COMMITTEE I
Unity of Science: Organization and
Change in Complex Systems

DRAFT - 8/15/85
For Conference Distribution Only

GRAVITATION AND THE ORIGIN OF LARGE STRUCTURES IN THE UNIVERSE

by

Jacob D. Bekenstein
Arnow Professor of Astrophysics
Ben-Gurion University
Beersheva, ISRAEL

The Fourteenth International Conference on the Unity of the Sciences
Houston, Texas November 28-December 1, 1985

1. THE EXPANDING UNIVERSE

The why of structure and organization in the physical world has ever fascinated mankind. Only in this century was enough understood about the quantum world to allow an explanation of organization in everyday objects: the order in a salt crystal, the precisely architectured shape of a benzene molecule, ... The organization shown by heavenly bodies also prompted early the question of how matter has segregated and organized into stars. Here not electromagnetism and quantum effects, but rather the long-range force of gravity is responsible. And who else but Isaac Newton should have been the first to put forth the idea that gravity has, over the eons, gathered matter, originally spread out homogeneously, into clumps which we see as stars (Koyre 1958). Newton was essentially right, but the actual situation is far more complicated than he envisaged; complete understanding of the organization apparent in the astronomical world still eludes us.

The first point to make is that the structure has arisen in an expanding universe, a fact unknown to Newton. This has manifold consequences for our understanding of the process. Around the turn of the century the British physicist Sir James Jeans developed in detail the mathematical theory of the process Newton had described two centuries earlier (Jeans 1902, 1929). He showed that Newton's hunch was right, and that even when account is taken of the pressure exerted by the medium pervading the universe, condensations will still form provided only that they involve a minimum mass (today we speak of the Jeans mass). In Jeans' scenario, the initial medium must contain "seed" inhomogeneities in order for condensations to appear at all. However, the growth in strength (or density contrast with respect to the surroundings) of the latter is exponential in time, so that weak seeds are also effective. We

know that in any seemingly homogeneous medium there must be some seeds, if only because of the particulate nature of matter and statistical fluctuations. Thus, if Jeans' study were the whole story, there would be nothing to add in the present paper.

The realization that our universe expands, first expressed by the great American astronomer Edwin Hubble (Hubble 1929) immeasurably complicated matters. A theory of the expansion became possible in terms of the theory of General Relativity completed by Albert Einstein in 1915 (Einstein 1916). Einstein had applied very early the theory to cosmology (Einstein 1917), and it is a well known tale how, committed as he was to the philosophical view of an unchanging universe, he missed the chance to predict the expansion of the universe. This honor was claimed by the Russian mathematician Alexander Friedmann (Friedmann 1922, 1924) who invented the cosmological models used today as a basis of the description of the universe's evolution. (On this see the accompanying paper by Sexl.)

Friedmann's models, which are based on General Relativity, indicate that the universe began in a highly dense state of virtually infinite density, and expanded thereafter by a very large factor. One can speak of the universe when it was a thousandth of its present size, or even a billion times smaller. The availability of General Relativity led the Russian physicist Evgenii Lifshitz to reconsider Jeans' calculation in the framework of Friedmann's models. He found that seeds still grow in strength, but no longer exponentially (Lifhitz 1946). In effect the expansion fights tooth and claw against local gravity and almost succeeds in neutralizing the Newton-Jeans growth. Lifhitz established (and this has been confirmed time and again by many others) that the strength of inhomogeneities grows in direct proportion to the size of the universe. For example, to achieve a

thousanfold magnification, a seed must have been present when the universe was 1000 times smaller than today (or as this is expressed in the jargon, it must have been present at redshift 1000).

So what? you say. Let us assume the seeds were present at a an early enough stage in the universe, so that enough expansion has elapsed till today to build those seeds into strong inhomogeneity. Actually, such answer would be acceptable if the universe contained only matter. But there is also radiation, not only starlight, and radio waves from active galaxies, and X-rays from neutron stars, but also the famous microwave radiation background. It was first observed in 1965 by Arno Penzias and Robert Wilson (Penzias and Wilson 1965) and won them the 1978 Nobel Prize in Physics (for a fuller account refer again to Sexl's paper in this volume). The radiation changes the terms of our problem entirely. Virtually all scientists agree that the background is radiation that severed contact with the emitting matter long ago, when the universe was much smaller than today (1000 times smaller is a good guess). Now in your imagination trace backwards the expansion of the universe. The matter in it becomes denser in inverse proportion to the cube of the universe's size. Not so the radiation; its energy density grows faster because photons gain energy as their wavelength shrinks with the universe. We need not go back very far in this imaginary odyssey before the density of energy in radiation overwhelms that of matter. The picture is now of a universe dominated in its early dynamics by radiation.

Now when Lifhitz's calculation is redone in the radiation-dominated universe, it predicts <u>no</u> growth of inhomogeneities. Growth sets in only late in history when matter begins to dominate the dynamics. That, coincidentally, also happened when the universe was about 1000 times smaller

than presently. Thus seeds can only have grown by a factor of 1000 in our universe. Were the statistical seeds large enough initially to give imhogeneities as we see them today? Far from it. We know that statistical fluctuations in the number of particles that find themselves by chance in an arbitrary volume are about the square root of the mean number. If we are interested, say, in the inhomogeneity destined to form a star, we must ask about the seed size for an amount of matter containing some 10^{57} particles (about a solar mass of atoms). Evidently the root of 10^{57} is a negligible fraction of itself: the seed strength in this case is well below the level of 1/1000. Statistical seeds could not have given rise to stars, or any larger objects, in the expanding universe!

The conclusion must be that early in the universe's expansion, when radiation still held sway, there must already have been seed inhomogeneities much larger than statistical seeds. This is the way cosmologists today look at the origin of structure in the universe. This viewpoint leaves much to be desired because it relies on "initial conditions" to beget organization. Yet, as a pragmatical philosophy it has proved fruitful, and of late a rationalization for it has emerged from the so called "inflationary" cosmological model which will be described briefly in Sec. 5.

2. THE REALM OF THE GALAXIES

Having mentioned the difficulty facing Newton's conjecture about the origin of organization because of the expansion of the universe, let us turn and ask what does the universe look like at present. As late as the 1920's it was accepted that stars are the basic building blocks in the heavens. In fact, when Einstein invented relativistic cosmology, he always imagined a

universe of homogeneously distributed stars. With Hubble's demonstration (Hubble 1926) that the spiral "nebulae" are distant analogues of our own stellar system, the Milky Way galaxy, it became evident that galaxies are the basic units in the universe. Stars may be likened to cells of an organism (galaxy). The analogy just drawn is not an idle one. Multiple lines of evidence suggest that stars in a galaxy were formed after their mother galaxy had become a separate entity in the medium pervading the universe. And just as cells in an organism die and are replaced, so stars in a galaxy may die (witness the supernovae), and stars are born continuously in a large fraction of the galaxies. So if galaxies are the units in the universe, we are faced with two questions: What does the inner structure of galaxies look like, and how did it arise? How are galaxies organized in the universe?

Let us take up the first question in this section. Just as the organisms we likened them to, galaxies are of many species and genera. To avoid getting lost in the "taxonomy" of galaxies (also originally due to Hubble), let it be said at the outset that the majority of galaxies are composed of a roundish component, the spheroid, and a disk rotating about the center of the spheroidal component. In spiral galaxies such as our own (see Figs. 1-4), this division is very appropriate, though here and there there are spirals whose visible spheroid is minute. Elliptical galaxies are almost pure spheroid (though there are a number of ellipticals sporting small disks in their central regions). To be sure a few percent of galaxies are irregular galaxies with no easily defined shape (see Fig. 4), and do not fit easily into the spheroid-disk paradigm. By contrast the large group of lenticular galaxies have clear spheroid-disk morphology even though lacking spiral structure entirely.

What forces shaped most galaxies according to the spheroid-disk motif?

The prevailing view among astrophysicists might be summarized thus. As in Newton's original proposal, the gas filling the universe began to grow condensations under the action of gravity, and the condensations then collapsed on themselves. These were protogalaxies, and they must have been roughly spherical. As the gas was compressed in the collapse and lost energy to radiation, the Jeans mass would have decreased rapidly meaning that each essentially smooth part of the galaxy was allowed to fragment into smaller lumps. The process may have repeated until the lumps took on stellar proportions and stars were formed. All this must have been accomplished as the collapse went on. The fresh stars, once formed, would move only under the influence of gravity and, in effect, would form a "gas" of stars. And just as a gas fully fills the receptacle confining it, so would the stars fill the entire volume occupied by the protogalaxy when star formation began. Thus was the spheroid formed.

No process is pefectly efficient: some of the protogalaxy's gas must have escaped condensation into stars, and continued to collapse. It stands to reason that the protogalaxy had some angular momentum. At least one mechanism is known, tidal interaction, which could have given angular momentum to the protogalaxies before they separated much (Peebles 1980). The leftover gas would share some of this angular momentum, and would thus be prevented, by centrifugal forces, from falling to the center of the spheroid. Instead, it must have settled into a flattened rotating disk in the plane perpendicular to the angular momentum vector. In the disk the gas cooled by radiation, and must also have started to form stars, tough in a protracted manner. In this way were disks formed.

Why are some galaxies (ellipticals) nearly all spheroid and no disk? The prevalent opinion is that the low angular momentum of these promoted very

efficient star formation during protogalaxy collapse, so that no gas was left for the disk. By contrast a high angular momentum would have converted the radial collapse into collapse to a disk. In this manner one can understand the nearly pure disk spiral galaxies.

We now turn to the question of spiral structure, surely the most aesthetically striking feature in galaxies. How does it arise, this highly organized structure? The first point to make is that spiral arms in galaxies are delineated by young stars and regions of gas; old stars are not found there. This immediately suggests that the spirals are not material structures but travelling waves. For were they material spirals, they should disappear soon since stars cannot remain young forever, and new stars cannot be formed at the high rate required to keep the bulk of the stellar population in the spiral arms young over billions of years. Further, any material structure inscribed on a rotating galaxy disk could not last long. All galactic disks rotate differentially, that is, the angular velocity steadily decreases with distance from the center, at least outside the very central regions. This uneven rotation would wind up material arms and destroy them over a period of some 10^8 years. Galaxies are suspected to be some 10^{10} years old, so spiral structure would be a rare occurrence if the spirals were material. This is belied by the facts: a major portion of disk galaxies have spiral structure.

The idea that the spirals are traveling waves first occurred to the Swedish astronomer Bertil Lindblad (Lindblad 1927). An elaborate theory of "spiral density waves" was worked out only much later by Chia Chiao Lin and Frank Shu (Lin and Shu 1964). The basic idea of this much developed theory is that a galactic disk made of stars and gas can serve as a propagation medium for spiral shaped waves which circulate around it rigidly (in contrast

with the stars and gas which orbit around it differentially), and at a steady angular velocity. The waves are density waves in the sense that gas swept up by them is compressed, and by Jeans' mechanism gives birth to stars. The newborn stars delineate the spiral. As they age, they are left behind and the wave induces new star formation to replace them. All this is very much like a conflagration sweeping through a dry forest. A distant observer, upon seeing the line of fire advancing through it, might regard it as some travelling material structure. In fact what advances is the front between the charred trees and the yet untouched ones. In a galaxy the spiral arms separate a region in which star "ignition" has just ended from one in which much gas awaits the chance to be turned into stars.

A striking confirmation of the theory is provided by the lenticular galaxies which are disk-spheroid galaxies with no spiral structure whatsoever. Optical and radiowave studies have verified that the disks of lenticulars are devoid of gas. Either early star formation was vigorous and consummed all the gas that fell onto the disk, or some catastrophe, like a near collision with a neighbor galaxy, has swept the gas out. At any rate, since there is no gas to make stars, no "conflagration" can propagate, and the spiral structure cannot express itself.

The mathematical theory of spiral waves (Toomre 1977) makes it clear that the waves propagate as a result of the interplay of gravitation, pressure and rotation in the disk. Without gravity there would be no spiral waves, just as without gravity no waves could propagate on the surface of a pond. Altough the spiral wave theory has had successes, it is still unable to give an account of the <u>origin</u> of the waves. Propagation of the spirals, once formed, is understood; the mechanism that triggers them is not. Of the various triggers suggested (a quickly rotating central bar in the galaxy,

perturbation from another galaxy...), none seems to be the universal answer.

It is significant that some of the internal organization in galaxies, i.e., the spheroid-disk organization, reflects material structures, while other organization reflects a traveling phenomenon, a wave. This dichotomy is unique and not found at lower or higher levels in the universe.

Altough we have stressed the galactic level of structure, it is well to point out that the stellar component of galaxies is not amorphous when examined at small scales. Stars are grouped into doubles, triplets and associations. And there are clusters of stars with populations ranging from hundreds to hundreds of thousands. There is, however, evidence that clusters and associations can disperse, so that the organization just mentioned may be ephemeral (Fall and Rees 1977). Let us thus turn attention from galaxy interiors outward.

3. THE FABRIC OF THE UNIVERSE

We now turn to the second question raised in the last section: how are galaxies organized in the universe? The early impression of astronomers was of a rather homogeneous distribution of galaxies over the sky if allowance was made for obscuration associated with the Milky Way itself. Out of this impression arose the Cosmological Principle which states that, on a very large scale, matter in the universe is distributed homogeneously. Of course on smaller scales matter is distributed irregularly. One of the key questions of cosmology is where to draw the line between these two regimes.

By the 1930's it was known that some galaxies appear in pairs, and that there are small groups as well as large clusters of galaxies. One of the large clusters, the Coma Cluster, played an important role in the early

discovery by Fritz Zwicky, that great American astronomer of Swiss origin, of the discrepancy between the mass seen in extragalactic systems, and the mass that should be there on dynamical grounds (Zwicky 1933). With the completion of the Palomar Observatory photographic survey of the sky in the 1950's it became crystal clear that clusters, far from being an occasional feature, are quite prevalent in the universe.

Much speculation attended the question of whether the hierarchy of clumping continues to higher level: clusters of clusters... In particular, one should mention Gerard de Vaucouleurs' farsighted belief (de Vaucouleurs 1953) that our own group of galaxies, the Local Group, is part of a large supercluster which also includes the populous Virgo Cluster at a distance of some 30 million light years. The concept of superclusters was not very popular in the 60's and early 70's. A well known cosmology text of that period, citing the giants of astronomy, claims "...the hierarchy stops at clusters of galaxies or at most at clusters of clusters of galaxies, and shows no evidence of inhomogeneities of larger scale..." (Weinberg 1972).

The tide started turning in the mid-1970's. At that time James Peebles at Princeton analyzed in a systematic way the correlation of positions of galaxies as seen in the sky and came to the conclusion that even loose galaxies are not sprinkled over the heavens at random (Peebles 1980). Rather, there is a clear tendency for galaxies to "hang together" even if the plain eye does not reveal a well defined group. The correlation function, which expresses this finding quantitatively, has by now become one of the basic tools for understanding organization in the universe, and is also regarded as a rich source of information about the early universe. The correlation function approach showed that there is more organization among galaxies than meets the eye. It did not reveal the full fabric of the

universe because it was based on a two-dimensional map of galaxies.

Up to the mid-1970's positions of galaxies projected on the heavenly sphere could easily be determined, but distances could be established moderately accurately only for a small minority. At that time the development of "mass production" techniques for measuring distances to galaxies via their redshift (Hubble's expanding universe hypothesis relates distance of a far object to its redshift) allowed astronomers to compose an extensive three-dimensional map of the universe (Huchra et.al. 1983) Two great surprises followed. First, it was found that galaxies and clusters tend to fall on chain-like or sheet-like structures. These were named superclusters (Oort 1983). Thus was de Vaucouleurs' insight verified and extended. To the best of present knowledge, superclusters are the largest structures in the universe. Many stretch out over distances of tens of millions of light years and encompass tens of thousands of galaxies. The superclusters form a veritable maze in space, making up the filaments of the cosmic fabric. The second discovery was that in between superclusters space is virtually empty: very few glaxies have been detected in these <u>voids</u>. The largest known void (Kirshner et.al. 1981) spans nearly 300 million light years of empty space.

This, then, is the fabric of the universe. How was it formed? The first question that must be confronted is a modern version of the proverbial querry: who came first, the chicken or the egg? Did galaxies form first and then clumped to form clusters which then grouped into superclusters, or did superclusters form first and then splintered into clusters which then fragmented into galaxies? Both schemes can be based on the Newton-Jeans process. The distinction between them must hark back to the scale of the important inhomogeneities in the matter filling the universe in primordial

times. If the salient inhomogeneities involved masses of the order of a galaxy's, it is a good guess that galaxies formed first and then clumped in response to their mutual attraction. If, by contrast, the salient inhomogeneities involved masses equivalent to tens of thousands of galaxies, it would be a safe guess that superclusters emerged first and galaxies were born from them by repeated fragmentation.

No unanimity has yet been reached by cosmologists discussing these possibilities. Two cosmological scenarios contend today for primacy in explaining the large scale organization of the universe. First we have the so called hierarchical scenario espoused by Peebles and many of his colleagues (Peebles 1980, Gott and Rees 1975). It assumes that the primeval inhomogeneities were matter density inhomogeneities in the early universe in which small scales were most salient, and large scales less, but with a smooth transition in strength from scale to scale (technically the density contrast is proportional to some inverse power of the scale). Under such conditions galaxies would form first from the small scale inhomogeneities, and would then begin to cluster under the influence of the weaker but larger scale inhomogeneities. It is even possible in this scenario that galaxies were not the first structures to form, but rather objects with masses similar to today's globular star clusters, which themselves went on to cluster and form galaxies (Fall and Rees 1977).

The rival scenario is the pancake scenario espoused by the prominent Soviet physicist Yakov Zel'dovich and his colleages (Zel'dovich 1972, Doroshkevich, Sunyaev and Zel'dovich 1974). The pancake scenario (the rationale for the name will be clear below) assumes that the primeval inhomogeneities were joint matter and radiation irregularities (remember that early on radiation was more intense than today). Small scale

inhomogeneities of this type were susceptible to erasure at the epoch when radiation severed direct interaction with matter as a result of the recombination of ions and electrons to atoms. Calculations first carried out by Joseph Silk (Silk 1968) showed that inhomogeneities involving masses less than some 10^{15} solar masses will not survive past the recombination epoch. Therefore, in this scenario the first structures to form have masses like the superclusters (thousands of galactic masses). Clusters and galaxies must appear later by repeated fragmentation.

How do these scenarios fare in explaining the evidence? One check of the hierarchical scenario is possible by numerical simulations which follow the motion of a large assemblage of particles (galaxies or smaller entities) subject to Newtonian gravitation and in a steadily expanding background. Numerous "N-body simulations" of this type have been carried out on large computers, and the general impression is that the combination of expansion and gravitational attraction is indeed responsible for the tendency of galaxies to form stable clumps as observed. The specific form of the correlation function found by Peebles from the data is also explained (Gott 1979). Some of the simulations even show voids, but the maze-like organization of the superclusters has not yet found a natural place in the hierarchical scenario.

Unlike the previous approach which studies motion of discrete particles, any simulation relevant to the pancake scenario has to take into account that the matter about to form a supercluster is still gaseous (no stars or galaxies exist yet by definition). The dynamics is thus complicated by fluid dynamic effects. These, together with the universal expansion, make it natural for the incipient supercluster to collapse not spherically, but to a highly flattened structure, a "pancake" in the jargon of the subject. A

shock ensues followed by cooling of the gas due to radiation, and both effects facilitate the fragmentation of the pancake into clusters and galaxies. Neighboring pancakes may intersect, and the dense loci of crossing are especially favorable to fragmentation. Thus chains of galaxies and voids between them have a natural place in the pancake scenario.

However, the pancake scenario has also had setbacks. Observational constraints on the magnitude of primeval joint matter-radiation inhomogeneities (see Sec. 4) forced theorists to modify the scenario by supposing that the "matter" the scenario deals with is mostly, not ordinary elements, but rather a gas of neutrinos endowed with rest mass. This hypothesis allows the constraints to be sidestepped neatly. But the modified scenario has lately been shown to be incapable of explaining the precise form of the galaxy correlation function (White et.al. 1984). The latest strategy for the pancake scenario now calls for matter mostly made up of esoteric massive elementary particles called axions. It is too early to pass final veredict on this approach, except to point out that it is typical of the current trend that weaves the physics of elementary particles and cosmology into a common cloth.

We may also mention Ostriker's maverick explosion scenario (Ostriker and Cowie 1981, Schwarz, Ostriker and Yahil 1975). It holds that the origin of the large linear structures is in the intersection of expanding shells expelled by violent explosions of a class of primordial supermassive objects. It is a sobering thought that detonations akin to those supernovae that end the lives of massive stars may turn out responsible for the birth of the superclusters of galaxies.

4. LOOKING INTO THE PAST

Observations of very distant objects in the universe are tantamount to a "time machine" that takes us back to long past times when the light we observe in our lifetime was just leaving those objects. Can such a look back to the era of galaxy formation tell us whether galaxies or superclusters came first? Can it help us verify that there were really inhomogeneities long before galaxies or superclusters became evident?

To answer the first question one would like to detect both galaxies and superclusters at very large distances (equivalently, at very large redshifts caused by the intervening expansion of the universe) to tell which kind of object is seen farther out. Unfortunately, present earthbound telescopes with existing detectors are only able to see isolated galaxies out to a redshift of about one (meaning that one sees them at the time the universe was half of today's size). By all accounts this is not reaching deep enough into the past to allow us to tell directly whether galaxies came first. Doubtlessly, with the projected launch in 1986 of the Hubble Space Telescope by NASA, the prospects for this approach will improve.

Even from Earth's surface progress may be made by other avenues. Quasars, those bright beacons shining from afar, can now be detected at redhifts approaching 4 (that is we see them as they were when the universe was a fifth of its present size). The weight of opinion now is that a quasar is an outburst in a galaxy's nucleus. If true this means we can indirectly see some galaxies at redshifts which are becoming relevant for testing our querry. One way to check whether superclusters came first has been developed by Patrick Osmer (Osmer 1983): find enough high redshift quasars to decide

whether they tend to cluster as if they belonged to superclusters. That program is still in its infancy, but may be able to enlighten us within a few years.

And now for our second question: is there evidence for the presence of inhomogeneities long before galaxies and superclusters took shape? To answer this we must evidently look out to very high redshifts, and the only messenger from those regions (and epochs) known to us is the microwave background radiation we have alluded to in Section 1.

A striking feature of this radiation is its purely thermal spectrum: it looks just like radiation emitted by a black body at a temperature of 2.7 degrees Kelvin. As mentioned in Section 1, we can infer that the energy density of the radiation (and its temperature) grow very rapidly as we go back in time. The equations of General Relativity allow us to extrapolate this trend almost to the point that the density and temperature were infinite and the universe had zero radius. Thus the beginning of the universe was associated with blinding intensity of radiation. From here the name "primeval fireball" sometimes given to the radiation background. At early times the radiation was in intimate contact with matter (in physicist's jargon there was thermodynamic equilibrium) and this led to its striking thermal character today, but the contact must have been severed as the primordial hot plasma turned into unionized gas at the epoch when the temperature fell to some 3000 degrees Kelvin. Since then the radiation must have been travelling unhindered through the expanding universe for long ages.

Any inhomogeneity present both in the matter and radiation before they decoupled must have left an imprint on the freed radiation. Today this should be reflected in variations of the radiation's intensity over the sky. In this way one can expect to see directly the inhomogeneities at very early

times, in fact at a redshift of 1000 corresponding to the decoupling epoch (This is much farther than can be probed by telescopes looking for galaxies). Yet, a second striking feature of the background radiation is its isotropy (extreme uniformity with respect to direction) on all angular scales. For example, recent measurements (Uson and Wilkinson 1984) have shown that over angles of a few arc-minutes the radiation intensity is uniform to an accuracy of a few thousands of a percent! Very smooth indeed. One can calculate that at redshift 1000 an angle of a few arc-minutes corresponds to a region encompassing a supercluster sized mass.

So, if the original pancake scenario were right in that primordially there were joint radiation and matter inhomogeneities, then the density contrast achieved by the matter's inhomogeneities, after the growth associated with a 1000-fold expansion of the universe, would be a few percent today. Yet the superclusters today have a much higher density contrast with respect to their surroundings. The scenario thus runs int trouble. It is precisely for this reason that the modification of the scenario involving massive neutrinos (see Sec. 3) was given serious consideration.

The hierarchical scenario is less badly hit by the observations because it posits pure matter inhomogeneity in primordial times, and the lack of variations in the radiation intensity does not directly clash with this assumption. However, to explain superclustering, if it can do that at all, this scenario must posit sizeable (of order 0.1%) matter inhomogeneities at redshift 1000. Such matter irregularities, when they go in motion, must willy-nilly induce irregularities in the radiation via the Doppler effect (Davis 1980). Thus it is generally agreed that a small improvement in the measurements of the radiation isotropy will bring woes for the hierarchical scenario too.

In such eventuality, the less accepted explosion scenario may come to the fore. It relies only on primeval inhomogeneities on very small scale (corresponding to the mass of a supermassive object, not even a galaxy's). Technically, it is very difficult to establish accurate isotropy limits on the required angular scales of arc-seconds, so it is unlikely that the scenario will fall on this score. In the explosion scenario the large scale structure comes from the ejected shells, and does not require primordial inhomogeneities on scales susceptible to accurate observational scrutiny. However, it remains to be seen, when details of the explosion scenario are worked out in depth, whether it does not run aground on unexpected effects.

5. AFTER THE BEGINNING

Assuming a particular type of initial inhomogeneity has proved a handy procedure for cosmologists in their attempt to understand organization in the heavens. Yet much thought has gone into trying to explain where the initial inhomogeneities came from. We mentioned that the simplest explanation, that they are statistical fluctuations in the distribution of particles, fails misserably. Alternative explanations have usually posited an irregularity in some other physical quantity (magnetic field strength, matter velocity...) which then "infects" the matter density. This is, evidently, only postponing the problem. At some point one must come to terms with the real issue: in fundamental terms what were the initial conditions in the universe like, and how did the required inhomogeneities arise from them?.

What did the universe look like immediately after emergence from the "beginning"? If we do not wish to introduce ad hoc assumptions, there is

very little leeway here: the universe must have started accurately homogeneous and isotropic (looking the same not only at all locations, but also in all directions). To have it otherwise would require specification of initial parameters describing the inhomogeneities and anisotropies. Since the universe is not just an example of a class of systems, but the only universe, such parameters would take on the role of physical laws. But having so many laws would clearly prove inimical to a rational picture of cosmology.

Altough I have presented this view matter of factly, it is actually a minority view today. Perhaps more popular is Misner's hypothesis (Misner 1969) that initially the universe was as chaotic as possible, and only became nearly smooth as a result of complex dissipative processes. It is only when we appeal to thermodynamics that the strength of the contrary hypothesis, "highly smooth initial universe", shines through clearly.

The striking lesson of thermodynamics is its second law: "in a closed system the entropy cannot decrease, and will usually increase if the system undergoes a change" (for the significance of the concept of entropy, refer to Sexl's review on entropy in this volume). We defer for the moment the question of whether the universe is a closed system in any sense, and presume the second law to be valid fo the universe. It is perhaps ironic that despite the wide applicability of the second law in science, any clearcut explanation of it in terms of dynamical laws has ever been thwarted by the fact that all physical dynamics (with a small exception - the superweak interaction responsible for the K mesons oscillations) are symmetric under time reversal. Hence the dynamics cannot be exclusively responsible for the increase of entropy. It is actually the boundary conditions that set the

"arrow of time", that is, the temporal sense in which entropy increases. In Ludwig Boltzmann's original H theorem (see Sexl's paper) which showed how entropy increase is enforced in molecular dynamics, it was his assumption of initial lack of correlations (molecular chaos) that set the arrow of time.

As emphasized long ago by David Layzer (Layzer 1971), and more recently by Roger Penrose (Penrose 1979), in the context of cosmology the appropriate boundary condition to impose so that entropy will increase as the universe expands is that the entropy be low initially. This means that the universe must be created in a highly regular and smooth state, one lacking irregularities which would translate into a contribution to the entropy. It is simplest to interpret this initial condition to apply both to the universe's gravitational field (or, what is equivalent, its spacetime geometry), and to its material contents.

The geometry will be smoothest if the universe, initially, is a perfect Friedmann model, by which we only mean homogeneous and isotropic (Penrose 1979). The matter will be smoothest if it is in a perfectly homogeneous quantum state. Statistical fluctuations in the number of particles in a given volume are antagonistic to perfect homogeneity. It follows that the matter quantum state should be <u>a vacuum</u>, one devoid of any particles. In view of recent developments in field theory (Birrell and Davies 1982), we know that a vacuum state in an expanding universe is not necessarily vacuous. For example, associated with it may be a nonzero energy density, a "vacuum energy". The absence of particles, including radiation quanta, from the vacuum justifies regarding it as a zero temperature state.

The proposal that the universe must have started cold apparently goes back to Layzer. It sounds paradoxical in view of the rampant belief,

documented in Secs. 1 and 4, that the early universe passed through a very hot state. Nevertheless, the need for such an initial cold homogeneous quantum state for the matter in the universe is clear. It can be reconciled with the evidence for an early hot era if at some very early epoch the cold smooth low-entropy state could transform itself into a hot state of higher entropy (this must take place at least as early as redhift 10^{10} to preserve the highly succesful picture of helium formatiom in the hot early universe).

This, then, is the prescription for making peace between thermodynamically suitable boundary conditions and the strong evidence for a hot early universe. It sounds farfetched, but has actually been advocated on entirely different grounds in the "inflationary" cosmological model propounded by Alan Guth (Guth 1981) and Andre Linde (Linde 1982) among others. In inflationary cosmology the universe starts with an accurately homogeneous and isotropic gravitational field (a de Sitter geometry, a special case of Friedmann model), and initially contains only a curious field, called Higgs field by particle physicists, in a cold vacuum state. This state of the field is directly responsible for a very early and exponentially rapid (inflationary) expansion of the universe, an expansion which plays an important role in solving several thorny problems of cosmology. What need concern us here is the view of inflationary cosmology that the rapid expansion terminates in conjunction with thermalization of the vacuum state of the Higgs field at a high redshift (about 10^{28}). Matter together with radiation are created at the expense of energy of the Higgs field; the new state is an high-entropy hot state. The cold low-entropy universe thus turns into a hot high-entropy universe as required

by our previous discussion.

A bonus of this scenario is that it provides an elegant genesis for the inhomogeneities required for the formation of large structures in the universe. According to detailed calculations (Bardeen, Steinhardt and Turner 1983) the quantum fluctuations of the Higgs field in its original vacuum state, fluctuations dictated by physical law, are transformed into inhomogeneities of the matter and radiation created upon collapse of the vacuum state. Not only that, but the calculated inhomogeneities have the right distribution by strength. Their strength decreases with increasing scale according to a law, first suggested by Edward Harrison (Harrison 1970) and Zel'dovich (1972), which, according to informed opinion, is most appropriate for the required initial spectrum of inhomogeneities. This is an unexpected and most welcome success of the inflationary cosmology. Still beclouding these endeavors is the fact that the overall strength of the inhomogeneities as predicted is too high, but there are signs that this problem may find a resolution (Hawking 1985).

6. A MEASURE OF ORDER

Thus the inflationary cosmology can explain the passage of the universe from thermodynamically reasonable initial conditions to a state pervaded by hot matter with some inhomogeneity. This state is to serve as raw material for the formation of large structures. A paradox appears at this stage. It is well known that a system in thermodynamic equilibrium, such as the matter at the epoch in question is most likely to be, has attained the maximum value of its entropy. According to the usual information-theoretic interpretation,

if the entropy takes on its maximum value, the information obtainable about the detailed state of the system is nil. One then wonders how the matter in question can eventually evolve into large structures (galaxies, galaxy clusters, superclusters) which are highly organized systems requiring a lot of information for their specification. Put another way, how can the matter in question go from a state of maximal entropy to one considerably below maximal entropy. Does not this violate the second law of thermodynamics?

A routine retort to a query of this kind is that the universe is an open system, so that the second law does not apply to it, and hence there is no paradox. I think this viewpoint obscures the real issue. It is true that if the universe is spatially infinite (and the empirical evidence leans in this direction), then it is not closed in a strict sense. However, another evident feature of the universe is that very distant objects recede from us with velocities very similar to those required by Hubble's picture of a uniformly expanding universe. This means that on a large scale (larger than superclusters) matter transfer between separate locations is negligible. If our position in the universe is not priviledged, we may learn from the low velocity of our galaxy with respect to the thermal background radiation (about 300 km/sec, small compared to relative galaxy velocities in the local supercluster) that all over the universe radiation hardly flows with respect to the matter. Thus there is no important radiation transfer between separate locations. The conclusion is that any large region in the universe, defined by the galaxies it contains rather than by volume, will not exchange much matter or heat with its surroundings. That region is thus nearly closed in a thermodynamic sense.

The law of entropy increase should thus operate for each such large

region in the universe. Since such regions today do contain highly organized structures, we come back to the paradox. How do organized systems arise from matter already at the maximal entropy level without running afoul of the second law? One possible resolution was pointed out long ago: the maximal level of entropy is no set once and for all, but is continually raised by the expansion of the universe (Tolman 1934). After all, in classical thermodynamics a maximal entropy level is determined in the context of particular energy and volume of the system. Raise that energy or expand that volume and you can reasonably expect to raise the maximal entropy. In the expanding universe the volume of a region grows and the energy of its contents decreases as a result of the work it performs on its surroundings (or, in the relativistic viewpoint, as a result of the redshift). It is thus reasonable to expect the maximal entropy to change.

That the maximal entropy is raised is made particularly clear by the well known theorem (Tolman 1934, Layzer 1971) that the expansion removes any system of particles, which are neither non-relativistic nor ultrarelativistic, from thermodynamic equilibrium. Since this equilibrium is a maximal entropy situation for given constraints, the removal from equilibrium can only be accomplished by a raising of the maximal entropy level. The level cannot be lowered because the actual entropy cannot decrease. In fact, because of the departure from equilibrium, the actual entropy is expected to increase somewhat.

As the maximal entropy level is raised, the region in question acquires a potential for information content measured by the difference between the maximal entropy and the actual entropy. Thus the expansion opens up the oportunity for organization in what would otherwise remain formless matter.

According to this viewpoint, galaxies, clusters of galaxies and superclusters are creatures, not just of the ability of gravitation to gather distant matter together, but also of the information generating ability of the Hubble expansion.

I believe that though the resolution just described, which has been offered at various times in the literature (Tolman 1934, Layzer 1971), has elements of the truth, it cannot be the whole truth. Much of the organization at the cosmic level appears after the universal expansion has lost its grip on the matter. For example, it is believed that galaxy disks and spiral structure in them become well developed only after the protogalaxies have become detached from the universal medium and collapsed upon themselves. Can it be claimed that the maximal entropy has grown during the collapse?

Evidently the volume of the system of interest decreases; only if we insist on considering a larger volume, parts of which are still expanding, can an increment of the maximal entropy be claimed on the score of volume. But is it reasonable to encompass the external volume? Yes if there is an agent which binds both volumes intimately. That agent is gravitation.

The bulk kinetic and thermal energies of the collapsing protogalaxy do increase, mostly has a result of the steady decrease of the (negative) gravitational potential energy. This can be taken as a factor promoting the increase in the maximal entropy. In fact, it is precisely the unbounded decrease of the potential energy which is held responsible for the "gravothermal catastrophe" which Newtonian assemblages of masses can undergo (Lynden-Bell and Wood 1968), a catastrophe which is expressed in unbounded growth of the entropy. Thus, the factors for promoting the raising of the

maximal entropy level are present in protogalaxy collapse, but it is important to note that they depend critically on gravitation.

This brings us to suggest that no picture of the growth of organization in sytems ruled by gravitation is complete unless it includes a "gravitational entropy" in its considerations. For it seems unreasonable to rely on gravitation to extend the entropy limits of the matter while denying it any part in the entropy of the system. Of course gravitational entropy was once an unthinkable concept, but with the general acceptance of black hole entropy (Bekenstein 1973, Hawking 1975) as a **bona fide** entropy in gravitational physics, the psychological hurdles have been removed. Another similar entropy may be associated with gravitational systems which do not include black holes, similar in that it is quantified by geometric properties of gravitation, rather than by properties of matter. If so, any argument about the growth of organization in gravitating systems will be incomplete if it fails to include gravitational entropy in the information-theoretic considerations we mentioned.

It was Penrose (1979) who first suggested, on different grounds, a local gravitational entropy quantified by the Weyl tensor, the measure of "wrinkling" of spacetime due to gravitation. To his arguments we may add one directly relevant to our subject. We stressed the importance of an homogeneous initial state for the matter to provide the right boundary conditions for operation of the second law. Now an homogeneous matter state is possible only in a homogeneous and isotropic spacetime, for any irregularities in its geometry would feed back to the matter. The homogeneous isotropic (Friedmann) spacetime has vanishing Weyl tensor. In later epochs the Weyl tensor departs from zero in regions where mass

congregates. The larger the departure from homogeneity, the larger will Weyl's tensor be for a given mass. It is easy to see in the growth evidenced by the Weyl tensor the increase expected of an entropy. It must be stressed, however, that no concrete and accepted formula relating gravitational entropy to Weyl tensor exists as yet, and, therefore, the concept has yet to be put to crucial test.

REFERENCES

Bardeen, J. M., Steinhardt, P. J. and Turner, M. S. 1983, Spontaneous Creation of Almost Scale Free Density Perturbations in an Inflationary Universe, Phys. Rev. D28, 679.

Bekenstein, J. D. 1973, Black Holes and Entropy, Phys. Rev. d7, 2333.

Birrell, N. D. and Davies, P. C. W. 1982, Quantum Fields in Curved Space (Cambridge: Cambridge University Press).

Davis, M. 1980, Lower Limits to Fluctuations in the Microwave Background Radiation Induced By Recombination, Physica Scripta 21, 717.

Doroshkevich, A. G., Sunyaev, R. G. and Zel'dovich, Ya. B. 1974., The Formation of Galaxies in Friedmannian Universes, in Confrontation of Cosmological Theories with Observational Data, M. Longair, Ed. (Dordrecht: Reidel), p. 213.

Einstein, A. 1916, Die Grundlage der allgemeinen Relativitetstheorie (The Foundation of the General Theory of Relativity), Ann. d. Physik 49, 769.

_____1917, Kosmologische Betrachtungen zur allgemeinen Relativitetstheorie (Cosmological Considerations about the General Theory

of Relativity), Sitzungs Berichte d. Preussische Akademie d. Wissenschaften 1, 142.

Fall, M. S and Rees, M. J. 1977, Survival and Disruption of Galactic Substructure, Mon. Not. Roy. Astron. Soc. 181, 37P.

Friedmann, A. 1922, Ubber die Krummung des Raumes (On the Curvature of Space), Zeitschrift f. Physik 10, 377.

_____1924, Ubber die Moglichkeit einer Welt mit Konstanter negativer Krummung des Raumes (On the Possibility of a World with Constant Negative Space Curvature), Zeitschrift f. Physik, 21, 326.

Gott, R. and Rees, M. J. 1975, A Theory of Galaxy Formation and Clustering, Astron. and Astroph. 45, 365.

Gott, R. 1979, Computer Simulation of the Universe, Comm. on Astroph. 8, 55.

Guth A. H. 1981, The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems, Phys. Rev. D23, 347.

Harrison, E. R. 1970, Fluctuations at the Treshold of Classical Cosmology, Phys. Rev. D1, 2726.

Hawking, S. W. 1975, Particle Creation by Black Holes, Commun. Math. Phys. 43, 199.

_____1985, Cambridge University preprint.

Hubble, E. P. 1926, Extragalactic Nebulae, Astroph. Journ. 64, 321.

_____1929, A Relation Between Distance and Radial Velocity Among Extra-Galactic Nebulae, Proc. Nat. Acad. Sci. (Wash.) 15, 168.

Huchra, J., Davis, M., Latham, D. W. and Tonry, J. 1983, A Survey of Galaxy Redshifts: IV. The Data, Astroph. Journ. Suppl. 52, 89.

Jeans, J. H. 1902, On the Stability of Spherical Nebulae, Phil. Trans. Roy.

Soc. London 199, 1.

_____1929, Astronomy and Cosmogony (Cambridge: Cambridge University Press).

Kirshner, R. P., Oemler, A. Schechter, P. L. and Shectman, S. A. 1981, A Million Cubic Megaparsec Void in Bootes?, Astroph. Journ. 248, L57.

Koyre, A. 1958, From the Closed World to the Infinite Universe (New York: Harper and Row), p. 185.

Layzer, D. 1971, Cosmogonic Processes, in Astrophysics and General Relativity, M. Chretien, S. Deser and J. Goldstein, Eds. (New York: Gordon and Breach), Vol 2, p. 155.

Lifhitz, E. M. 1946, On the Gravitational Stability of the Expanding Universe, Zhurn. EKsp. Teor. Fiz. 16, 187 [Journ. of Phys. JETP 10, 116].

Lin, C. and Shu, F. 1964, On the Spiral Structure of DisK Galaxies, Astroph. Journ. 140, 646.

Lindblad, B. 1927, The Small Oscillations of a Rotating Stellar System and the Development of Spiral Arms, Medd. Astron. Obs. Uppsala 19, 1.

Linde, A. D. 1982, A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Isotropy and Primordial Monopole Problems, Phys. Letters 108B, 389.

Lynden-Bell D. and Wood R. 1968, The Gravo-thermal Catastrophe in Isothermal Spheres and the Onset of Red-Giant Structure for Stellar Sysyems, Mon. Not. Roy. Atronom. Soc. 138, 495.

Misner, C. W. 1968, The Isotropy of the Universe, Astroph. Journ. 151, 431.

Oort, J. 1983, Superclusters, Ann. Rev. Astron. astroph. 21, 373.

Osmer, P. 1983, Quasars and Superclusters, in Early Evolution of the Universe and its Present Structure, G. O. Abell and G. Chincarini, Eds. (Dordrecht: Reidel), p. 189.

Ostriker, J. P. and Cowie, L. L. 1981, Galaxy Formation in an Intergalactic Medium Dominated by Explosions, Astroph. Journ. 243, L127.

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

Penrose, R. 1979, Singularities and Time-Asymmetry, in General Relativity - An Einstein Centennary Survey, S. W. Hawking and W. Israel, Eds. (Cambridge: Cambridge University Press), p.581.

Penzias, A. A. and Wilson, R. W. 1965, A Measurement of Excess Antenna Temperature at 4080 MHz, Astroph. Journ. 142, 419.

Schwarz, J., Ostriker, J. P. and Yahil, A. 1975, Explosive Events in the Early Universe, Astroph. Journ. 202, 1.

Silk, J. 1968, Cosmic Black-Body Radiation and Galaxy Formation, Astroph. Journ. 151, 459.

Tolman, R. C. 1934, Relativity, Thermodynamics and Cosmology (London: Oxford University Press).

Toomre, A. 1977, Theories of Spiral Structure, Ann. Rev. Astron. Astroph. 15, 437.

Uson J. M. and Wilkinson, D. T. 1984, Small Scale Isotropy of the Cosmic Microwave Background at 19.5 GHz., Astroph. Journ. 283, 471.

Vaucouleurs, G. de 1953, Astron. Journ. 58, 30.

Weinberg, S. 1972, Gravitation and Cosmology (New York: Wiley), p.408.

White, S. D. M., Davis, M. And Frenk, C. S. 1984, The Size of Clusters in a Neutrino Dominated Universe, Mon. Not. Roy. Astron. Soc. 209, 27P.

Zel'dovich, Ya. 1972, A Hypothesis, Unifying the Structure and Entropy of the Universe, Mon. Not. Roy. Astron. Soc. 160, 1P.

Zwicky, F. 1933, Die Rotverschiebung von Ekstragalaktische Nebeln (The Redhift of Extragalactic Nebulae), Helvetica Physica Acta 6, 110.

FIGURE CAPTIONS

Fig.1 NGC 2903, distant some 15 million light years from us, is a disk galaxy with well developed spiral structure. Reproduced with permission from a plate exposed by Dr. A. Meisels with the 1 meter telescope of Wise Observatory, Israel.

Fig.2 M 94, a spiral galaxy dominated by its spheroidal component, is a member of the nearby Canes Venatici I cloud of galaxies. From a plate by the author exposed at Wise Observatory.

Fig.3 M 65 (upper) and NGC 3627 are two spiral members of the small M 66 group of galaxies. Note the extensive spheroidal component of M 65 and the disturbed shape of NGC 3627. Reproduced from a Wise Observatory plate by Dr. A. Meisels with his permission.

Fig.4 NGC 4657 (left lower corner), an example of an irregular galaxy, and NGC 4631 (center), a spiral galaxy seen edge-on, are members of the Canes Venatici II cloud of galaxies. Also visible is NGC 4627, a small satellite elliptical galaxy, visible as a small roundish smudge just above NGC 4631. The streak through the photograph is the trail of a meteor that traversed the telescope field in the course of the 3 hour exposure of the plate by O. Lahav at Wise Observatory. Reproduced with his permission.

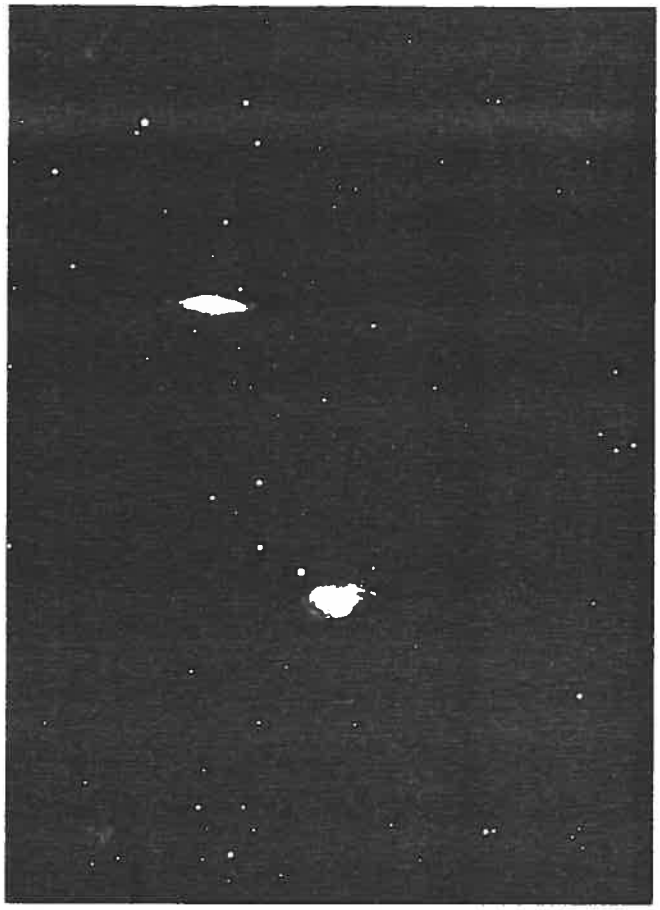
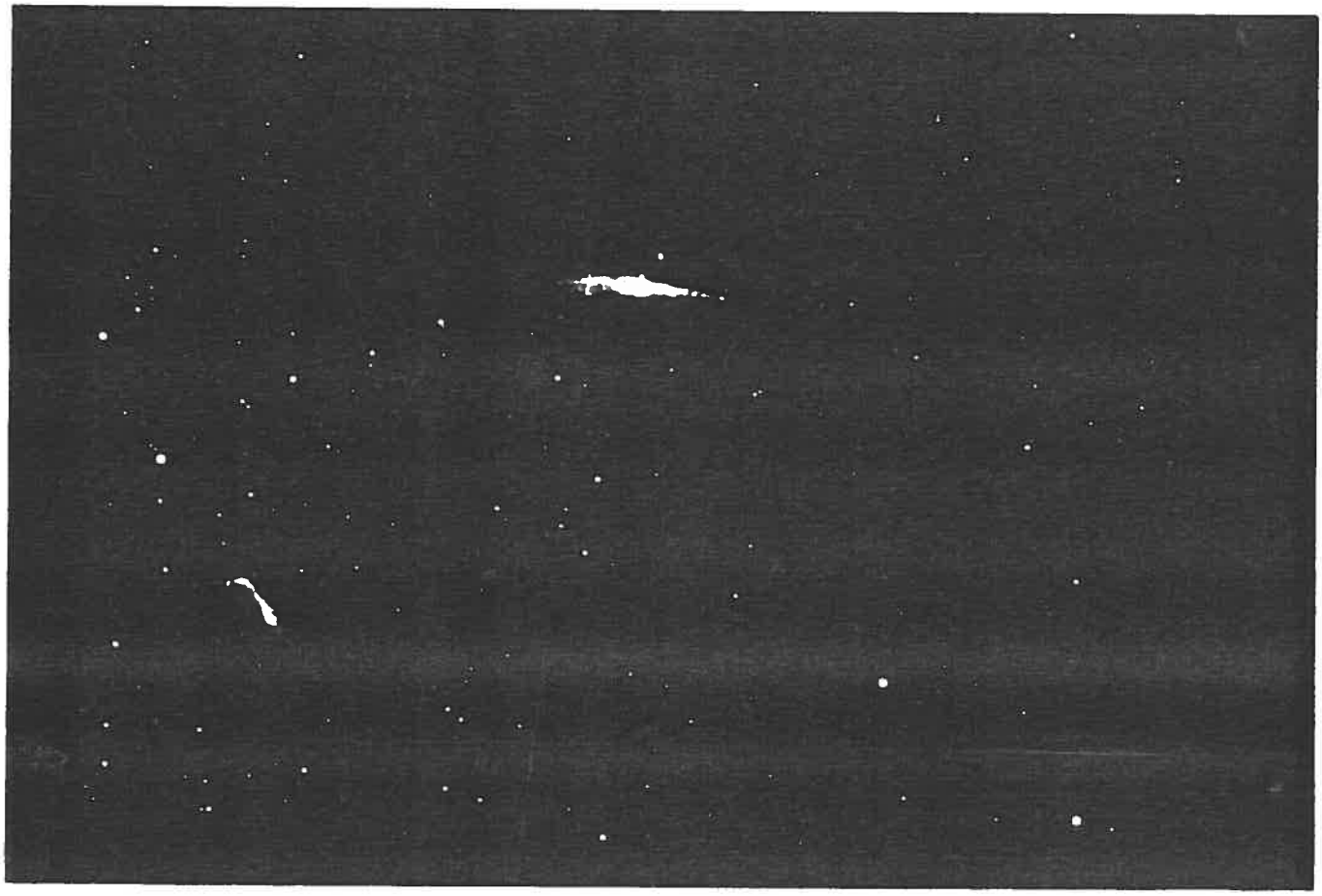


Fig. 3

Fig. 4



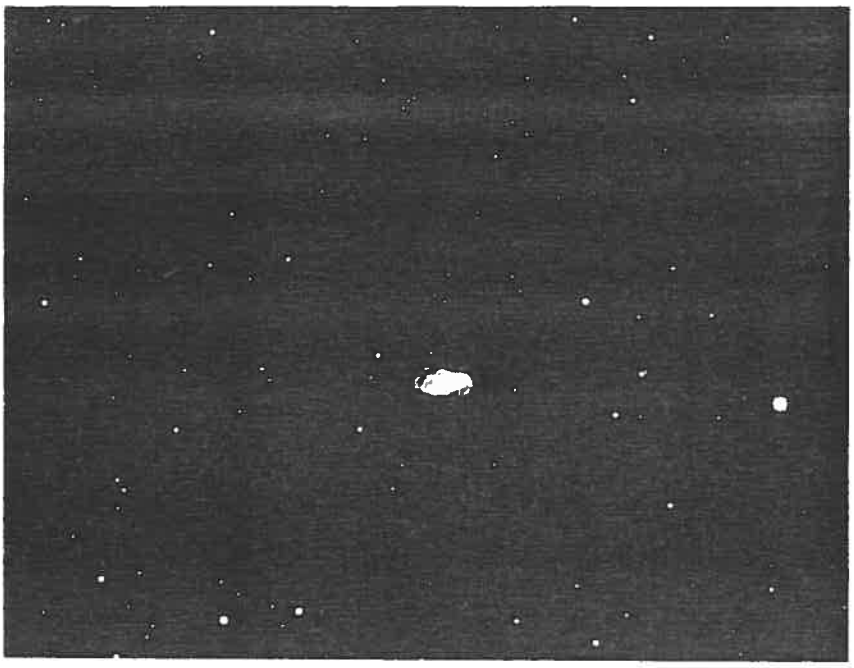


Fig. 1

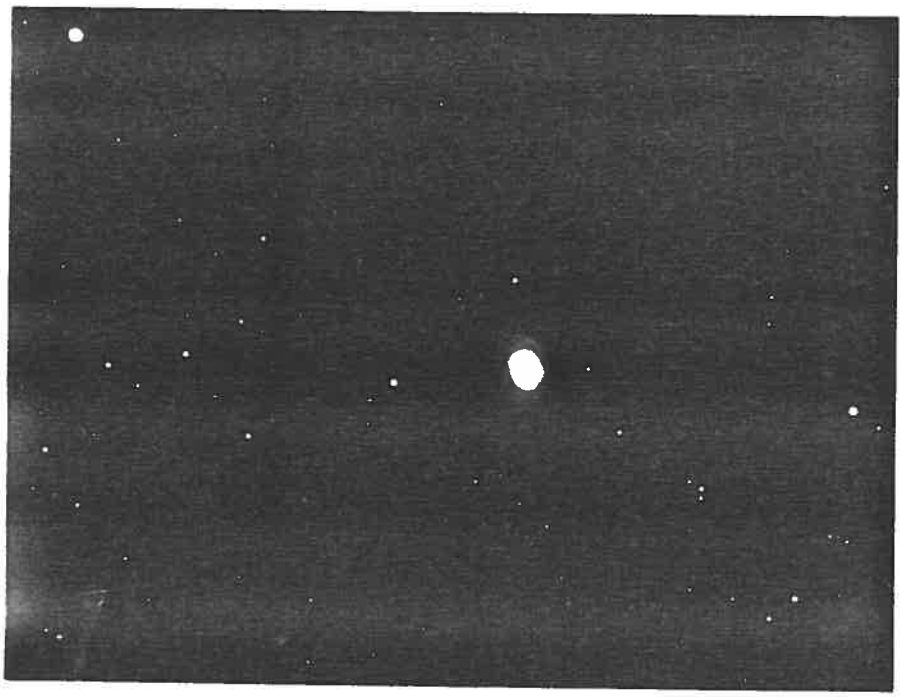


Fig. 2