ will be about 10^{10} years old, and looking for "young" galaxies requires that we look at much greater red-shifts. Of course, galaxies may have formed in a manner quite different from that envisaged here and "young" galaxies may be found locally. Whichever is the case, it is necessary to know what to look for: by which criteria are we to decide that an object is significantly less than 10^{10} years old? The two problems of looking for young galaxies at $z \gg 1$, and $z \ll 1$ are quite different and so will be discussed separately. It should be borne in mind that, notwithstanding the large body of opinion that most galaxies formed at redshifts $z > 1$, it is important to see whether there are any young galaxies with $z < 1$. Finding such an object would probably necessitate a serious revision of our "orthodox" ideas concerning the origin of galaxies.

Consider first the problem of identifying "young" galaxies with $z < 1$. Burbidge *et al.* (1963) considered the problem of locating such systems and discussed several criteria which might be indicators of youth. The difficulty encountered here is that a galaxy may have youthful attributes and yet still be 10^{10} years old: there can be no conclusive proof that a particular galaxy is young. Consider, for example, the presence in a galaxy of a large population of young stars. This may be due to a recent burst of star formation, or it may indicate that the stellar birth function is constant, but weighted in favor of high-mass stars. As another example, a galaxy may look irregular either because it is on the verge of collapse towards forming a structured, regular system, or because it has suffered tidal disruption in the relatively recent past. These arguments have been presented by Sargent and Searle (1972), who conclude that the existence of "young" galaxies can only be established by statistical analysis of data on large samples of galaxies.

An important example of such a statistical study has been presented by Searle *et al.* (1973), who computed "evolutionary tracks" for the integrated stellar populations of model galaxies, and plotted them on the $(B-V)$: $(U-B)$ plane. Two such tracks were computed using different assumptions: (a) that the rate of star formation is the same at all times, and (b) that star formation occurred in a burst when the galaxy formed. It is expected that the actual situation is intermediate between these two extremes. In other words, the rate of star formation is a steadily declining function of time. The two computed tracks lie remarkably close together on the $(B-V)$:$(U-B)$ diagram, so comparison with observation does not prefer either hypothesis. However, an important point does emerge: *In a sample of 148 ordinary galaxies chosen for comparison with these tracks, none was younger than* 3×10^{9} *years old, and none was older than* 3×10^{10} *years*. It is clear that to find young galaxies in this way, we must examine classes of galaxies with abnormal colors. There exist a number of galaxies which are anomalously blue (that is, they give the impression that their starlight is dominated by recently formed hot stars). Such objects have been listed by Haro (1956), Markarian (see Ulrich, 1971), and Zwicky (see Sargent, 1970). While these seem to be better candidates for "young" galaxies than ordinary galaxies, it is not as yet possible to rule out the hypothesis that they are galaxies which are (for some reason) undergoing anomalously active periods of star formation. It would be interesting to identify similar objects which are not undergoing this activity, but which went through this hypothetical active phase some time in the past, or have yet to go through such a phase. If the active phase is indeed a rare event in the life of such objects, there should be many more inactive objects in this class than active ones.

Let us now turn to the problem of detecting forming galaxies at great redshifts. It is difficult to detect an ordinary galaxy at a redshift of $z\sim1$, so any hope of looking back to redshifts $z\sim10$ must rest on the hypothesis that galaxies go through an anomously bright phase while they make their first stars. If a galaxy of mass M converts a fraction ΔZ of its mass into heavy elements over a period Δt, its total luminosity would be of the order of

$$\mathcal{L}\sim7\times10^{-3}\,Mc^2(\Delta Z/\Delta t)\ .$$

With $\Delta Z\sim10^{-2}$ and $\Delta t\sim10^8$ years, this yields

$$\mathcal{L}\sim10^{46}\ \text{ergs}\cdot\text{sec}^{-1}\ .$$

This is some two orders of magnitude brighter than the total bolometric luminosity of our Galaxy, but of course this estimate depends sensitively on Δt. Partridge and Peebles (1967a) and Weyman (1967) argued for a value $\Delta t\sim10^7$ years, whereas Δt as large as 10^9 years has been suggested (Salpeter, 1959). Tinsley (1972a, b) has evolved composite stellar populations in time and finds that $\Delta t\sim10^8$ years. The possibility of detecting such a galaxy depends further on the redshift of formation of the galaxy, the kinds of stars that are made during this initial burst, and the volume of the protogalaxy that is bright. The redshift of formation comes in several ways: it affects the integrated brightness of the image, it shifts the peak of the intrinsic radiation spectrum, and it also affects the size of the galaxy image. The effective temperature of the protogalaxy depends on what kinds of stars are formed, and the volume in which the initial burst of star formation takes place determines whether the galaxy will appear starlike or diffuse.

Several models of forming galaxies are now available, on the basis of which we can judge the prospects for finding such objects. The models of Partridge and Peebles (1967a) take star formation to occur throughout a volume 30 kpc in diameter. This assumption leads us to search for diffuse objects. The model adopted by Weymann (1967) assumes a bright volume 3 kpc in radius. This leads to starlike objects. More recently, Meier (1975) has pointed out that on the basis of Larson's (1974) models for the collapse of protogalaxies, one would expect the bright phase to take place at the time of formation of the nucleus of the galaxy. Again this indicates that young galaxies at great redshifts would appear starlike, and the only hope of identifying these objects lies in observing their spectra. Meier (1975) computed a typical spectrum for this kind of young galaxy on the basis of stellar population synthesis (see Tinsley, 1972a).

There have been two searches for diffuse Partridge–Peebles-type young galaxies by Partridge (1974) and Davis and Wilkinson (1974). Neither search revealed any objects that could clearly be labeled as primeval galaxies, though they do serve to put limits on the parameters of the Partridge–Peebles models. Of course, if primeval galaxies look like stars it is not surprising that these searches yielded only negative results; a search for stellar-type primeval galaxies may prove more fruitful. The first candidates one thinks of in this respect are QSOs ("quasi-stellar objects"), though just where QSOs fit into the general scheme of galaxies and clusters of galaxies is not certain. It has been suggested on quite different grounds that epochs corresponding to redshifts $z\sim2$ to 4 may be the time of galaxy formation (Sunyaev, 1971). [The observed counts of radio sources and QSOs suggest this as the period of formation of radio sources and QSOs (see, for example, Longair, 1966, and Doroshkevich *et al.*, 1970), though there is no strong reason to associate galaxy formation with quasar activity (see, for example, Field, 1964, and Lynden-Bell, 1967a, in this connection).] Meier (1975) has pointed out that the spectra of two quasars accord well with his theoretically predicted spectrum for a forming protogalaxy. Notwithstanding the limitations of his procedure (for example, the use of Larson's models for protogalaxy collapse, certain simple assumptions concerning the mass function of the forming stars and the neglect of dust), the agreement is impressive and will undoubtedly lead to further theoretical and observational work in this direction.

IX. OTHER THEORIES OF GALAXY FORMATION

The discussion of the preceding sections has been centered about the by now orthodox point of view that galaxies have grown from initially small deviations in an otherwise homogeneous and isotropic Friedmann–Lemaitre type cosmos. This picture has evolved out of a wish to preserve, as far as possible, the edifice of the hot big bang theory for the evolution of the universe. (As remarked earlier, this theory in its simplest form provides a ready explanation for both the observed 25% (by mass) cosmic abundance of helium, and the observed cosmic radiation field with its 2.7 °K Planckian spectrum.) However, if one abandons the assumption that the universe was less lumpy in the past than now, or if one goes so far as to abandon the underlying cosmological picture, it becomes possible to construct a number of rather different theories for the formation of galaxies. The only real constraint on these theories is that they should be consistent with observational data. Other constraints one may wish to impose, such as requiring that the model should involve only known physics, are merely aesthetic restrictions (though, naturally, one would not wish to go so far as to invent new physics for the sole purpose of explaining the existence of galaxies).

There are a number of models for galaxy formation in which galaxies are formed as a result of material *outflow* from what may be termed a "singularity."[34] In modern times, the first suggestion that such an outflow may play an important part in the formation of galaxies may be attributed to Jeans (1928, p 352):

> "[it is] difficult to resist a suspicion that the spiral nebulae are the seat of types of forces entirely unknown to us, forces which may possibly express novel and unsuspected metric properties of space. The type of conjecture which presents

[34]The use of the term "singularity" here is simply meant to convey the impression of a compact region, the origin of which we do not know. The term is of course more commonly used to describe space-time singularities occurring in the general theory of relativity.

itself, somewhat insistently, is that the centers of the nebulae are of the nature of 'singular points,' at which matter is poured into our universe from some other, and entirely extraneous, spatial dimension, so that, to a denizen of our universe, they appear as points at which matter is being continually created".

Of course, such a statement in itself does not qualify this idea as a theory. It was perhaps Hoyle and Narlikar (Hoyle, 1965; Hoyle and Narlikar, 1966a, 1966b) who did most to quantify this idea within the framework of the steady state theory (see Hoyle, 1965, and references therein), and Ambartsumian (1958; 1961; 1965) who stressed the observational attributes of such a concept.[35] One must beware of the danger of rejecting these theories simply on the grounds that the steady state concept is at present hard to maintain in the face of the observations. Such models could, without difficulty, be incorporated into the Friedmann–Lemaitre-type of cosmology.

It is not necessary to depart from the theory of General Relativity to build outflow models for forming galaxies. Such relativistic models have been developed by Novikov (1964), Ne'eman (1965), Ne'eman and Tauber (1967) (from the point of view of explaining quasar-type phenomena), and Harrison (1970 a, b, c; 1971). In these models, galaxies form from parts of the universe that have not evolved far from the cosmic singularity in comparison with the universe as a whole. They may thus be referred to as "lagging core" or "white hole" models for the origin of galaxies.

Just as with the orthodox point of view, the link between the lagging cores and galaxies as we see them now is a tenuous one, and it is difficult to think of a convincing observational test that might distinguish the two ideas. Evidence for a truly young galaxy would of course tip the balance against the orthodox view. Bahcall and Joss (1972) have examined testable consequences of the outflow hypothesis and conclude that it is consistent with the view that compact emission line systems like I *Zw* 0930 + 55 and II *Zw* 0553 + 03 (Sargent and Searle, 1970; Chamaraux *et al.*, 1970) are in the process of formation. The trouble is that the inflow hypothesis is equally consistent with this view.

It might be difficult to understand the origin of galactic angular momentum on a simple outlfow model like that of Novikov (1964), Ne'eman (1965), or Ne'eman and Tauber (1967). Hoyle's (1953) suggestion, that the tidal torque exerted on the protogalaxy by a neighboring cluster of galaxies might induce spin, cannot work here since these protogalaxies were more compact in the past (see Sec. IV). Harrison's (1971) model avoids this problem by allowing the lagging cores to possess primordial spin. According to this hypothesis there will be a clear distinction between spin-dominated systems, which might be identified as protospirals, and low-spin systems, which might be identified with protoellipticals.

These lagging-core-type models are inherently more complicated than the orthodox models for galaxy formation. There are at least two reasons for this. Firstly, the cores are not small perturbations to Friedmann–Lemaître universes, and consequently perturbation theory cannot be used to obtain solutions to the Einstein field equations. The exact Bondi–Tolman–Lemaître (Bondi, 1947; Cahill and Taub, 1971) solutions are a good starting point for discussions of this kind if one is prepared to restrict one's attention to spherically symmetric dust models; however, no solutions are available for more general situations. Secondly, the cores will accrete material from surrounding regions. The accretion process will lead to a core–halo-type structure, and, as remarked by Hoyle and Narlikar (1966a), this may be a desirable feature of these models. The accreted envelope can be discussed in the same kind of approximation as the adiabatic density perturbation theory (see Sec. III). Thus a $10^{10}\,M_\odot$ core embedded in an otherwise homogeneous and isotropic universe will look like a $\delta\rho/\rho = 0.1$ fluctuation in density on scales of $10^{11}\,M_\odot$, and a $\delta\rho/\rho = 0.01$ fluctuation on scales of $10^{12}\,M_\odot$. In an Einstein–de Sitter universe, the $10^{11}\,M_\odot$ region would start to fall back onto the core at a redshift $z\sim 100$, $z\sim 100$, and the $10^{12}\,M_\odot$ region would separate out at a redshift of $z\sim 10$. *It is difficult to make a clear distinction between this model and the gravitational instability theory*, at least as far as the outer regions of the galaxy are concerned. The lagging core has merely provided the additional, and perhaps desirable, feature of an active nucleus (though the lack of any obvious correlation between the scale of the nuclear regions and the galaxy as a whole may be difficult to understand).

Hoyle and Narlikar (1966a) calculated the expected falloff in density as a function of radius in the accreted envelope. They found, from simple arguments,[36] a law of the form $\rho(r)\propto r^{-8/3}$, which is somewhat shallower than the observations of the outer regions of elliptical galaxies would indicate. [Over much of the outer regions of elliptical galaxies, $\rho(r)\propto r^{-3}$; this corresponds to Hubble's surface brightness distribution (see Sec. VIII.B).] However, in regions of the galaxy where the stellar orbital periods are short compared with the age of the universe, one would expect relaxation process to modify the run of density with radius (since during the collapse of a stellar system the energy of a stellar orbit is not a conserved quantity). In the outer parts of the galaxy where relaxation processes take too long, the assumptions leading to the law $\rho\propto r^{-8/3}$ break down.

[35]It is interesting that Ambartsumian *predicted* the apparent instability of groups of galaxies by drawing an analogy with the expansion of stellar associations. Thus an explosive origin for galaxies was strongly suggested. A considerable amount of observational effort has gone into examining this hypothesis (see for example Burbidge and Sargent, 1970). At present, the problem of the dynamics of these small compact groups of galaxies is still unresolved.

[36]Hoyle and Narlikar (1966a) consider a spherical shell containing mass $M(r)$, surrounding a central core of mass $\mu \ll M$. The initial condition imposed on the motion of the shell is that when its radius is $r_0 = (2GM/c^2)(M/\mu)$, its expansion velocity should be $\dot{r}_0 = (2GM/r_0)^{1/2}$. Under these conditions, the shell expands a further factor M/μ before collapsing back towards the core. The maximum radius is then $r_m = (2GM/c^2)(M/\mu)^2$, so $M\propto r_m^{1/3}$ for each shell making up the protogalaxy. Each shell is assumed to collapse and relax to an equilibrium configuration independently of other shells. The density run in the final galaxy then follows $\rho\propto M/r^3\propto r^{-8/3}$.

The Hoyle–Narlikar calculation cannot be taken straight over to the problem of cores accreting in a Friedmann-type universe where there is a radiation pressure-dominated era; this is because the conditions assumed by Hoyle and Narlikar are inappropriate there. The mass of baryons contained within the horizon at t_{eq} exceeds the mass of a galaxy, hence protogalaxies come within the horizon during the radiation-dominated era. Thereafter, accretion onto the core is inhibited by the radiation pressure until after the universe has recombined. (There is not a period of continuous growth as in the Hoyle–Narlikar model.) Although some growth may take place before the protogalaxy comes within the horizon, the evolution at such epochs is not well understood (see Sec. III. A.4) and the initial conditions might as well be specified at the recombination epoch. For simplicity we can consider the case of a core of mass μ embedded in an otherwise homogeneous and isotropic universe. The density perturbation on scale M is therefore $\delta\rho/\rho = \mu/M$ and scales of mass M separate out from the universe at a redshift $z_M \sim (\mu/M) z_{rec}$. Following the argument of Hoyle and Narlikar, this is seen to lead to a density-radius relationship for the galaxy of $\rho(r) \propto r^{-9/4}$, somewhat shallower than the Hoyle–Narlikar model.[36] However, two effects conspire to modify this relationship. Firstly, relaxation processes will modify this where the stellar crossing times are short compared with the Hubble time. Secondly, in the outer regions of the galaxy, different degrees of relaxation, depending systematically on radius, will be achieved. It can be shown that this latter effect will produce a falloff in density like $\rho(r) \propto r^{-k}$ with $k \simeq 2.8$ in the outer regions.

It seems difficult to rule out such a model by direct comparison with observed systems. However, there are two points that require attention if such models are to be accepted at the same level as, say, the gravitational instability picture (cf. Sec. III). The first point, raised by Peebles (1974d), concerns the origin of large-scale structure. The statistical analysis of the distribution of galaxies on the sky by Peebles and co-workers (cf. Sec. VIII. D) has revealed the existence of structure on scales up to 30 h^{-1} Mpc. How can such structure be explained on the basis of a lagging-core theory? To suggest that the cores about which galaxies have formed are themselves clustered in a nonrandom fashion seems somewhat *ad hoc*. It may be that galaxies are born by some kind of "calving" process in the vicinity of other galaxies (Hoyle, 1965); whether such a process could operate on this kind of scale is not clear. The second problem concerns the observed shape of the spectrum of the cosmic microwave background radiation field: it does not deviate significantly from a Planck spectrum over the wavelength range 0.06 cm to 50 cm (Woody *et al.*, 1975). As shown by Zel'dovich and Sunyaev (1969), injection of energy into the universe, prior to recombination, may result in observable deviations from the Planck law. Conversely, the accuracy with which the observations fit a Planck law constrains the pre-recombination thermal history and so it is not possible to arbitrarily introduce hot bodies (such as lagging cores) into the universe. It should be possible to place restrictions on the size and temperature of the lagging cores since we know that there must be at least as many such cores as there are galaxies. This calculation has not as yet been done.

There are numerous other theories for the origin of galaxies which one may fit into the non-orthodox category. However, it is hoped that the one non-orthodox theory that I have chosen to discuss in some detail, the lagging-core type of theory, is sufficient to demonstrate that non-orthodoxy is not by any means grounds for rejection! To narrow the field of possible explanations for the origin of galaxies will involve a great amount of theoretical work on a large number of ideas, coupled with detailed consideration of observational data. There are two classes of theory which I have not discussed and which the reader may find of interest. First, there are approaches using the statistical mechanics of self-gravitating systems developed by Layzer (1963a, b; 1964; 1968) and by Saslaw (1968; 1969; 1970). The latter series of papers attempts to discuss the formation of galaxies as a phase transition (a condensation phenomenon) in an expanding gas wherein the particles interact gravitationally. Second, there are the Alven–Klein–Omnes cosmogonies in which the universe is considered to have been made up equally of baryons and antibaryons. The annihilation of the baryons and antibaryons produces the 3 °K cosmic radiation field (Omnes, 1969, 1974), leaving behind sufficient material to coalesce into galaxies and antigalaxies. (For a review of this and other baryon-symmetric cosmologies, see Steigman, 1974.) A model for the formation of galaxies in this cosmology has been presented by Stecker and Puget (1972). Conceptually, this problem is a difficult one, but if one accepts the picture as presented by these authors, the observed constraints on the Planckian shape of the cosmic background radiation field are violated (Jones and Steigman, 1976).

X. CONCLUDING REMARKS

Three aspects of galaxy formation theories have been selected for discussion in this review. (1) Current theories for the formation of protogalaxies, (Secs. III-VI). (2) The impact of these theories on our understanding of cosmology (Sec. VII). (3) Their confrontation with observation (Sec. VIII). It is hoped that this article has, among other things, demonstrated the close interrelation between these aspects of the subject.

The discussion has neglected a considerable body of literature on the evolution of protogalaxies towards the presently observed structures. However, the viewpoint adopted here is that it is necessary to understand first how protogalaxies, having appropriate scale and spin, could have formed from the expansion of the cosmos out of its singularity. The advantage of such a restrictive point of view is that the relevant physical processes are relatively well understood, while at this simple level there is nonetheless sufficient contact with the observational data to make the theories more than mere flights of theoretical fancy. (The theories are, in the sense of Karl Popper, falsifiable.) Indeed, the advances of recent years are such that the study of galaxy formation is now in a position to motivate specific observational programs which might eventually provide strong arguments, one way or the other, for various of the present theories.

As promising as this seems, it is nevertheless important to remember that all is still not clear at the theoretical level. There is still a controversy about the origin of vorticity in the gravitational instability picture (Sec. IV), and there are numerous outstanding questions concerning the various aspects of the cosmic turbulence theory (Sec. VI). Furthermore, as stated earlier, there is a considerable gap in our knowledge regarding the collapse of a protogalaxy to form a bound and relaxed system of the kind observed at present. It is on this latter front that future research will undoubtedly proceed, and a firm foundation has already been laid by the models of Larson (1974), Tinsley (1972a, b), and others. Considering the complexity of this phase of galaxy formation which is dominated by the processes of star formation, stellar relaxation, supersonic gas dynamic, and a host of other imaginable but poorly understood phenomena, the success achieved by these people is quite remarkable. A common feature of these investigations tends to be an extensive amount of computing. The development of bigger and faster machines together with improved numerical techniques will undoubtedly add impetus to this direction of research. One should be wary, however, of burying a clear understanding of the underlying physical processes in vast amounts of computer output.

This phase of galaxy evolution is almost in the arena of direct observation; it would, for example, be of considerable value to find a truly young galaxy. There is unfortunately little hope of directly observing the protogalactic phase of galaxy evolution, though several ways have been suggested for putting limits on the amount of primordial inhomogeneity by observing the spectrum of the cosmic microwave background radiation field or the small scale anisotropy of the last scattering surface. While attempts are still being made to establish the degree of isotropy of the radiation field, it is unfortunate that there is little interest in better establishing the detailed shape of the spectrum.

Thus in the ten years since the establishment of the hot big bang theory of the universe, considerable progress has been made towards an understanding of the development of cosmic structure. There are indeed problems to face, but these are by no means intractable. There is a diversity of opinion, but hopefully observations will play a vital role in discriminating amongst the various theories. Notwithstanding this optimism, however, it is perhaps prudent to bear in mind the remark made by Jeans towards the end of his work—"We may well be the most ignorant cosmogonists in the whole of space."

ACKNOWLEDGMENTS

My thanks go particularly to Professor P. J. E. Peebles for numerous discussions over a period of two years on the subject of galaxy formation. The structure of this article is based largely on a discussion we had prior to a meeting at Yale on galaxy formation.

During the past two years I have enjoyed numerous discussions with Kwing Lam Chan, Margaret Geller, Michael Fall, Janet Jones, Susan Lea, Don Olson and Joe Silk. These people have contributed significantly towards my understanding of the subject and to the organization of this article.

I owe thanks to the Berkeley Astronomy Department for their award of their Parisot Post-Doctoral Fellowship, their hospitality, and their help in producing this work.

APPENDIX A: SOME BASIC NOTIONS OF MODERN COSMOLOGY

A. The "big bang" theory

Our modern view of the universe is based mainly on a few important observations. Firstly, there is the deduction that *the nebulae are systems external to our own galaxy* (Hubble 1925 a, b, c, 1926). This together with the accompanying realization that our galaxy is a fairly typical spiral galaxy in a fairly homogeneous distribution of galaxies moves us one step further from believing we are at the center of the universe.

Secondly, there is the striking discovery of Hubble (1929) that *the more distant galaxies are receding from us faster than the nearby ones*. The linearity of the recession velocity—distance relationship was first established by Hubble out to the distance of the Virgo cluster of galaxies (whose distance is currently thought to be about $10h^{-1}$ Mparsec), and has now been extended almost a hundred times deeper into space. "Hubble's Law," as it is appropriately called, surely ranks as one of the outstanding discoveries of modern physics. Taken at face value, it requires no deep arguments to deduce that the universe was denser in the past than now, a result immediately suggestive of the existence of a unique cosmic event in our past. Indeed, the Hubble law, taken with the General Relativity cosmological models of Friedmann and Lemaitre and with conventional assumptions regarding the properties of matter, indicates that this event was a *singularity* a finite time in the past when the density of the universe was infinite.

Thirdly, there is the observation that radio sources were more numerous in the past than now (see Doroshkevich *et al.*, 1970, or Weinberg 1971 for a review of this). The datum points to an *evolutionary cosmology* consistent with the "big bang" cosmology, but which would be difficult to understand on the basis of a "steady state" cosmology (see Hoyle, 1965) which requires that the universe be isotropic not only spatially, but also temporally.

Fourth, there is the discovergy of an *isotropic cosmic radiation field with an apparently Planck spectrum at* $\sim 3\,°K$ by Penzias and Wilson (1965). This radiation is thought to be a relict of a *hot* dense phase in the past history of the universe (Dicke *et al.* 1965), and was predicted by Gamow (1953). The isotropy of the radiation and its Planckian spectrum would be difficult to understand on the basis of a cosmology that did not resemble the "hot big bang" theory over most of the past 10^{10} years of its history (see footnote 1). There are other data which support the "big bang" picture; Peebles' book (Peebles 1971) provides an excellent overview in this respect.

Finally, there is the realization that the oldest stellar systems have ages comparable with the inferred "age" of the universe. The observations just described constitute a fairly compelling chain of reasoning leading up to our present view of the universe.

The force that dominates the dynamics of the universe is gravity, and so Einstein's theory of General Relativity is the appropriate framework within which to construct cosmological models. A homogeneous and isotropic cosmological model is characterized by a single parameter, the *cosmic scale factor*, $R(t)$, which is a function of the cosmic time, t. Physically, $R(t)$ relates the separation of two co-expanding points at different epochs: $d(t)/R(t) = \text{const}$. The Einstein field equations, together with an equation of state, determine $R(t)$. In the case of homogeneous and isotropic model universe, they reduce to two equations which are essentially expressions of the local conservation of momentum and energy

$$3\frac{\ddot{R}}{R} = -4\pi G(\rho + 3p/c^2) ,$$
$$\frac{\dot{R}^2}{R^2} + \frac{k}{R^2} = \tfrac{8}{3}\pi G\rho , \qquad \text{(A1)}$$

where k is a constant measuring the "curvature" of space-time. A universe with $k > 0$ will expand to a maximum value of R and then collapse back upon itself. A $k < 0$ universe will expand forever. The *Einstein–de Sitter universe* is characterized by $k = 0$. For a given equation of state, $p = p(\rho)$, these can be solved to give $R(t)$, $p(t)$, and $\rho(t)$. The "flat" $k = 0$ universes have particularly simple solutions:

$$R(t) \propto t^{2/3}, \quad p = 0 ,$$
$$R(t) \propto t^{1/2}, \quad p = \tfrac{1}{3}\rho c^2 . \qquad \text{(A2)}$$

Conservation of baryons requires that the density of the baryon component, ρ_m, follows

$$\rho_m \propto R^{-3} . \qquad \text{(A3)}$$

(A co-expanding volume of space increases in volume like R^3 and contains a fixed number of particles.)

It should be noted that in the second of equations (A1), the k/R^2 term decreases like R^{-2}, while the $\frac{8}{3}\pi G\rho$ term decreases like R^{-3} (in the case of a universe containing matter rather than radiation). At sufficiently early epochs, the latter term dominates and the expansion of the universe is well approximated by the $k = 0$ solution: $R(t) \propto t^{2/3}$. In other words, the early expansion of a $k \neq 0$ universe looks rather like an Einstein–de Sitter universe. At later epochs, the k/R^2 term dominates and the dynamical behavior for $k < 0$ is approximately described by $R(t) \propto t$ (this is the $\rho = 0$ solution). This corresponds to the phase of *undecelerated expansion*. The transition from $k = 0$ like behavior to undecelerated expansion plays an important role in galaxy formation theories, and the epoch at which this happens can be estimated as follows. If at present $|k/R^2| \gg \frac{8}{3}\pi G\rho_0$, then $|k/R^2|$ will be approximately $(\dot{R}/R)^2$. This latter quantity is the square of the present cosmic expansion rate and can be estimated on the basis of the observed expansion of the system of galaxies. $\dot{R}/R$ is in fact the Hubble constant appearing in the recession velocity: distance relationship, $v = Hr$. Thus the present ratio of the terms $|k/R^2|$ and $\frac{8}{3}\pi G\rho_0$ is approximately $8\pi G\rho_0/3H^2$, a dimensionless quantity given the symbol Ω. This ratio increases in proportion to R^{-1} as we go into the past and attains unity when $R_1/R_0 = 8\pi G\rho_0/3H^2$. As will be shown later, this corresponds to a "redshift"

$$1 + z_1 = 3H^2/8\pi G\rho_0 = \Omega^{-1} .$$

Since $R \propto t$ in the undecelerated regime, we can say that this transition took place at a time $t_1 \doteqdot \Omega t_0$.

Consider now a model universe containing only radiation. It can be shown that all photons are reduced in frequency ("redshifted") as the universe expands, and that the radiation temperature falls off as R^{-1}

$$\nu \propto R^{-1} ,$$
$$T_r \propto R^{-1} . \qquad \text{(A4)}$$

The decrease in frequency can be viewed as a stretching of the photons wavelength by the cosmic expansion: $\lambda \propto R$. The *redshift* factor of a photon, z, is defined as the relative frequency change between the time of its being emitted and the time it is received. Thus

$$z = \frac{\nu_0 - \nu_1}{\nu_1} = \frac{R_1}{R_0} - 1 , \qquad \text{(A5)}$$

where R_1 and R_0 are the scale factors at the epochs t_1 and t_0 when the photon is received and when it is emitted. In cosmology, one is often concerned with such early epochs that $z \gg 1$, and then $z \propto R^{-1}$.

Consider a model universe containing only matter. In the absence of heat sources, the temperature will fall off as

$$T_m \propto R^{-2} . \qquad \text{(A6)}$$

This corresponds to the adiabatic expansion of a gas whose thermal velocities fall off as R^{-1} and whose density falls off as R^{-3}. In a universe containing both matter and radiation, the thermal history depends on the coupling between the components and on the ratio of their specific heats. Observations of the present cosmic radiation spectrum indicate a temperature near to 3 °K for the radiation component, and the ratio of the specific heats of the two components is

$$\sigma = \frac{4aT_r{}^3 m_N}{3k} = 1.3 \times 10^8 (\Omega h^2)^{-1} .$$

The heat capacity of the radiation vastly outweighs that of the matter, and so as the universe expands $T_r \propto R^{-1}$ (the presence of the matter hardly affects the cooling even if there is close thermal coupling between the matter and radiation). The magnitude of σ also implies that when the matter and radiation are coupled thermally, the matter and radiation temperatures are equal and $T_m \propto R^{-1}$.

At sufficiently early epochs, then, the universe was, according to this view, hot and dense. The cosmic matter would have been completely ionized (photoionizations would have outweighed recombinations) and thermal contact between the matter and radiation would be maintained via Compton scattering of photons off the free electrons. However, there comes a time during the expansion when the photoionization rate fails to compete with the recombinations and the cosmic matter becomes neutral. The process of recombination is quite rapid (taking only 20 per cent or so of the cosmic expansion time) and thus one speaks of the *epoch of recombination*. The disappearance of the electrons at this time means that the universe becomes optically thin to Compton scattering and thermal contact between the matter and radiation is lost. The matter temperature then falls off as $T_m \propto R^{-2}$ until heat sources such as galaxies or quasars appear. Such a reheating of the cosmic matter may cause it to become ionized once again, thereby re-establishing thermal coupling between the two phases.

The epoch at which the cosmic background radiation photons were last scattered defines the *surface of last scattering*. This is the earliest epoch about which we can derive information from the angular variations in the radiation temperature arising from hypothetical inhomogeneities in the distribution of matter. If there has been no reheating of the matter since recombination, then we can "look" back to the epoch of recombination itself. In most cosmological models that are consistent with observation, this corresponds to a redshift

$$z_{\rm rec} = 1500 \ .$$

If there has been a reheating of the matter which has not condensed into galaxies, it may only be possible to look back to redshifts of two or so with the cosmic background radiation. (This statement depends sensitively on the density of the left-over material and the details of the thermal history of the heat sources. Quasars have been observed with redshifts in excess of three, so we know that the universe is optically thin at least that far out.)

With the result that $T_r \propto R^{-1}$, we see that the energy density of the radiation field has evolved as

$$\rho_r = \frac{aT_r^4}{c^2} \propto R^{-4} \ .$$

So even though at present $\rho_m \gg \rho_r$, there was a period in the past when $\rho_r > \rho_m$. This is referred to as the *radiation-dominated era*. The epoch when $\rho_r = \rho_m$ corresponds to a redshift

$$z_{\rm eq} \doteqdot 4\times 10^4 (\Omega h^2) \ .$$

For most values of Ωh^2 consistent with observation, $z_{\rm eq} \geq z_{\rm rec}$.

B. Some units and symbols

The unit of distance most frequently used in cosmology is the Megaparsec (Mparsec):

$$1 \text{ Mparsec} = 3.086\times 10^{22} \text{ m} \ .$$

Velocities of galaxies are usually quoted in km·sec^{-1}. There is an observed relationship between the distances of galaxies and their recession velocities:

$$\left(\frac{v}{1\text{ks}^{-1}}\right) = \left(\frac{H_0}{1\text{ks}^{-1}\text{ Mparsec}^{-1}}\right)\left(\frac{r}{1\text{ Mparsec}}\right) \ . \qquad \text{(A7)}$$

This law describes the phenomenon of the expansion of the universe. H_0 is called "Hubble's Constant". Throughout this *review* I follow the convention adopted by Peebles (1971a) and use a dimensionless parameter h defined as the value of H_0 in units of 100 ks^{-1} Mparsec^{-1}.

$$h = H_0/100\text{ ks}^{-1}\text{ Mparsec}^{-1} \ . \qquad \text{(A8)}$$

The use of h is becoming fairly widespread and it is hoped that everyone will use the same normalization. (It should be noted that H_0 is the *present* value of a parameter H which measures the rate of expansion of the universe at different cosmic epochs. H is a function of cosmic time in the evolutionary cosmological models considered here.)

A so-called Einstein–de Sitter universe has the property that the present density of matter in the universe is related to the present value of the Hubble parameter, H_0, by

$$\rho_c = \frac{3H_0^2}{8\pi G} = 2\times 10^{-29}\text{ h}^2\text{ g}\cdot\text{cm}^{-3} \ . \qquad \text{(A9)}$$

(cgs units have been used in quoting densities in almost all of the literature cited in this *review*, and have accordingly been used here. In future it seems probable that this practice will be abandoned in favor of *S.I.* units, in which system $\rho_c = 2\times 10^{-26}\text{ h}^2\text{kgm}^{-3}$.) The actual density of the universe, ρ_0, may differ from ρ_c and it is customary to introduce a *density parameter*, Ω such that

$$\rho_0 = \Omega\rho_c = 2\times 10^{-29} (\Omega h^2)\text{ g}\cdot\text{cm}^{-3} \ . \qquad \text{(A10)}$$

Here Ω is time dependent and in a model universe with zero cosmological constant is twice the *deceleration parameter*, $q_0 = -(1 + \dot{H}/H^2)$. The amount of visible matter in the universe corresponds to a density $\Omega = 0.014$ (Shapiro 1971). Unless there are vast amounts of unseen material, the universe is "open" in the sense that there is not enough matter to gravitationally halt the expansion.

The system used by astronomers for describing the apparent brightness of an object is the somewhat archaic *magnitude scale*. A total flux ℓ erg·cm^{-2}·s^{-1} received from an object is ascribed a *bolometric magnitude* $m_{\rm bol}$ given by

$$m_{\rm bol} = -2.5\log_{10}\left(\frac{\ell}{2.52\times 10^{-5}\text{ erg cm}^{-2}\text{s}^{-1}}\right) \ . \qquad \text{(A11)}$$

(For a detector with a finite bandwidth, the magnitude in that band is given by a similar equation using the detected flux. In the optical region special bands called the U, B, and V band have been selected to define a UBV magnitude system.) The *absolute magnitude*, M, of an object is defined as the magnitude the object would have if placed at a distance of 10pc. Thus an object of apparent magnitude m at a distance d has absolute magnitude

$$M = m - 5\log_{10}\left(\frac{d}{10\text{ parsec}}\right) \ . \qquad \text{(A12)}$$

The *intrinsic luminosity* of the object is then

$$\mathfrak{L}/\mathfrak{L}_\odot = 10^{-0.4(M - M_\odot)} , \qquad \text{(A13)}$$

where $\mathfrak{L}_\odot$ is the luminosity of the Sun (in the relevant waveband), and $M_\odot$ is the Sun's absolute magnitude.

An important feature of the development of astronomy has been the cataloging and classifying of astronomical objects: we recognize stars, clusters of stars, galaxies of stars and clusters of galaxies as well as nebulae and other forms of diffuse matter. The detailed "taxonomy" of these objects has led to a considerable understanding of the physics of these objects and to the determination of cosmic distances. The modern *cosmic distance ladder* by means of which we estimate the distances of galaxies has been summarized by Weinberg (1971). Once the distance to a galaxy is known, its mass can be estimated if we can measure its rotation speed and assuming that the system is in dynamical equilibrium. In units of solar masses and solar luminosities, typical galaxies are thought to have mass and luminosity of the order

$$M=10^{11}M_{\odot}=2\times10^{44}\ \mathrm{g}$$

$$\mathcal{L}=10^{10}\mathcal{L}_{\odot}=4\times10^{33}\ \mathrm{erg\,s^{-1}}\ .$$

Clusters of galaxies have populations ranging from "groups" of a few galaxies up to systems having over a thousand members.

APPENDIX B: HYDRODYNAMICS

A. Description of the flow

Here the discussion follows the presentation of Batchelor (1970). The flow of a fluid is characterized by by specifying the velocity field $\vec{u}(\vec{x},t)$. The local state of the fluid is characterized by its density ρ, pressure p, ionization x_e, temperature T, and other such state variables. Locally, the velocity field can be resolved into components having different physical interpretations. In some small neighborhood of a point (which is chosen as the origin) the velocity field can be written

$$u_i(\vec{r},t)=u_i(0,t)+r_k\left.\frac{\partial u_i}{\partial x_k}\right|_0\ . \tag{B1}$$

The tensor $\partial u_i/\partial x_k$ can be split into its symmetric and skew-symmetric parts

$$e_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)$$
$$\zeta_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}-\frac{\partial u_j}{\partial x_i}\right)\ . \tag{B2}$$

With each part we can identify velocity field components $\delta u_i^{(s)}$ and $\delta u_i^{(a)}$

$$\delta u_j^{(s)}=r_i e_{ij}\ ,\quad \delta u_j^{(a)}=r_i\zeta_{ij}. \tag{B3}$$

Here $\delta\vec{u}^{(s)}$ represents pure straining motion: a spherical volume of fluid at a point $\vec{x}$ is deformed into an ellipsoid with axes changing at rates determined by the three eigenvalues of e_{ij}. The axes of the ellipsoid are in fixed directions and in an incompressible flow the volume of the ellipsoid is constant.

The velocity field $\delta\vec{u}^{(a)}$ is related to the local *vorticity*, $\vec{\omega}$ by

$$\delta\vec{u}^{(a)}=\tfrac{1}{2}\vec{\omega}\wedge\vec{r}\ , \tag{B4}$$

where

$$\vec{\omega}=\nabla\wedge\vec{u}\ , \tag{B5}$$

Here (B4) shows that the velocity field $\delta\vec{u}^{(a)}$ is that associated with a solid body spinning with angular velocity $\frac{1}{2}\vec{\omega}$. The angular momentum of a *spherical* element at $\vec{x}$ is

$$\vec{H}=\tfrac{1}{2}I\vec{\omega}\ ,$$

where I is the moment of inertia of the element about an axis through the center. This accords with the local description of the flow as having a component of solid body rotation with angular velocity $\frac{1}{2}\vec{\omega}$. Although this expression for $\vec{H}$ does not involve the $\delta\vec{u}^{(s)}$ components, that is only a result of the assumption of a spherical volume: $\delta\vec{u}^{(s)}$ would contribute to the angular momentum of a nonspherical elemental volume.

B. The Navier Stokes equations

The dynamics of the fluid flow are governed by the so-called Navier-Stokes equations, which are discussed at length in all standard hydrodynamics texts. The local acceleration of the fluid is determined by the equation

$$\rho\left[\frac{\partial\vec{u}}{\partial t}+(\vec{u}\cdot\nabla)\vec{u}\right]=\rho\vec{F}-\nabla\left[p-\mu\left\{\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}-\frac{2}{3}\delta_{ij}\frac{\partial u_k}{\partial x_k}\right\}\right], \tag{B6}$$

where μ denotes the fluid viscosity, and $\vec{F}$ represents the external force field in which the fluid moves. Notice that the expression in curly braces involving the spatial derivatives of $\vec{u}$ is just the trace-free part of tensor e_{ij}. The operator

$$\frac{D}{Dt}\equiv\frac{\partial}{\partial t}+\vec{u}\cdot\nabla$$

is the *time derivative following the motion* of a particular fluid element. (It is also referred to as the *convective derivative*).

The conservation of fluid is expressed by the so-called *equation of continuity*:

$$\frac{\partial\rho}{\partial t}+(\vec{u}\cdot\nabla)\rho+\rho\nabla\cdot\vec{u}=0\ . \tag{B7}$$

An *incompressible flow* is one in which the density of a given fluid element remains constant in time: $D\rho/Dt=0$. This implies $\nabla\cdot\vec{u}=0$.

In addition to these equations there is an equation describing the exchange between the internal energy of the fluid and other forms of energy

$$T\frac{DS}{Dt}=c_p\frac{DT}{Dt}-\beta\frac{T}{\rho}\frac{Dp}{Dt}$$
$$=\frac{2\mu}{\rho}\left[e_{ij}e_{ij}-\frac{1}{3}\left(\frac{\partial u_k}{\partial x_k}\right)^2\right]+\frac{1}{\rho}\frac{\partial}{\partial x_i}\left(k\frac{\partial T}{\partial x_i}\right). \tag{B8}$$

Here, S is the entropy of the fluid, c_p is the specific heat at constant pressure, β is the coefficient of thermal expansion, and k is the coefficient of thermal conductivity. Other terms may appear on the right hand side describing radiative diffusion or other phenomena. To complete the set of equations, an equation of state relating p, ρ, and T is needed.

C. Kelvin's circulation theorem

The circulation around a closed *material* curve, Γ, in the fluid is defined as

$$C(t)=\oint_\Gamma\vec{u}\cdot d\vec{l}\ . \tag{B9}$$

Physically, C can be regarded as an average tangential component of velocity around the curve Γ. The rate of change of C is governed both by the change of $\vec{u}$ with time and the change in the tangential elements $d\vec{l}$ as the curve Γ is carried around with the fluid

$$\frac{dC}{dt}=\oint_\Gamma\left(\frac{D\vec{u}}{Dt}\right)\cdot d\vec{l}+\oint_\Gamma\vec{u}\cdot(d\vec{l}\cdot\nabla\vec{u})\ .$$

The second integral arises from the fact that the rate of change of an elemental vector $\vec{\delta l}$ representing a material line element is the difference between the velocities at the two ends of the element, that is

$$\frac{d}{dt}(\vec{\delta l}) = (\vec{\delta l}\cdot\nabla)\vec{u} .$$

The equation of motion can be used to manipulate the first integral, while the integrand of the second integral is just $\frac{1}{2}\vec{dl}\cdot\nabla u^2$. Thus

$$\frac{dC}{dt} = \oint_\Gamma \left[\vec{F} - \frac{1}{\rho}\nabla p + \frac{1}{\rho}\nabla\left\{\mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3}\delta_{ij}\frac{\partial u_k}{\partial x_k}\right)\right\}\right]\cdot\vec{dl} . \tag{B10}$$

In a viscous flow, this will not necessarily be zero. However, in the limit of zero viscosity

$$\frac{dC}{dt} = \oint_\Gamma \left(\vec{F} - \frac{1}{\rho}\nabla p\right)\cdot\vec{dl} . \tag{B11}$$

If the external forces are derivable from a potential such that $\vec{F} = \nabla\Psi$, and if the flow is *homentropic* (the density is a function of pressure alone and $(1/\rho)\nabla p = \nabla\int(dp/\rho)$, then

$$\frac{dC}{dt} = \oint_\Gamma \nabla\left(\Psi - \int\frac{dp}{\rho}\right)\cdot\vec{dl} = 0 . \tag{B12}$$

The circulation around Γ is conserved. This is the *Kelvin Circulation Theorem*. If the flow is not homentropic (which might occur if there were a heat source) then the circulation would not be conserved.

It should be noticed that the shear component of the flow field does not contribute to the circulation since that component is, by definition, curl free. Thus the Kelvin Circulation theorem does not imply the conservation of angular momentum for a closed loop of fluid.

In the astrophysical problem of tidal transfer of angular momentum between neighboring protogalaxies, the circulation is the physical quantity of interest. The local vorticity of a fluid element remains zero if it is initially zero, under the same conditions which lead to the conservation of circulation (Batchelor, 1970, p. 277). However, these theorems say nothing about the conservation of angular momentum for fluid elements, or closed material curves.

D. Turbulence

Everyone has a naive impression of "turbulent flow" as a set of random motions of a fluid which can be viewed as a collection of forming and dissolving eddies on a great variety of lengthscales. Such a flow can best be described in statistical terms. The character of the velocity field, for example, can be described at the simplest level by giving the mean square velocity in various directions. The velocity autocorrelation functions provide more detail. The cross-correlations between the pressure and velocity variations, or any other physical properties of the flow, provide a good first-order description which can be compared with experiment. The dynamics of turbulent flow may then be described in terms of the time dependence of these various stochastic functions.

Considerable simplification results from restricting attention to incompressible turbulent flow. The flow parameters can be split up into their spatial averages plus a fluctuating (random) component having zero mean:

$$\begin{aligned} u_i &= \mathring{u}_i + u_i' , \quad \mathring{u}_i = \overline{u}_i , \overline{u_i'} = 0 \\ p_i &= \mathring{p}_i + p' , \quad \mathring{p} = \overline{p}, \overline{p'} = 0 . \end{aligned} \tag{B13}$$

The fluctuating velocity autocorrelation function plays an important role in the stochastic theory of homogeneous and isotropic turbulence. The "longitudinal" and "lateral" velocity autocorrelation functions

$$\begin{aligned} f(r) &= \frac{\overline{u_p(\vec{x})u_p(\vec{x}+\vec{r})}}{\overline{u_p^2}} , \\ g(r) &= \frac{\overline{u_n(\vec{x})u_n(\vec{x}+\vec{r})}}{\overline{u_n^2}} , \end{aligned} \tag{B14}$$

are particularly convenient to consider. Here the averages are taken over all positions $\vec{x}$; the assumption of statistical homogeneity and isotropy implies that f and g are functions of the magnitude of $\vec{r}$. In these equations, u_p denotes the component of the velocity parallel to $\vec{r}$, and u_n is the component of velocity perpendicular to $\vec{r}$. The functions f and g are not independent [see Eq. (B19)].

It is possible to define several characteristic lengthscales in terms of the function f. The *integral lengthscale*

$$L_p = \int_0^\infty f(r)\,dr \tag{B15}$$

characterizes the extent of a region over which velocities are appreciably correlated. Another lengthscale can be defined by noting that for small r, f must have a power series expansion of the form

$$f(r) = 1 - (r^2/2\lambda^2) + 0(r^4) . \tag{B16}$$

[The absence of a linear term is because $f(r) = f(-r)$.] The factor 2 in the denominator of the r^2 term is a matter of convention. Here λ is called the *Taylor microscale*. (Mathematically, λ is the radius of curvature of the correlation function at $r = 0$.)

The general two-point velocity correlation function

$$R_{ij}(\vec{r}) = \overline{u_i(\vec{x})u_j(\vec{x}+\vec{r})} \tag{B17}$$

is related to f and g by

$$R_{ij}(\vec{r}) = \frac{1}{3}\overline{u_i^2}\left[\frac{f-g}{r^2}r_i r_j + g\delta_{ij}\right] \tag{B18}$$

where f and g themselves are related by the equation

$$g = f + \frac{1}{2}r\frac{df}{dr} \tag{B19}$$

(see Batchelor, 1970, p. 46).

There is an interesting interpretation of the lengthscale λ in terms of the two-point vorticity correlation function. It can be shown without much difficulty that if $\vec{\omega} = \nabla_\wedge\vec{u}$, and if the flow is incompressible, then

$$\overline{\omega_i(\vec{x})\omega_i(\vec{x}+\vec{r})} = -\nabla^2 R_{ij}(r) .$$

The mean square vorticity is therefore

$$\overline{\omega_i^2} = [-\nabla^2 R_{ij}(r)]_{r=0} = \frac{15\overline{u^2}}{\lambda^2} . \tag{B20}$$

Thus λ is a lengthscale defined from the ratio of the mean square velocity fluctuation to the mean square vorticity fluctuation.

Some simple arguments can be made showing some qualitative features of turbulent flow. Suppose that, in a turbulent flow, one observes velocity fluctuations of amplitude u_l in a region a size l. (A more precise definition of u_l can be given in terms of the velocity correlation function, but this is not necessary in the present context.) If l is sufficiently large that viscous dissipation is not an important factor, then the characteristic lifetime of the velocity fluctuations is $\tau_l \sim l/u_l$. The kinetic energy is therefore transferred out of this scale at a rate $\epsilon_l \sim u_l^2 u_l / l$. This energy is used to generate motion on smaller scales. If the "energy spectrum" u_l^2 is to preserve its shape, then all scales must transfer their kinetic energy at the same rate, and then $\epsilon_l = \epsilon$, a constant. Thus the requirement that the turbulence energy spectrum be shape preserving at all times leads to the *Kolmogorov Law*

$$u_l \sim (\epsilon l)^{1/3} . \tag{B21}$$

The quantity ϵ characterizes the strength of the turbulence, and may be defined in terms of the motion of the largest eddies as $\epsilon \sim u_{L_0}^3 / L_0$. [Note that there is no *a priori* justification for taking $L_0 \sim L_p$; in terms of the correlation function f the largest eddies are those for which $f(L_0) = 0$.]

Another lengthscale can be constructed from ϵ and the kinematic viscosity ν:

$$\lambda_0 \sim \left(\frac{\nu^3}{\epsilon}\right)^{1/4} . \tag{B22}$$

We can interpret this scale physically by noting that the rate of viscous dissipation from a scale l is $\nu u_l^2 / l^2$. Equating this to ϵ defines a particular lengthscale λ_0 which is roughly the scale of motion at which friction dissipates the spectral energy flux ϵ. Thus the Kolmogorov Law holds over a range of scales

$$u_l \propto l^{1/3}, \quad L_0 \gg l \gg \lambda_0 \tag{B23}$$

This range is called the *inertial range*.

To go beyond this phenomenological description of the flow is difficult and involves writing the equations of motion for the fluctuating component of the flow and the various stochastic descriptors of the flow. Since further advances in the theory of cosmic turbulence will require something more than qualitative arguments it is appropriate to summarize here some aspects of the more detailed theory of turbulence that may pertain to the evolution of cosmic turbulence. The following discussion will therefore concentrate on the mechanism of turbulent energy transfer.

If the flow is resolved into its mean and fluctuating components, the fluctuating component of the Navier-Stokes equations can be extracted

$$\left[\frac{\partial u_i'}{dt} + (\mathring{u}_k + u_k') \frac{\partial u_i'}{\partial x_k}\right] + u_k' \frac{\partial \mathring{u}_i}{\partial x_k} = -\frac{1}{\rho}\frac{\partial p'}{\partial x_i} + \frac{1}{\rho}\frac{\partial}{\partial x_k}(\sigma_{ki}' - \tau_{ik}) , \tag{B24}$$

where

$$\sigma_{ki}' = \rho\nu \left(\frac{\partial u_k'}{\partial x_i} + \frac{\partial u_i'}{\partial x_k}\right), \quad \nu = \mu/\rho ,$$

and

$$\tau_{ik} = -\rho \overline{u_i' u_k'} . \tag{B25}$$

The term in square brackets will be recognized as the derivative of u_i' following the random fluid motion. σ_{ki}' is the fluctuating part of the stress tensor defined earlier, multiplied by the viscosity (this is referred to as the "viscous stress"). τ_{ik} is the important *Reynolds Stress* term.

A comparison between the viscous and Reynolds stresses is important since, if the Reynolds stresses dominate, the fluid viscosity can be ignored from the point of view of u_i'. If we tentatively assign a value u to the characteristic value of u_i', and L to a lengthscale over which u_i' varies, the ratio

$$\mathrm{Re} = uL/\nu \tag{B26}$$

is a dimensionless number reflecting the relative importance of inertial forces to friction forces. This number is called the *Reynolds Number*. In pipe flows, it is found that laminar flows with Reynolds numbers somewhat less than 10^3 are stable, whereas for $\mathrm{Re} \gg 10^3$, they are unstable to the generation of turbulence. (Laminar flows can exist at extremely high Reynolds number, but these are highly unstable.)

The *turbulent energy equation* can be derived by contracting (B24) on u_i' and averaging:

$$\rho \frac{dE'}{dt} = -\tau_{ik} \frac{\partial \mathring{u}_i}{\partial x_k} - \frac{1}{2}\mu \overline{\left(\frac{\partial u_i'}{\partial x_k} + \frac{\partial u_k'}{\partial x_i}\right)^2} + \frac{\partial}{\partial x_i}\left[\overline{p' u_i'} + \overline{u_k' \sigma_{ki}'} - \tfrac{1}{2}\rho \overline{u_i' u_k u_k}\right], \tag{B27}$$

where

$$E' = \tfrac{1}{2}(\overline{u_1^2} + \overline{u_2^2} + \overline{u_3^2}) = \tfrac{1}{2}\overline{u_i^2} \tag{B28}$$

is the mean kinetic energy per unit mass associated with the fluctuating motions (the *turbulent pressure*). The first term on the right hand side can be interpreted as the conversion of kinetic energy from the mean flow into kinetic energy of random motions. This is an important aspect of the Reynolds stresses. The second term on the right describes the transformation of the kinetic energy of the fluctuating motion into heat by the action of viscosity. The third term can be shown to be negligible in comparison with the others; it contributes little to the total energy balance of a finite volume of fluid as it is purely a surface-term. This equation enables us to introduce an *eddy viscosity* μ_{turb} describing the effect of removal of energy from the mean flow by the turbulence. The stress tensor for the mean motion is

$$\Phi_{ik} = \frac{\partial \mathring{u}_i}{\partial x_k} + \frac{\partial \mathring{u}_k}{\partial x_i} , \quad \Phi_{kk} = 0 . \tag{B29}$$

If we make an *ansatz* that the Reynolds stresses are proportional to Φ_{ik} then we are led to write

$$\overline{u_i' u_k'} = \tfrac{2}{3} E' \delta_{ik} - \mu_{\mathrm{turb}} \Phi_{ik} . \tag{B30}$$

(The term proportional to δ_{ik} is determined by the definition of E' and $\Phi_{kk} = 0$). The energy equation then reads

$$\frac{dE'}{dt} \simeq \frac{1}{2}\,\mu_{\text{turb}}\left(\frac{\partial \mathring{u}_i}{\partial x_k}+\frac{\partial \mathring{u}_k}{\partial x_i}\right)^2-\frac{1}{2}\,\mu\overline{\left(\frac{\partial u_i'}{\partial x_k}+\frac{\partial u_k'}{\partial x_i}\right)^2}\,, \qquad \text{(B31)}$$

which shows the appropriateness of interpreting μ_{turb} as an eddy viscosity. Of course, this does not tell us how to determine μ_{turb}. Note that under stationary conditions, where $dE'/dt=0$, we have

$$\epsilon \equiv \frac{1}{2}\,\mu\overline{\left(\frac{\partial u_i'}{\partial x_k}+\frac{\partial u_k'}{\partial x_i}\right)^2}=-\,\overline{u_i' u_k'}\,\frac{\partial \mathring{u}_i}{\partial x_k}\,. \qquad \text{(B32)}$$

This is the rate of viscous dissipation of energy, and also the rate at which turbulence is fed from the mean flow. It should be noted that the term $\tau_{ik}\,\partial\mathring{u}_i/\partial x_k$ appearing in equation (B32) need not always be negative; contrary situations are encountered, for example, when the turbulence is driven by an external source of energy (like a temperature gradient) and results in the turbulence driving the mean flow.

REFERENCES

Abell, G. O., 1958, Astrophys. J. Suppl. Ser. **3**, 211.
Abell, G. O., 1965, Annu. Rev. Astron. Astrophys. **3**, 1.
Aitken, R. G. 1906, Publ. Astron. Soc. Pac. **18**, 111.
Aizu, K., S. Hayakawa, H. Sato, and M. Taketani, 1971, Prog. Theor. Phys. Suppl. No. 49, pp. 1–10.
Alexander, S., 1852, Astron. J. 2, 95.
Alpher, R. A., G. Gamow, and R. Herman, 1967, Proc. Natl. Acad. Sci. USA **58**, 2179.
Ambartsumian, V. A., 1958, 11th Solvay Conference, "Structure and Evolution of Galaxies," (Stoops, Brussels), p. 241.
Ambartsumian, V. A., 1961, Astron. J. **66**, 536.
Ambartsumian, V. A., 1965, 13th Solvay Conference, "Structure and Evolution of Galaxies," (Interscience, London), p. 1.
Aristotle, c. 340 B. C., *De Coelo et Mundo*.
Aristotle, c. 340 B. C., *Naturalis Auscultatio*.
Arons, J., and J. Silk, 1968, Mon. Not. R. Astron. Soc. **140**, 331.
Arp, H., 1966, Astrophys. J. Suppl. Ser. **14**, 1.
Bahcall, N., 1974, Astrophys. J. **187**, 439.
Bahcall, J. N., and P. C. Joss, 1972, Comments Astrophys. Space Phys. **4**, 95.
Balko, A. V., 1971, Sov. Phys.–JETP **33**, 237.
Bardeen, J., 1968, Astrophys. J. **73**, S164.
Batchelor, G. K., 1951, Proc. Camb. Phil. Soc. **47**, 359.
Batchelor, G. K., 1970a, *Homogeneous Turbulence* (Cambridge University, Cambridge).
Batchelor, G. K., 1970b, *An Introduction to Fluid Dynamics* (Cambridge University, Cambridge), Sec. 5.3.
Belinski, V. A., I. M. Khalatnikov, and E. M. Lifschitz, 1970, Adv. Phys. **19**, 525.
Belton, M. J. S., and J. C. Brandt, 1963, Ann. Astrophys. **26**, 229.
Binney, J., 1974, Mon. Not. R. Astron. Soc. **168**, 73.
Bogart, R. S., and R. V. Wagoner, 1973, Astrophys. J. **181**, 609.
Bok, B. J., 1934, Harvard Obs. Bull. No. 895.
Bondi, H., 1947, Mon. Not. R. Astron. Soc. **107**, 410.
Bonnor, W. B., 1956, Z. Astrophys. **39**, 143.
Boyanovitch, D., and J. R. M. Radok, 1964, Theoretical Hydromechanics (Wiley-Interscience, New York).
Brosche, P., 1970, Astron. Astrophys. **6**, 240.
Brosche, P., 1973, Astron. Astrophys. **23**, 259.
Brown, F. G., 1964, Mon. Not. R. Astron. Soc. **127**, 517.
Brown, F. G., 1968, Mon. Not. R. Astron. Soc. **138**, 527.
Burbidge, E. M., G. R. Burbidge, and F. Hoyle, 1963, Astrophys. J. **138**, 873.
Burbidge, E. M., and W. L. W. Sargent, 1970, in *Nuclei of Galaxies*, edited by D. J. K. O'Connell (North-Holland, Amsterdam), p. 351.
Burbidge, G. R., 1971, Nature **233**, 36.
Burbidge, G. R., 1976, "Masses of Galaxies" in *Stars and Stellar Systems*, edited by A. R. Sandage and M. Sandage (University of Chicago, Chicago), Vol. IX.
Cahill, M., and A. Taub, 1971, Commun. Math. Phys. **21**, 1.
Chamaraux, P., J. Heidmann, and R. Lanqué, 1970, Astron. Astrophys. **8**, 424.
Chamberlain, T. A., 1900, J. Geol. **8**, 58.
Chamberlain, T. A., 1901, Astrophys. J. **14**, 17.
Chan, K. L., and B. J. T. Jones, 1976, in preparation.
Chandrasekhar, S., 1949, Astrophys. J. **110**, 329.
Chernin, A. D., 1969, Astrofizika **5**, 656.
Chernin, A. D., 1970, Nature **226**, 440.
Chernin, A. D., 1971, Astrophys. Lett. **8**, 31.
Chernin, A. D., and E. D. Eidel'man, 1969, Astrofizika **5**, 654.
Chibisov, G. V., 1972a, Sov. Astron.-AJ **16**, 56.
Chibisov, G. V., 1972b, Sov. Astron.-AJ **16**, 235.
Chitre, P. M., 1972, Phys. Rev. D **6**, 3390.
Collins, C. B., and S. W. Hawking, 1973, Astrophys. J. **180**, 317.
Collins, C. B., and J. M. Stewart, 1971, Mon. Not. R. Astron. Soc. **153**, 419.
Crighton, D. G., 1969, Proc. Camb. Phil. Soc. **65**, 557.
Curtis, H. D., 1919, J. Wash. Acad. Sci. **9**, 217.
Dallaporta, N., and F. Lucchin, 1972a, Astron. Astrophys. **19**, 123.
Dallaporta, N., and F. Lucchin, 1972b, Astron. Astrophys. **26**, 325.
Darwin, G. H., 1908, *Scientific Papers* (Cambridge University, Cambridge), Vol. II.
Davis, M., and D. T. Wilkinson, 1974, Astrophys. J. **192**, 251.
de Vaucouleurs, G., 1953, Astron. J. **58**, 30.
de Vaucouleurs, G., 1959, Handb. Phys. **53**, 311.
de Vaucouleurs, G., 1961, in *Problems of Extragalactic Research*, edited by G. C. McVittie (Macmillan, New York), p. 3.
de Vaucouleurs, G., 1970, Astrophys. J. **159**, 435.
de Vaucouleurs, G., 1971, Publ. Astron. Soc. Pac. **83**, 113.
de Vaucouleurs, G., 1976, "Groups of Galaxies" in *Stars and Stellar Systems*, edited by A. R. Sandage and M. Sandage (University of Chicago, Chicago), Vol. IX.
de Vaucouleurs, G., and A. de Vaucouleurs, 1964, *Reference Catalogue of Bright Galaxies* (University of Texas, Austin).
de Vaucouleurs, G., and W. L. Peters, 1968, Nature **220**, 868.
Dearborn, D. S., 1973, Astrophys. J. **179**, 49.
Descartes, R., 1647, in *Principes de la Philosophie* (trans. Picot): *Oeuvres*, Vol. X (see for example Part 3, Art. 30).
Dicke, R. H., P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson, 1965, Astrophys. J. **142**, 414.
Dickens, R. J., and J. V. Peach, 1972, in IAU Symposium No. 44, *External Galaxies and Quasi-Stellar Objects*, edited by D. S. Evans (Reidel, Dordrecht/Boston), p. 89.
Doroshkevich, A. G., 1966, Astrofizika **2**, 37.
Doroshkevich, A. G., 1971, Astrofizika **6**, 581.
Doroshkevich, A. G., 1973, Astrophys. Lett. **14**, 11.
Doroshkevich, A. G., M. L. Longair, and Ya. B. Zel'dovich, 1970, Mon. Not. R. Astron. Soc. **147**, 139.
Doroshkevich, A. G., Ya. B. Zel'dovich, and I. D. Novikov, 1967, Sov. Astron.-AJ **11**, 233.
Eggen, O. J., D. Lynden-Bell, and A. Sandage, 1962, Astrophys. J. **136**, 748.
Emerson, D. T., and J. E. Baldwin, 1973, Mon. Not. R. Astron. Soc. **165**, 9P.
Epicurus, C., c. 310 B. C., *Fragmenta de Natura*, XI (translated by C. Bailey, *Epicuros*, Oxford, 1926).
Fath, E. A., 1914, Astron. J. **28**, 75.
Faye, H., 1885, *Sur l'Origine du Monde* (Gauthier-Villars, Paris), 4th ed., Chap. 13 and 14.
Field, G. B., 1964, Astrophys. J. **140**, 1434.
Field, G. B., 1967, in *Stars and Stellar Systems*, edited by A. R. Sandage and M. Sandage (University of Chicago, Chica-

go), Vol. IX.
Field, G. B., 1971, Astrophys. J. **165**, 29.
Field, G. B., 1972, Annu. Rev. Astron. Astrophys. **10**, 226.
Field, G. B., H. Arp, and J. Bahcall, 1975, *The Redshift Controversy* (Addison-Wesley, Reading).
Field, G. B., and W. C. Saslaw, 1971, Astrophys. J. **170**, 199.
Field, G. B., and L. C. Shepley, 1968, Astrophys. Space Sci. **1**, 309.
Fish, R., 1964, Astrophys. J. **139**, 284.
Forets, G. P. des, and J. Schneider, Astron. Astrophys. **26**, 397.
Freeman, K. C., 1970, Astrophys. J. **160**, 811.
Gamow, G., 1949, Rev. Mod. Phys. **21**, 367.
Gamow, G., 1951, *Creation of the Universe* (Viking, New York).
Gamow, G., 1952, Phys. Rev. **86**, 231.
Gamow, G., 1953, K. Dan. Vidensk. Selsk. Mat.-Fys. Medd. **27**, No. 10.
Gamow, G., 1956, in *Vistas in Astronomy,* edited by A. Beer (Pergamon, New York), Vol. II, p. 1726.
Gamow, G., and E. Teller, 1939, Phys. Rev. **55**, 654.
Gamow, G., S. Ulam, and N. Metropolis, 1948, unpublished.
Geller, M. J., and P. J. E. Peebles, 1973, Astrophys. J. **184**, 329.
Gott, J. R., 1973, Astrophys. J. **186**, 481.
Gott, J. R., G. T. Wrixon, and P. Wannier, 1973, Astrophys. J. **186**, 777.
Gowdy, R. H., 1971a, Phys. Rev. Lett. **27**, 826.
Gowdy, R. H., 1971b, Phys. Rev. Lett. **27**, 1102.
Hagedorn, R., 1970, Astron. Astrophys. **5**, 184.
Haro, G., 1956, Bol. Obs. Tonantzintla Tacubaya No. 14, **2**, 8.
Harrison, E. R., 1967, Rev. Mod. Phys. **39**, 862.
Harrison, E. R., 1970a, Mon. Not. R. Astron. Soc. **147**, 279.
Harrison, E. R., 1970b, Phys. Rev. D **1**, 2726.
Harrison, E. R., 1970c, Mon. Not. R. Astron. Soc. **148**, 119.
Harrison, E. R., 1971, Mon. Not. R. Astron. Soc. **154**, 167.
Harrison, E. R., 1973, Annu. Rev. Astron. Astrophys. **11**, 155.
Hauser, M. G., and P. J. E. Peebles, 1973, Astrophys. J. **185**, 757.
Hawking, S. W., 1966, Astrophys. J. **145**, 544.
Hawking, S. W., 1971, Mon. Not. R. Astron. Soc. **152**, 75.
Hawley, D., and P. J. E. Peebles, 1976, Astrophys. J. (to be published).
Heidmann, J., 1968, Astrophys. Lett. **3**, 168.
Heisenberg, W., 1948a, Z. Phys. **124**, 628.
Heisenberg, W., 1948b, Proc. R. Soc. Lond. **195**, 402.
Heisenberg, W., and C. F. von Weizsacker, 1948, Z. Phys. **125**, 290.
Helmholtz, H. von, 1858, Crelle's J. **55**.
Herschel, Sir W., 1811, Phil. Trans. R. Soc. Lond. **101**, 269.
Hills, J. K., 1973, Astrophys. J. **179**, 371.
Hoerner, S. von, 1951, in *IAU Symposium on the Motion of Gaseous Masses of Cosmical Dimensions:* Problems of Cosmical Aerodynamics (Central Air Documents Office), p. 201.
Hoerner, S. von, 1958, Fortschr. Phys. **6**, 335.
Holmberg, E., 1964, Ark. Astron. **3**, 387.
Holmberg, E., 1969, Ark. Astron. **5**, 305.
Hoyle, F., 1951, in *IAU Symposium on the Motion of Gaseous Masses of Cosmical Dimensions:* Problems of Cosmical Aerodynamics (Central Air Documents Office), p. 195.
Hoyle, F., 1965, *Galaxies, Nuclei, and Quasars* (Heinemann, London).
Hoyle, F., 1975, Astrophys. J. **196**, 661.
Hoyle, F., and J. V. Narlikar, 1966a, Proc. R. Soc. Lond. **290A**, 177.
Hoyle, F., and J. V. Narlikar, 1966b, Proc. R. Soc. Lond. **290A**, 162.
Hoyle, F., and J. V. Narlikar, 1974, *Action-at-a-distance in Physics and Cosmology* (Freeman, San Francisco).
Hu, B. L., and T. Regge, 1972, Phys. Rev. Lett. **29**, 1616.
Hubble, E. P., 1925a, Astrophys. J. **62**, 409.
Hubble, E. P., 1925b. Pub. Amer. Astron. Soc. **5**, 261.
Hubble, E. P., 1925c, Observatory **48**, 139.
Hubble, E. P., 1926, Astrophys. J. **63**, 236.
Hubble, E. P., 1929, Astrophys. J. **69**, 103.
Hubble, E. P., 1934, Astrophys. J. **79**, 8.
Hubble, E. P., 1936a, Astrophys. J. **84**, 158.
Hubble, E. P., 1936b, Astrophys. J. **84**, 517.
Hubble, E. P., 1936c, *The Realm of Nebulae* (Yale University, New Haven).
Hubble, E. P., and M. L. Humason, 1931, Astrophys. J. **74**, 43.
Humason, M. L., N. U. Mayall, and A. Sandage, 1956, Astron. J. **61**, 97.
Hunter, C., 1970, Astrophys. J. **162**, 445.
Icke, V., 1973, Astron. Astrophys. **27**, 1.
Innanen, K., 1966, Astrophys. J. **143**, 125.
Jaki, S. L., 1973, *The Milky Way* (David and Charles, North Pomfret, Vt.).
Jaschek, C. O. R., 1957, Z. Astrophys. **44**, 23.
Jeans, J., 1902, Phil. Trans. **199A**, 1.
Jeans, J., 1918, *Cosmogony and Stellar Dynamics* (Cambridge University, Cambridge).
Jeans, J., 1928, *Astronomy and Cosmogony* (Cambridge University, Cambridge).
Jones, B. J. T., 1970, Ph.D. Thesis, University of Cambridge.
Jones, B. J. T., 1972, unpublished essay.
Jones, B. J. T., 1973a, Bull. Am. Phys. Soc. **18**, 101.
Jones, B. J. T., 1973b, Astrophys. J. **181**, 269.
Jones, B. J. T., 1976, in preparation.
Jones, B. J. T., and K. L. Chan, 1976, in preparation.
Jones, B. J. T., and G. Steigman, 1976, in preparation.
Jones, J. E., 1976, in preparation.
Kant, I., 1755, *General Natural History and Theory of the Heavens,* (London).
Karman, Th.v., and L. Howarth, 1938, Proc. R. Soc. Lond. **164A**, 192.
Keeler, J. E., 1900, Astrophys. J. **11**, 346.
Kelvin, Lord, 1869, Trans. R. Soc., Edinburgh **25**.
Kepler, J., 1609, *Astronomia Nova.*
King, I. R., 1966, Astron. J. **71**, 64.
King, I. R., 1971, Publ. Astron. Soc. Pacific **83**, 377.
King, I. R., and R. Minkowski, 1972, in IAU Symp. No. 44 *"External Galaxies and Quasi-Stellar Objects,"* edited by D. S. Evans (Reidel, Boston), p. 87.
Kolmogorov, A. N., 1941, Comptes Rendus (Doklady) Acad. Sci. URSS **30**.
Kundt, W., 1971, in *General Relativity and Cosmology* edited by R. K. Sachs (Academic, New York).
Lambert, J., 1761, *Kosmologischen Briefe* (Ausberg).
Landau, L. D., and E. M. Lifshitz, 1959, *Fluid Mechanics* (Pergammon, New York).
Laplace, Marquis de, 1884, Oeuvres Vol. VI, pp. 498–509 (Gauthier-Villars, Paris). This is more complete than the 1976 edition of *Système du Monde*.
Larson, R. B., 1974, Mon. Not. R. Astron. Soc. **166**, 585.
Layzer, D., 1963a, Astrophys. J. **137**, 351.
Layzer, D., 1963b, Astrophys. J. **138**, 174.
Layzer, D., 1964, Ann. Rev. Astron. Astrophys. **2**, 241.
Layzer, D., 1968a, Astrophys. Lett. **1**, 99.
Layzer, D., 1968, Proc. Brandeis Summer School.
Lea, S., 1974, Ph.D. Thesis, Univ. of California, Berkeley.
Lemaître, G., 1950, *The Primeval Atom* (van Nostrand, New York).
Lifschitz, E. M., 1946, J. Phys. USSR **10**, 116.
Lifschitz, E. M., and L. M. Khalatnikov, 1963, Adv. in Phys. **12**, 185.
Lighthill, M. J., 1952, Proc. R. Soc. Lond. A **211**, 564.
Lighthill, M. J., 1953, in *Gas Dynamics of Cosmic Clouds* (IAU Symposium No. 2: Cambridge, England), p. 121.
Ligondes, R., 1897, *Formation Mechanique du Systeme du Monde* (Gauthier-Villars, Paris).

Limber, D. N., 1953, Astrophys. J. **117**, 116.
Limber, D. N., 1954, Astrophys. J. **119**, 655.
Limber, D. N., 1957, Astrophys. J. **125**, 9.
Lin, C. C., 1947, in *Proceedings of the 1st Symposium in Applied Mathematics* (American Mathematical Society, New York, 1949), p. 81.
Lin, C. C., and W. M. Reid, 1963, Handb. Phys. **8**, 438.
Lundmark, K., 1920, K. Sven. Vetenskapsakad. Handl. **60**, No. 8.
Lundmark, K., 1926a, Uppsala Medd. No. 8.
Lundmark, K., 1926b, Uppsala Medd. No. 16.
Lundmark, K., 1927a, Uppsala Medd. No. 30.
Lundmark, K., 1927b, Uppsala Medd. No. 41.
Lynden-Bell, D., 1967a, Observatory, **87**, 163.
Lynden-Bell, D., 1967b, Mon. Not. R. Astron. Soc. **136**, 101.
Mackay, C. D., 1971, Mon. Not. R. Astron. Soc. **151**, 421.
Matsuda, T., H. Sato, and H. Takeda, 1971, Prog. Theor. Phys. **46**, 416.
Matsuda, T., H. Takeda, and H. Sato, 1973, Prog. Theor. Phys. **49**, 1770.
Matzner, R. A., and C. W. Misner, 1972a, Astrophys. J. **171**, 415.
Matzner, R. A., and C. W. Misner, 1972b, Astrophys. J. **171**, 433.
Mayall, N. V., 1934, Lick Obs. Bull. **16**, 177.
Meier, D. L., 1975, M. A. Thesis, University of Texas, Austin.
Michie, R., 1967, unpublished.
Miller, R. H., and K. H. Prendergast, 1962, Astrophys. J. **136**, 713.
Misner, C. W., 1968, Astrophys. J. **151**, 431.
Moulton, F. R., 1900, Astrophys. J. **11**, 103.
Moulton, F. R., 1905, Astrophys. J. **22**, 165.
Mowbray, A. G., 1938, Publ. Astron. Soc. Pac. **50**, 275.
Nariai, H., 1956a, Sci. Report Tohoku Univ., Ser. 1, **39**, 213.
Nariai, H., 1956b, Sci. Report Tohoku Univ., Ser. 1, **40**, 40.
Nariai, H., 1970, Prog. Theor. Phys. **44**, 110.
Nariai, H., 1971, Prog. Theor. Phys. **45**, 61.
Nariai, H., and K. Tomita, 1971, Prog. Theor. Phys. Suppl. No. 49, p. 83.
Nariai, H., K. Tomita, and S. Kato, 1967, Prog. Theor. Phys. **37**, 60.
Ne'eman, Y., 1965, Astrophys. J. **141**, 1303.
Ne'eman, Y., and G. Tauber, 1967, Astrophys. J. **150**, 755.
Newton, I., 1644, *Principia Philosophiae* (Amsterdam).
Newton, I., 1687, *Principia* Vol. II, Prop. 52, Cor. 6.
Neyman, J., T. L. Page, and E. L. Scott, editors, 1961, Conference on the Instability of Systems of Galaxies, in Astron. J. **66**, 533.
Neyman, J., and E. L. Scott, 1959, Handb. Phys. **53**, 416.
Neyman, J., and E. L. Scott, 1961, in Proc. 4th Berkeley Symp. Math. Stat. and Probability **3**, 261.
Neyman, J., E. L. Scott, and C. D. Shane, 1952, Astrophys. J. **117**, 92.
Neyman, J., E. L. Scott, and W. Zonn, 1962, Astron. J. **67**, 583.
Noerdlinger, P., 1970, Astrophys. J. Lett. **159**, L179.
Nordseick, K. H., 1973a, Astrophys. J. **184**, 719.
Nordseick, K. H., 1973b, Astrophys. J. **184**, 735.
Novikov, I. D., 1965, Sov. Astron.-AJ **8**, 857.
Obhukoff, A., 1941, C. R. (Dokl.) Acad. Sci. URSS **32**, 19.
Oemler, A., 1973a, Astrophys. J. **180**, 11.
Oemler, A., 1973b, Ph.D. Thesis, California Institute of Technology.
Olsen, D. W., and R. K. Sachs, 1973, Astrophys. J. **185**, 91.
Omnès, R., 1969, Phys. Rev. Lett. **23**, 38.
Omnès, R., 1974, in *Physics of Dense Matter,* edited by C. J. Hansen (Reidel, Dordrecht/Boston), p. 301.
Onsager, L., 1945, Phys. Rev. **68**, 286.
Oort, J. H., 1965, in *Galactic Structure,* edited by A. Blaauw and M. Schmidt (University of Chicago, Chicago), p. 445.
Oort, J., 1970, Astron. Astrophys. **7**, 381.
Ostriker, J. P., and P. J. E. Peebles, 1973, Astrophys. J. **186**, 467.
Ozernoi, L. M., and A. D. Chernin, 1968a, Sov. Astron.-AJ **11**, 907.
Ozernoi, L. M., and A. D. Chernin, 1968b, Sov. Astron.-AJ **12**, 901.
Ozernoi, L. M., and G. V. Chibisov, 1971a, Sov. Astron.-AJ **14**, 615.
Ozernoi, L. M., and G. V. Chibisov, 1971b, Astrophys. Lett. **7**, 201.
Ozernoi, L. M., and G. V. Chibisov, 1972, Sov. Astron.-AJ **15**, 923.
Page, T. L., 1960, Astrophys. J. **159**, 791.
Parker, L., 1969, Phys. Rev. **183**, 1057.
Parker, L., 1971a, Phys. Rev. D **3**, 346.
Parker, L., 1971b, Phys. Rev. D **3**, 2546.
Parker, L., 1972, Phys. Rev. Lett. **28**, 705, 1497.
Partridge, R. B., 1974, Astrophys. J. **192**, 241.
Partridge, R. B., and P. J. E. Peebles, 1967a, Astrophys. J. **147**, 868.
Partridge, R. B., and P. J. E. Peebles, 1967b, Astrophys. J. **148**, 377.
Peebles, P. J. E., 1965, Astrophys. J. **142**, 1317.
Peebles, P. J. E., 1966, Phys. Rev. Lett. **16**, 410.
Peebles, P. J. E., 1967a, lectures presented at the 4th Texas Symposium on Relativistic Astrophysics (unpublished).
Peebles, P. J. E., 1967b, Astrophys. J. **147**, 859.
Peebles, P. J. E., 1968a, Astrophys. J. **153**, 1.
Peebles, P. J. E., 1968b, Nature **220**, 237.
Peebles, P. J. E., 1969a, J. R. Astron. Soc. Canada **63**, 4.
Peebles, P. J. E., 1969b, Astrophys. J. **155**, 393.
Peebles, P. J. E., 1970, Phys. Rev. D **1**, 397.
Peebles, P. J. E., 1971a, *Physical Cosmology* (Princeton University, Princeton).
Peebles, P. J. E., 1971b, Astron. Astrophys. **11**, 377.
Peebles, P. J. E., 1971c, Astrophys. Space Sci. **11**, 443.
Peebles, P. J. E., 1972, Comments Astrophys. Space Phys. **4**, 53.
Peebles, P. J. E., 1973, Astrophys. J. **185**, 413.
Peebles, P. J. E., 1974a, Astrophys. J. **189**, L51.
Peebles, P. J. E., 1974b, Astrophys. J. Suppl. Ser. No. 252.
Peebles, P. J. E., 1974c, Astron. Astrophys. **32**, 197.
Peebles, P. J. E., 1974d, in Proceedings IAU Symposium No. 58 (Canberra).
Peebles, P. J. E., and R. H. Dicke, 1969, Astrophys. J. **154**, 891.
Peebles, P. J. E., and E. J. Groth, 1975, Astrophys. J. **196**, 1.
Peebles, P. J. E., and M. G. Hauser, 1974, Astrophys. J. Suppl. Ser. No. 252.
Peebles, P. J. E., and J. T. Yu, 1970, Astrophys. J. **162**, 815.
Penzias, A. A., and R. W. Wilson, 1965, Astrophys. J. **142**, 419.
Perko, T. E., R. A. Matzner, and L. C. Shepley, 1972, Phys. Rev. D **6**, 969.
Poincaré, H., 1894, *Leçons sur les Hypothèses Cosmogoniques* (Hermann, Paris).
Prendergast, K. H., and E. Tomer, 1970, Astron. J. **75**, 674.
Press, W. H., and P. Schechter, 1974, Astrophys. J. **187**, 425.
Rees, M. J., 1971, in *General Relativity and Cosmology,* edited by R. K. Sachs (Academic, New York).
Rees, M. J., 1972, Phys. Rev. Lett. **28**, 1669.
Rees, M. J., and D. W. Sciama, 1969a, Comments Astrophys. Space Phys. **1**, 140.
Rees, M. J., and D. W. Sciama, 1969b, Comments Astrophys. Space Phys. **1**, 153.
Rees, M. J., and J. Silk, 1970, Sci. Am. **222**, 26.
Reiz, A., 1941, Ann. Lund Obs. **9**, 1.
Roberts, M. S., 1969, Astrophys. J. **74**, 859.
Roberts, M. S., 1976, "Neutral Hydrogen in Galaxies," in *Stars and Stellar Systems,* edited by A. R. Sandage and

M. Sandage (University of Chicago, Chicago), Vol. IX.
Roberts, M. S., and A. M. Rots, 1973, Astron. Astrophys. **26**, 483.
Robson, E. I., D. G. Vickers, J. S. Huizinga, J. E. Beckman, and P. E. Clegg, 1974, Nature **251**, 591.
Rood, H. J., 1969, Astrophys. J. **158**, 657.
Rood, H. J., 1974, Publ. Astron. Soc. Pac. **86**, 99.
Rood, H. J., T. L. Page, E. C. Kintner, and I. R. King, 1972, Astrophys. J. **175**, 627.
Rood, H. J., V. C. A. Rothman, and B. E. Turnrose, 1970, Astrophys. J. **162**, 411.
Rood, H. J., and G. N. Sastry, 1971, Publ. Astron. Soc. Pac. **83**, 313.
Rood, H. J., and G. N. Sastry, 1972, Astrophys. J. **77**, 451.
Rood, H. J., and G. A. Welch, 1971, Astrophys. J. **165**, 225.
Rosse, Lord, 1850a, Mon. Not. R. Astron. Soc. **10**, 21.
Rosse, Lord, 1850b, Phil. Trans. 499.
Ruban, V. A., and A. D. Chernin, 1972, Sov. Astron.-AJ **16**, 363.
Rubin, V. C., 1954, Proc. Nat. Acad. Sci. **40**, 541.
Ruzmaikina, T. V., 1972, Sov. Astron.-AJ **16**, 991.
Sachs, R. K., 1974, lectures given at Banff Summer School, edited by W. Israel (Reidel, Boston).
Sachs, R. K., and A. M. Wolfe, 1967, Astrophys. J. **147**, 73.
Salpeter, E. E., 1959, Astrophys. J. **129**, 608.
Sandage, A., 1961, *The Hubble Atlas of Galaxies* (Carnegie Institution of Washington, Washington, D. C.).
Sandage, A., K. C. Freeman, and N. R. Stokes, 1970, Astrophys. J. **160**, 83.
Sandage, A., G. Tamman, and E. Hardy, 1972, Astrophys. J. **172**, 253.
Sargent, W. L. W., 1970, Astrophys. J. **160**, 405.
Sargent, W. L. W., and L. Searle, 1970, Astrophys. J. Lett. **162**, L155.
Sargent, W. L. W., and L. Searle, 1972, Comments Astrophys. Space Phys. **3**, 111.
Saslaw, W. C., 1968, Mon. Not. R. Astron. Soc. **141**, 1.
Saslaw, W. C., 1969, Mon. Not. R. Astron. Soc. **143**, 437.
Saslaw, W. C., 1970, Mon. Not. R. Astron. Soc. **147**, 253.
Sastry, G. N., 1968, Publ. Astron. Soc. Pac. **80**, 252.
Sato, H., 1971, Prog. Theor. Phys. **45**, 370.
Sato, H., T. Matsuda, and H. Takeda, 1971, Prog. Theor. Phys. Suppl. No. 49, p. 11.
Sciama, D. W., 1971, in *General Relativity and Cosmology*, edited by R. K. Sachs (Academic, New York).
Seares, F. H., 1925, Astrophys. J. **62**, 168.
Searle, L., W. L. W. Sargent, and W. Bagnuolo, 1973, Astrophys. J. **179**, 429.
See, T. J. J., 1910, *Researches on the Evolution of Stellar Systems: The Capture Theory of Cosmical Evolution* (Nichols, Tulsa), Vol. II, Chap. 19.
Sexl, R., and H. K. Urbankte, 1969, Phys. Rev. **179**, 1247.
Shane, C. D., and C. A. Wirtanen, 1950, Proc. Amer. Phil. Soc. **94**, 13.
Shane, C. D., and C. A. Wirtanen, 1954, Astron. J. **59**, 285.
Shane, C. D., and C. A. Wirtanen, 1967, Publ. Lick Obs. **22** (Pt. 1), 49.
Shapiro, S. L., 1971, Astron. J. **76**, 291.
Shapley, H., 1919, Publ. Astron. Soc. Pac. **31**, 261.
Shapley, H., 1932, Harvard Bull. **890**.
Shapley, H., 1933, Proc. Nat. Acad. Sci. USA **19**, 591.
Shapley, H., and A. Ames, 1932, Ann. Harvard Coll. Obs. **88**, 41.
Silk, J., 1968, Astrophys. J. **151**, 459.
Silk, J., 1973a, Comments Astrophys. Space Phys. **5**, 9.
Silk, J., 1973b, in IAU Symposium No. 63, "Confrontation of Cosmological Theories with Observational Data," edited by M. S. Longair (Reidel, Dordrecht, and Boston), p. 175.
Silk, J., 1974, Private communication.
Silk, J., and S. Ames, 1972, Astrophys. J. **178**, 77.
Silk, J., and S. Lea, 1973, Astrophys. J. **180**, 669.
Stecker, F. W., and J. L. Puget, 1972, Astrophys. J. **178**, 57.
Steigmann, G., 1974, in IAU Symposium No. 63, "Confrontation of Cosmological Theories with Observational Data," edited by M. S. Longair (Reidel, Dordrecht, and Boston), p. 347.
Stein, R. W., 1974, Astron. Astrophys. **35**, 17.
Stewart, J. M., 1969, Mon. Not. R. Astron. Soc. **145**, 347.
Stewart, J. M., and D. W. Sciama, 1967, Nature **216**, 748.
Sunyaev, R. A., 1971, Astron. Astrophys. **12**, 190.
Sunyaev, R. A., and Ya. B. Zel'dovich, 1970, Astrophys. Space Sci. **7**, 1.
Sutherland, W., 1911, Astrophys. J. **34**, 251.
Swedenborg, E., 1734, *Principia Rerum Naturalium* (Dresden).
Takase, B., and H. Kinoshita, 1967, Publ. Astron. Soc. Jpn. **19**, 409.
Tarter, J., and J. Silk, 1974, Q. J. R. Astron. Soc. **15**, 122.
Taylor, G. I., 1935, Proc. R. Soc. Lond. **151** A.
Taylor, G. I., 1936, Proc. R. Soc. Lond. **156** A.
Taylor, G. I., 1938, Proc. R. Soc. Lond. **164** A, 476.
Tinsley, B. M., 1972a, Astron. Astrophys. **20**, 383.
Tinsley, B. M., 1972b, Astrophys. J. **178**, 319.
Tomita, K., 1969a, Prog. Theor. Phys. **37**, 60.
Tomita, K., 1969b, Prog. Theor. Phys. **37**, 831.
Tomita, K., 1969c, Prog. Theor. Phys. **42**, 9.
Tomita, K., 1971, Prog. Theor. Phys. **45**, 1747.
Tomita, K., 1972a, Prog. Theor. Phys. **47**, 416.
Tomita, K., 1972b, Prog. Theor. Phys. **48**, 1503.
Tomita, K., 1973, Publ. Astron. Soc. Jpn. **25**, 291.
Tomita, K., H. Nariai, H. Sato, T. Matsuda, and H. Takeda, 1970, Prog. Theor. Phys. **43**, 1511.
Totsuji, H., and T. Kihara, 1969, Publ. Astron. Soc. Jpn. **21**, 221.
Turnrose, B. E., and H. J. Rood, 1970, Astrophys. J. **159**, 773.
Ulrich, M. H. J., 1971, Astrophys. J. **163**, 441.
van den Bergh, S., 1962, Z. Astrophys. **55**, 21.
van den Bergh, S., 1969, Astrophys. J. Suppl. Ser. **19**, 145.
Vandervoort, P., 1970, Astrophys. J. **162**, 453.
Veneziano, G., 1968, Nuovo Cimento A **57**, 190.
Vorontsov-Vel'yaminov, 1966, Sov. Astron.-AJ **10**, 184.
Wagoner, R. V., W. A. Fowler, and F. Hoyle, 1967, Astrophys. J. **148**, 3.
Weber, J., 1970, Phys. Rev. Lett. **25**, 180.
Weinberg, S., 1971, Astrophys. J. **168**, 175.
Weinberg, S., 1972, *Gravitation and Cosmology* (Wiley, New York).
Weizsacker, C. F. von, 1943, Z. Astrophys. **22**, 319.
Weizsacker, C. F. von, 1948, Naturwissenschaften **35**, 188.
Weizsacker, C. F. von, 1951, Astrophys. J. **114**, 165.
Weizsacker, C. F. von, 1951, in Proceedings of the *IAU-IUTAM Symposium on the Motion of Gaseous Masses of Cosmical Dimensions:* Problems of Cosmical Aerodynamics, (Central Air Documents Office), p. 201.
Weymann, R., 1966, unpublished.
Wilczyinskii, E. J., 1896, Astrophys. J. **4**, 97.
Wilson, C., 1973, Ph.D. Thesis, University of California.
Wolf, M., 1901, Astron. Nachr. **155**, 127.
Wolf, M., 1905, Astron. Nachr. **170**, 211.
Woody, D. P., J. C. Mather, N. S. Nishoika, and R. L. Richards, 1975, Phys. Rev. Lett. **34**, 1036.
Wright, T., 1750, *An Original Theory, or New Hypothesis of the Universe founded upon the laws of Nature, etc.* (London).
Yahil, A., and J. P. Ostriker, 1973, Astrophys. J. **185**. 787.
Yu, J. T., and P. J. E. Peebles, 1969, Astrophys. J. **158**, 103.
Zel'dovich, Ya. B., 1965, Adv. Astron. Astrophys. **3**, 241.
Zel'dovich, Ya. B., 1970, Astron. Astrophys. **5**, 84.
Zel'dovich, Ya. B., 1971, Comments Astrophys. Space Phys. **3**, 179.
Zel'dovich, Ya. B., 1972, Mon. Not. R. Astron. Soc. **160**, 1P.
Zel'dovich, Ya. B., V. G. Kurt, and R. A. Sunyaev, 1969, Sov. Phys.—JETP **28**, 146.
Zel'dovich, Ya. B., and I. D. Novikov, 1967, Annu. Rev.

Astron. Astrophys. **5**, 627.
Zel'dovich, Ya. B., and I. D. Novikov, 1969, Astron. Zh. **46**, 960 [Sov. Astron.-AJ **13**, 754 (1970)].
Zel'dovich, Ya. B., and I. D. Novikov, 1970, Astrofyzika **6**, 379.
Zel'dovich, Ya. B., and A. A. Starobinshky, 1972, Sov. Phys.-JETP **34**, 1159.
Zel'dovich, Ya. B., and R. A. Sunyaev, 1969, Astrophys. Space Sci. **4**, 301.
Zwicky, F., 1957, *Morphological Astronomy* (Springer, Heidelberg).
Zwicky, F., 1966, Astrophys. J. **143**, 192.
Zwicky, F., E. Herzog, P. Wild, M. Karpowicz, and C. T. Kowal, 1961–1968, *Catalogue of Galaxies and Clusters of Galaxies* (California Institute of Technology, Pasadena).