THE FORMATION OF GALAXIES

Richard B. Larson

1. INTRODUCTION

Many properties of galaxies cannot be understood without some knowledge of the way in which they were formed. Such gross structural properties as the radial distribution of light in elliptical galaxies and the disc-to-bulge ratios of spiral galaxies, for example, cannot have changed greatly since their time of formation, since the time scale for stellar dynamical relaxation processes in galaxies is longer than the age of the universe. Also, while tidal interactions may modify the structure of some galaxies, they can hardly explain all of their observed properties. A number of similarities and differences exist among the structural properties of galaxies; for example, most elliptical galaxies have very similar light profiles, but their ellipticity profiles may be very different. Understanding the significance of these results requires some theoretical understanding of the processes by which galaxies are formed. In addition, there are correlations between the morphological types of galaxies and the surroundings in which they are located; for example, elliptical galaxies are preferentially found in dense clusters, while spirals are mostly found in the field or in loose groups. These differences must reflect differences in the way in which spiral and elliptical galaxies are formed, and if they could be interpreted in terms of a picture for galaxy formation, they might provide important information about the processes and conditions that are relevant for the formation of different types of galaxies.

It is also important to understand the chemical composition of galaxies and the variation of chemical abundances within and between galaxies. For example, elliptical galaxies have radial abundance gradients which must be explained by any theory of the formation process. Also, the overall metal abundances of elliptical galaxies are correlated with their masses, and this must reflect differences in the way in which small and large galaxies are formed. In our own galaxy, the variations in chemical composition among the different stellar populations are evidently closely related to the way in which the galaxy formed; indeed, there is now an extensive literature attempting to interpret observed abundances in terms of models for the formation and evolution of our galaxy.

A further reason for wishing to understand how galaxies form is the current interest in searching for "primeval galaxies", i.e. galaxies which are still forming or are in very early stages of evolution. Searches for primeval galaxies and attempts to measure the cosmic background light bear closely on problems of the formation and early evolution of galaxies, and will eventually provide important constraints on theoretical models. Also, it is possible that some peculiar galaxies are young or have recently been rejuvenated by the addition of fresh material; if so, these objects will provide important observational information about how galaxies form. Perhaps the most interesting possibility is that some of the more spectacular peculiarities, such as quasar activity, may be closely related to gas infall processes and the formation of dense nuclei in galaxies. It has long been suspected that quasars are somehow related to the formation and early evolution of galaxies, and it would be very important to clarify the nature of such a connection.

The Origin of Galaxies

The origin of galaxies is closely related to cosmological problems of the structure and evolution of the universe at early times. At present, some form of "big bang" cosmology is almost universally accepted, and it is believed that galaxies formed at some stage in the expansion of the universe by the recollapse of the densest parts of an inhomogeneous pregalactic medium. These dense regions or "protogalaxies" may have originated through the growth of small density fluctuations which were present from a very early stage of the expansion, as in the "gravitational instability" picture of galaxy formation, or they may have been generated by turbulence or other processes at a later stage of the expansion. Both gravitational instability and turbulence theories of the origin of galaxies have been extensively discussed (see Jones 1976), but unfortunately only very limited predictions are as yet offered by these theories, and it is not even clear that they can explain such basic facts as the masses of typical galaxies. In any case it may be difficult to relate present-day galaxies to primordial density fluctuations, since it is not known how much coalescence or disruption of the initial density fluctuations may have taken place prior to the appearance of the galaxies which we observe. At present it is probably better for modeling purposes to separate the later stages of the galaxy formation process from the little understood early evolution of primordial density fluctuations, and postulate as an initial condition the existence of protogalaxies with masses and angular momenta comparable to those of the presently observed galaxies.

Although some form of gravitational collapse or condensation picture seems the most promising way of explaining in detail the many observed properties of galaxies, gravitational theories of galaxy formation can take a number of forms (Larson 1975b). The simplest idealization,which has formed the basis of most model calculations, is that a protogalaxy begins as a discrete spherical cloud of gas (and/or stars) with uniform density and uniform initial rotation. It is likely, however, that this is a considerable oversimplification of the real situation, and that protogalaxies are less distinct from their surroundings and quite inhomogeneous in structure. For example, it is possible that a galaxy begins as a collection of smaller subsystems which later coalesce into a single larger system; the subsystems could conceivably range in size from globular clusters to two smaller galaxies that collide and merge into one. Large galaxies may grow from smaller beginnings by accreting surrounding matter, either in the form of primordial uncondensed gas or as already formed neighboring galaxies. These various possible detailed models of galaxy formation all share the basic features that the material forming a galaxy is collected together by gravity and that some form of dissipation is required to make the material stick together in a relatively compact system. Thus it should be possible to study some of the basic features of the formation process even with relatively idealized collapse models, although the possibility of continuing accretion of surrounding matter for a time long compared with the initial collapse time may be an additional important effect. Accordingly, we shall in these lectures be concerned mainly with the evolution of idealized discrete model protogalaxies with well-defined sizes and masses, which can be treated in isolation from the rest of the universe.

Given the concept of a discrete isolated protogalaxy, some crude constraints can be set on its characteristic size and collapse time. If galaxies form by a collapse process, the radius of a protogalaxy must be at least as large as the presently observed radius of a typical galaxy, i.e. $\gtrsim$ 30 - 50 kpc. For a system of mass 10^{11} $M_{\odot}$, a radius of 30 kpc implies a free-fall time of 3 x 10^{8} yr, and since it takes at least one free-fall time for a protogalaxy to expand to its maximum radius before beginning to collapse, a lower limit on the time of collapse is at least $\sim$ 6 x 10^{8} yr after the big bang. It is more difficult to set an upper limit on the collapse time, but if we assume that most galaxies collapsed at redshifts of 3 or more and are therefore at least 80 per cent as old as the universe, the corresponding maximum radius for a protogalaxy of mass 10^{11} $M_\odot$ is $\sim$ 100 kpc, and the collapse occurs within $\sim 4 \times 10^9$ yr after the big bang. There is no a priori reason why the collapse time could not be even longer than this and why galaxies could not continue to form at the present time; an interesting possibility to be discussed later is that some peculiar galaxies may represent galaxies which are still forming or are in very early stages of evolution.

Different authors have made different assumptions about the initial state of the matter in a protogalaxy before it begins to collapse. In some models it is assumed that a protogalaxy initially consists only of gas or gas clouds, and that stars form from the gas as the system collapses. At the other extreme, it has been assumed in some calculations that the gas is early turned into stars, either in a smooth distribution or in discrete smaller galaxies, before the system collapses. The latter models are the simplest in that they consider only stars and stellar dynamics, whereas the models that begin with gas must consider the complexities of gas dynamics and star formation as well. Thus it will be convenient to discuss separately the processes of stellar dynamics and those of gas dynamics which are relevant for collapsing protogalaxies. Also, since the process of star formation is central to an understanding of the formation and evolution of galaxies, we shall devote some attention to the possible mechanisms of star formation which may be relevant for the formation and evolution of galaxies.

2. DYNAMICS OF PROTOGALAXIES

(a) STELLAR DYNAMICS

An extensive literature exists on the dynamics of star clusters, and most of the dynamical processes relevant to the evolution of star clusters are by now well understood (see, for example, IAU Symposium No. 69, Dynamics of Stellar Systems, ed. A. Hayli). Because of the relative elegance and simplicity of stellar dynamics as compared with most of astrophysics, and because an understanding of the formation and dynamics of galaxies must depend at least in part on stellar dynamics, one approach has been to attempt to understand the structure of galaxies using as far as possible only the concepts of stellar dynamics. This approach might be applicable for the elliptical galaxies, most of which contain no detectable gas and can thus be regarded as purely stellar systems.

Two-Body Relaxation

Most ordinary star clusters have probably experienced considerable dynamical evolution caused by gravitational encounters between the stars, which tend to produce relaxation toward a Maxwellian velocity distribution. This relaxation causes the core of the cluster to contract and the envelope to expand as the stars with the highest velocities escape from the core and build up an extended envelope, or even escape from the cluster. The centrally condensed and symmetrical structure of many star clusters is probably largely a result of such processes, and various theoretical models incorporating relaxation effects have claimed to provide satisfactory representations of the structure of star clusters. Some authors (eg. King 1966) have noted that the same models which satisfactorily represent the structure of globular clusters are also able, with somewhat different choices of parameters, to represent the structure of elliptical galaxies; thus it has been said that elliptical galaxies "look relaxed", and efforts have been made to find a basis for applying relaxation models to elliptical galaxies.

The time scale required for relaxation to significantly alter the orbits of stars is given approximately by

$$T_r \sim 2 \times 10^8 \frac{V^3}{m^2 N} \text{ yr} \qquad (1)$$

(eg. Hénon 1973), where V is the mean stellar velocity in km s^{-1}, N is the number of stars per pc^3, and m is the average stellar mass in $M_\odot$. When we plug in typical values for an elliptical galaxy, such as $V \sim 200$, $N \sim 10^{-2} - 10^{+1}$, $m \sim 1$, we find relaxation times of $T_r \sim 2 \times 10^{14} - 2 \times 10^{17}$ yr, much too long for ordinary two-body relaxation effects to be of any importance, except possibly in extremely dense nuclei. Thus star cluster models cannot be directly applied to elliptical galaxies. However, this does not mean that relaxation concepts and stellar dynamics have no relevance for the formation of galaxies. We note from equation (1) that the relaxation time depends inversely on the mass m of the interacting objects, so it is possible that relaxation could be important if, at an early stage of evolution, E galaxies are made up of smaller systems much more massive than ordinary stars, such as massive star clusters, gas clouds, or small galaxies which later break up and merge into a single galaxy. In systems containing $\lesssim$ 100 particles the relaxation time is approximately equal to the orbital crossing time, so that significant two-body relaxation would be expected to occur even within one orbital time scale. If the subsystems all have about the same mass, the time required for the formation of a pronounced core-halo structure is $\sim$ 10-20 relaxation times, but it is problematical whether the subsystems could survive for this long. However, if there is a large spread in mass among the interacting objects, as there is in star clusters with a realistic stellar mass spectrum, the evolution is much faster and the most massive objects rapidly form a dense core after only $\sim$ 1-2 relaxation times. In effect, the massive objects experience a strong "dynamical friction" (see below) due to the background of less massive objects, and they fall rapidly to the center. Some situations in which dynamical friction may play a role in the formation and evolution of galaxies are considered further below.

Violent Relaxation

A second possibility for explaining the structure of E galaxies using only stellar dynamics is provided by the concept of "violent relaxation" (Lynden-Bell 1967). If the stellar orbits are controlled mainly by the mean gravitational field of the whole system, and if the mean field fluctuates sufficiently rapidly with time, as might occur during an initial violent collapse of the system, then the energies of individual stars are not conserved, and the resulting reshuffling of stellar energies may simulate a relaxation process and produce an approach to a most probable equilibrium distribution. Lynden-Bell worked out the statistical mechanics of this problem and concluded that the resulting equilibrium system would be basically an isothermal sphere with a Maxwellian velocity distribution modified by a cutoff at the escape velocity similar to that in the star cluster models of King (1966). Thus Lynden-Bell's violent relaxation theory might provide some justification for the application of King's cluster models to elliptical galaxies, and also for the generalized King models with rotation that were studied by Wilson (1975) and shown to represent well some of the basic structural properties of E galaxies.

Unfortunately the applicability of violent relaxation theory is somewhat unclear, in view of the fact that various numerical experiments designed to test the theory have shown only partial agreement with it; in all of these experiments, significant numbers of particles are thrown into high energy orbits for which relaxation is ineffective and the experimental energy distributions disagree with the predicted ones. This suggests that violent relaxation theory may not apply to the halos of E galaxies.

A more direct application of the violent relaxation concept to galaxy models has been provided by Gott's(1973,1975) calculations of the collapse of rotating axisymmetric systems of stars, the results of which will be discussed in more detail later. In these models, a galaxy is represented by a system of 2000 discrete mass rings which are initially distributed in a uniform sphere with uniform rotation and small random velocities, and are then allowed to collapse under the gravity of the system. The orbits of all the rings are calculated individually and the evolution of the system is followed through an initial collapse and several oscillation periods until it has reached a steady equilibrium state. The influence of violent relaxation is seen in a broadening of the energy distribution, and this leads to a centrally condensed final configuration in which the low energy stars form a dense core and the high energy stars form an extended halo. Thus the rapid collapse of a system of stars can produce a centrally condensed system qualitatively resembling an elliptical galaxy, but as will be seen below, the quantitative agreement is poor unless additional effects like "infall" are considered.

It is conceivable that the symmetries built into the Gott calculations make violent relaxation less effective than it might be in a less idealized situation. Another model for the formation of E galaxies which involves stellar dynamics and violent relaxation effects is the collision model of Toomre (1974,1976), in which two disc galaxies collide and violently disrupt each other and eventually merge into a single elliptical-like system. This model yields a density distribution that more closely resembles real E galaxies in structure, although the observed degree of central concentration is still not reproduced.

Dynamical Friction

As a star or other massive body moves through a background of other stars, its gravitational field tends to concentrate the background stars into a "gravitational wake" behind it, whose gravitational pull exerts a decelerating force on it. The time required for this "dynamical friction" to significantly decelerate a massive object of mass M >> m moving with a typical velocity V is given by

$$T_d \sim \frac{2m}{M} T_r \sim 4 \times 10^8 \frac{V^3}{mMN} \text{ yr} \qquad (2)$$

(eg. Hénon 1973); since this is shorter than the relaxation time T_r by approximately a factor m/M, T_d can be interestingly short if M is large. With the parameters assumed above to estimate T_r in an elliptical galaxy, a mass $M \gtrsim 4 \times 10^4$ - $4 \times 10^7 M_\odot$ is required to make T_d less than 10^{10} years. This suggests that dynamical friction may be important for massive globular clusters and may cause them to spiral in toward the nucleus of a galaxy within 10^{10} years; Tremaine et al. (1975) and Tremaine (1976) have suggested that this mechanism may be important in explaining the formation of small sharp nuclei in galaxies. (Against this suggestion, however, we note that the high metal abundances of galactic nuclei do not seem consistent with formation from globular clusters. Other observational arguments against this mechanism for the formation of galactic nuclei have been given by van den Bergh (1976a).)

Since the structure of elliptical galaxies is probably established within only a few free-fall times of the beginning of the collapse (or coalescence) process, dynamical friction can be important for the structure of E galaxies if its effects become significant within a time interval of, say, 5×10^8 yr. This could be the case if proto-elliptical galaxies contain massive condensations or subsystems with masses $M \gtrsim 10^6 - 10^9 M_\odot$, which is not implausible if elliptical galaxies form by the coalescence of smaller systems. These hypothetical smaller systems could have been either massive star clusters, small galaxies, or massive gas clouds in the process of collapsing into stars. However, in view of our almost complete lack of knowledge about the detailed structure of proto-galaxies, it is difficult to assess the likely importance of relaxation or dynamical friction effects for the formation of elliptical galaxies.

Dynamical friction may also play a role in causing some galaxies to grow by the accretion of nearby companion galaxies. If two galaxies interpenetrate during a close passage, dynamical friction will reduce the orbital energy of the colliding pair, and if enough energy is dissipated in this way into random stellar motions, the two galaxies may spiral together and merge into a single galaxy. Some large galaxies in groups and clusters, particularly those in a dominating central position, may thus continue to grow slowly with time by accreting other galaxies which pass nearby. Also, if much of the mass in large clusters like the Coma cluster is in extended galactic halos or intergalactic stars, dynamical friction will tend to make the galaxies spiral in toward the center. The growth in mass and luminosity of the central galaxy may be great enough to affect cosmological tests (Ostriker & Tremaine 1975), and accretion may also noticeably affect the structure of the central galaxy, possibly accounting for the phenomenon of "cD" galaxies with extremely extended envelopes (Richstone 1976). However, calculations which are sufficiently reliable to predict the quantitative details of such processes are still generally lacking, except for a calculation by Toomre (1974, 1976) of the special case of two equal disc galaxies colliding head-on and merging into a single system resembling an elliptical galaxy. Further numerical calculations of dynamical friction and accretion process will be an important area for continuing research in stellar dynamics.

(b) GAS DYNAMICS

For several reasons, it is likely that gas dynamics plays an important role in the collapse of protogalaxies. It seems very unlikely that star formation processes can be so efficient during the earliest stages of galaxy formation that the gas is all turned into stars before protogalaxies begin to collapse; some primordial gas must almost certainly remain and play a role in the dynamics of the collapse. Even if the primordial gas were completely exhausted, various stellar mass loss processes would provide a continuing supply of gas to the system, and the fate of this gas must be considered. In a small galaxy with small binding energy much of the gas may be lost at an early stage of evolution, but in a massive galaxy the gas may be retained and may sink toward the center to form later generations of stars in a highly condensed, metal rich nucleus, as outlined schematically by Spitzer (1971). Indeed, a picture involving infall and simultaneous metal enrichment of residual gas seems the most promising way to account for the metal abundance gradients that are present in most galaxies; a galaxy collapse model involving only stellar dynamics offers no natural way to explain such gradients.

While both stellar dynamics and gas dynamics may play a role in the formation of spheroidal stellar systems, it is clear that the formation of highly flattened disc systems requires strong dissipation, and must therefore involve gas rather than stars. To the extent that there is continuity between disc systems and less flattened spheroidal systems, as appears to be the case among the stellar populations in our galaxy, it seems probable that gas dynamics has been important for the less flattened spheroidal systems as well.

Physical State of the Protogalactic Gas

It is relevant to consider first the physical state of the protogalactic gas, for example whether it is smooth and quiescent or clumpy and turbulent. A number of considerations suggest that the protogalactic gas in fact contains large inhomogeneities and large internal motions. The separation out of protogalaxies from the expanding pre-galactic medium cannot have been a completely smooth and gentle process, because protogalaxies have somehow experienced sufficiently strong perturbing effects to produce the large angular momenta which galaxies are observed to have. Whatever the origin of this angular momentum - tidal torques, pre-galactic turbulence, or coalescence of smaller systems with large random motions - it seems unlikely that it would leave a protogalaxy in a very homogeneous and quiescent state. Any initial inhomogeneities are likely to be amplified during the collapse of a protogalaxy, both because the denser parts will tend to collapse first and because the internal motions are likely to be supersonic and generate shock waves which heat and compress the gas.

If cooling is not effective, it is possible that shock heating could heat much of the gas to a virial temperature and create a hot, thermally supported gas distribution. However, this situation can persist only if the radiative cooling time for such a hot gas cloud exceeds the free-fall time. Consider for example a uniform sphere of gas with mass 10^{11} $M_\odot$ and radius 50 kpc, for which the mean particle density is 6×10^{-3} cm^{-3} and the virial temperature is about 2×10^5 °K; under these conditions, the cooling time (Cox & Tucker 1969) is about 6×10^5 yr for a "cosmic" abundance of heavy elements and 3×10^7 yr with no heavy elements, in which case helium is the major coolant. Only if both the mass and the radius of the protocloud are substantially larger does the cooling time become longer than the free-fall time. Thus in most circumstances cooling will probably be rapid compared with the free-fall time, and a hot thermally supported protogalactic gas cloud is not possible.

If the temperature of the protogalactic gas is much less than the virial temperature, all of the available calculations for collapsing clouds suggest that the protogalaxy will soon begin to break up into clumps with about the Jeans mass. The Jeans mass depends sensitively on the temperature, being given approximately by

$$M_J \sim 1.9\,\rho^{-1/2}\,(\mathcal{R}T/G)^{3/2} \qquad (3)$$

(Larson 1974a). Since the cooling rate is high for $T > 10^4$ °K but drops sharply for $T < 10^4$ °K, hot gas cools rapidly to $\sim 10^4$ °K and then remains near this temperature for a relatively long time before cooling further; likewise, if any steady heat sources are present, a wide range of heating rates will produce an equilibrium temperature near 10^4 °K. If we therefore assume that most of the protogalactic gas has a temperature near 10^4 °K, the Jeans mass at protogalactic densities is in the range $\approx 10^7 - 10^9\ M_\odot$, i.e. much smaller than the mass of a typical galaxy, so that protogalaxies may contain many Jeans-mass condensations. Again this suggests that protogalaxies are quite inhomogeneous and clumpy- although we note that since the Jeans mass is only a lower limit for the size of a gravitationally bound concentration of matter, the protogalactic gas could well possess structure on all scales from the Jeans mass up to the size of the galaxy itself.

If the density distribution and the gravitational field of a protogalaxy possess large irregularities, these will generate internal random motions in the protogalaxy, even if none were present to start with. Also, processes such as supernova explosions or winds from hot stars may generate or help sustain turbulent motions in the protogalactic gas. Observationally, large random motions of the protogalactic gas seem required to account for the large velocity dispersions of the halo stars and globular clusters in our galaxy, many of which even have retrograde orbits (Oort 1965).

A simple model which has often been used to represent the dynamics of a gas with large inhomogeneities and random motions assumes that the gas is all distributed in discrete spherical clouds with independent random motions; the clouds are assumed to collide inelastically, thus dissipating the kinetic energy of the random motions. Models of this type for gaseous collapsing protogalaxies were studied by Larson (1969) and Brosche (1970), both of whom assumed arbitrarily that the clouds occupy about one-tenth of the volume (and hence have ten times the mean density). The dissipation rate then depends on the mass assumed for the clouds. Larson (1969) assumed that the cloud mass is equal to the Jeans mass for a temperature of 10^4 °K; this implies that the cloud mass is initially of the order of 10^9 $M_\odot$ and decreases as the system collapses and the density rises. On the other hand, Brosche (1970) estimated that a protogalaxy typically contains $\sim$ 10 clouds, if it is assumed that the angular momentum of a galaxy is the sum of the purely random angular momenta of it constituent clouds. The resulting dissipation time scales do not differ greatly in the two cases, being of the same order as the free-fall time of the system. A somewhat similar model with inelastically colliding particles has been used by Brahic (1975) to study the formation of flattened disc systems. In all cases, the presence of dissipation plays an important role in allowing at least part of the system to become much more condensed or more flattened than would be possible for a purely stellar system.

Dynamics of a System of Gas Clouds

Two different approaches have been tried for calculating the dynamics of a system of randomly moving gas clouds. The one which is most free from assumptions and approximations involves calculating individually the orbits of all of the gas clouds, as in n-body simulations of the dynamics of star clusters, but with the important difference that the gas clouds have finite cross sections and collide inelastically with each other. Because of the large computing effort required, this approach has not yet been applied in a realistic way to models of collapsing proto-galaxies, but Brahic (1975) has made calculations of this type for the special case of a system of massless test particles orbiting around a central point mass. These calculations are promising in that they show the formation of flat disc systems like the discs of spiral galaxies, but the results are also sensitive to the way in which the collisions are treated; the collisions must be very inelastic if a thin disc is ever to be formed. Further work on models of this type will clearly be valuable.

The other approach, which has been developed in some detail, is to treat the system of gas clouds hydrodynamically as a fluid whose constituent particles are the clouds. The system is then described by fluid variables such as the mean density, mean velocity, and velocity dispersion, whose time dependences are determined by moment equations derived from the Boltzmann equation. This approach has been applied to spherical systems by Larson (1969, 1974b) and to axisymmetric rotating systems by Larson (1975c, 1976). The hydrodynamical equations governing the time dependence of the mean density, mean velocity, and velocity dispersion are given in detail in these papers, and will not be repeated here. The main assumption made in deriving these equations is that the velocity distribution of the clouds is symmetrical about the mean velocity and has no third-order moments, so that the system of moment equations can be closed at the second order. The existence of a non-zero third-order moment would imply a non-zero energy flux (i.e., a heat flow) carried by the random motions; however, while this effect may be important for the long term evolution of systems dominated by two-body relaxation (Larson 1970), it is probably not very important in collapsing protogalaxies. A more serious possible deficiency of the fluid-dynamical treatment is the assumption that the velocities of the gas clouds are random and uncorrelated, and that there are enough clouds in each volume element to define meaningful values for the various moments of the velocity distribution (mean density, mean velocity, and velocity dispersion). In practice these assumptions are difficult to justify, and it is possible that larger scale structures and correlations may exist on all scales up to that of the galaxy itself, in which case hydrodynamic quantities become difficult to define because their values depend on the size of the region over which averages are taken. If this is the case, hydro-dynamical models can provide only a crude qualitative description of the system, but this may in any case be all that is warranted by our still very limited understanding of the detailed processes of gas dynamics and star formation involved.

In a hydrodynamical model, the rate at which the kinetic energy of random motions is dissipated appears as a "cooling" rate in the equations, and it is straightforwardly calculated from the properties of the colliding clouds. In the model of Larson (1969), for example, the dissipation time is given by

$$\tau_{dis} = \frac{\alpha_g}{d\alpha_g/dt} \sim 450\, \rho_g^{-1/2}\, \alpha_g^{-1/2} \qquad (4)$$

where ρ_g is the mean gas density and α_g is the mean square random velocity in one coordinate direction. In this equation the units of mass, length, and time are 1 $M_\odot$, 1 pc, and 10^6 yr respectively. Note that τ_{dis} depends on the velocity dispersion as well as on the density of the gas, and it is equal to the free-fall time $\tau_{ff} = 8.1\ \rho_g^{-1/2}$ for a velocity dispersion of $\alpha_g^{1/2} \sim$ 55 km s^{-1}. Because it is difficult to justify convincingly the details of this particular colliding cloud model, an alternative assumption has also been considered (Larson 1974b), which is that the dominant scale of the gas motions in a protogalaxy is comparable with the size of the system, so that the dissipation time scale is comparable to the overall dynamical or free-fall time scale; more specifically, it has been assumed that

$$\tau_{dis} = D\ \tau_{ff}\ , \tag{5}$$

where

$$\tau_{ff} = 8.1\ \bar{\rho}^{-1/2} \tag{6}$$

is the free-fall time for the average density $\bar{\rho}$ of matter interior to the point considered, and D is a constant of order unity for which values in the range $1 \lesssim D \lesssim 2$ have been considered. With the assumed values of the parameters, equations (4) and (5) predict comparable values for τ_{dis} during the early stages of the collapse, but the value of τ_{dis} given by equation (4) increases strongly during later stages as ρ_g decreases, while the value given by equation (5) remains nearly constant.

In addition to the dissipation effect discussed above, a second important property of a system of randomly moving gas clouds is a large viscosity. Like the molecules in a classical gas, the clouds transport momentum (or angular momentum) by their random motions, and in general there is a net flux of momentum or angular momentum from one part of the system to another. In a rotating protogalaxy, this "turbulent viscosity" generally transfers angular momentum outward, and this allows the central part of the system to become highly condensed without becoming highly flattened (Larson 1975c). Physically, this can be understood in terms of the effect pointed out by McCrea (1960) in connection with his floccule theory of star formation: because of their random motions the clouds possess a wide range of angular momenta, and the clouds with little angular momentum will tend to fall toward the center and concentrate there, while the clouds with high angular momenta remain in the outer parts of the system; thus there is a segregation of low and high angular momentum material, leading to an effective outward transport of angular momentum.

Two approaches to treating the turbulent viscosity in a hydrodynamic model have been considered (Larson 1975c). If the number of fluid particles (gas clouds) is large and their mean free paths are small compared with the size of the system, the effect of viscosity is given by the classical Navier-Stokes equations with a kinematic viscosity coefficient

$$\nu = \frac{1}{3} \overline{v} \lambda \qquad (7)$$

where $\overline{v}$ is the average random velocity and λ is the mean free path of the particles. For the cloud model considered above, the kinematic viscosity is

$$\nu \sim 260 \rho_g^{-1/2} \alpha_g^{1/2} . \qquad (8)$$

The time scale for viscous transfer of angular momentum is given approximately by

$$\tau_{vis} \sim \frac{r^2}{\nu} \sim 4 \times 10^{-3} \, r^2 \, \rho_g^{\,1/2} \, \alpha_g^{\,-1/2} \qquad (9)$$

and is typically of the same order as the free-fall time τ_{ff} during the early stages of the collapse. The Navier-Stokes equations and equations (7) - (9) are strictly applicable only if the cloud mean free path λ is much smaller than the scale length r of the system, but in fact λ and r are typically of the same order of magnitude, so this approach is only approximately valid. However, it may be compatible with the other approximations made in the hydrodynamic treatment of the evolution of a system of gas clouds, and it gives models which reproduce closely the observed structure of elliptical galaxies.

If the mean free path of the gas clouds is comparable to or greater than the size of the system, the viscosity can no longer be treated as a purely local effect related to the shear of the velocity field, and it is even possible that the application of the Navier-Stokes equations can give the wrong sign for the viscosity effect. However, it is still possible to define the components of the viscous stress tensor in terms of moments of the velocity distribution, and these moments can in principle be calculated by solving the appropriate moment equations. An example of such a set of moment equations for a system with cylindrical symmetry has been given by Larson (1975c, Appendix), where approximate solutions for the limiting cases of long and short mean free paths have been discussed. In the limit of short mean free paths these equations reduce to the Navier-Stokes equations, but in the limit of long mean free paths (for example, in a purely stellar system) the viscosity vanishes for an equilibrium system and is appreciable only if the system is dynamically evolving, as in an initial rapid collapse. In this situation the viscosity again acts to transfer angular momentum outward, qualitatively as predicted by the Navier-Stokes equations. The magnitude of the viscosity is given by the second-order moment $\delta = \langle (u - \langle u \rangle)(w - \langle w \rangle) \rangle$, where u is the radial component and w the transverse (rotational) component of velocity. The simplest dimensional expression which predicts qualitatively the right behavior for δ during the collapse of a system of gas clouds is $\delta \sim - \langle u \rangle \langle w \rangle$, which implies that the outward transport of angular momentum by random motions is comparable to the inward transport of angular momentum by the mean flow. The time scale for viscous transport of angular momentum implied by this assumption for δ is

$$\tau_{vis} \sim - \frac{r}{\langle u \rangle} , \qquad (10)$$

which is of the order of the free-fall time during the initial rapid collapse but becomes longer during later phases of evolution when the remaining gas settles toward an equilibrium disc and the contraction velocity $\langle u \rangle$ becomes small. In this respect this approach is probably more realistic than the use of a kinematic viscosity given by equation (8), which implies that the viscous time scale decreases as the gas density decreases.

Heating of the Gas

Although we have seen that it is unlikely that all of the gas in a protogalaxy could be maintained at a very high temperature, it is nevertheless possible that part of the gas could be heated to a high temperature and form a hot diffuse medium; if the energy input is sufficient, this hot gas may even stream out of the galaxy in a hot "galactic wind" analogous to the solar wind. The most important source of heating is probably supernova explosions, although other effects such as stellar winds and thermalization of mass motions probably also contribute. With conventional values of the parameters, the supernova heating rate is sufficient to drive a hot wind and exhaust the gas from elliptical galaxies (Mathews & Baker 1971) and probably from the bulge components of spirals (Faber & Gallagher 1976). This effect may become important even at an early stage in the evolution of a galaxy and may result in significant mass loss, especially for small galaxies (Larson 1974c).

The effect of a supernova explosion occurring in a uniform gas of typical interstellar density is to create a hot "bubble" of low density gas surrounded by an expanding thin dense shell of swept-up ambient material. If the supernova rate is sufficiently low or the ambient density sufficiently high, most of the supernova energy is radiated away from the dense shell before neighboring supernova remnants intersect, but if the supernova rate is higher or the gas density lower, the hot dilute bubbles generated by different supernovae can merge and form an extended hot dilute medium before the supernova remnants have been completely dissipated (Cox & Smith 1974). The interstellar medium will then be divided into (at least) two distinct phases: cool dense "cloud" gas originating from the compressed outer shells of supernova remnants, and hot dilute "intercloud" gas which is directly heated by the supernova explosions. Since a galactic wind will presumably involve mostly the hot dilute component, more detailed models for the gas dynamics during early stages of galactic evolution will probably have to consider a multi-phase gas;this has not yet been done, but will be an important direction for further research.

Some simple estimates of the effect of supernova heating and gas loss on the early evolution of galaxies have been made by Larson (1974c). According to these estimates, under typical conditions about 90% of the initial supernova energy of $\sim 10^{51}$ ergs is radiated away, and 10% or $\sim 10^{50}$ ergs is available in the form of thermal energy to drive a hot galactic wind. This amount of energy is still sufficient to remove a significant amount of gas from a galaxy, particularly for galaxies of small mass and hence small escape velocity. This can have important effects on both the structure and the composition of small galaxies by removing the chemically enriched gas which would otherwise have condensed to form a metal rich dense core or disc component (see below).

3. STAR FORMATION

A central role in the formation and evolution of galaxies is played by the process of star formation. Model calculations such as those to be discussed below emphasize the importance of the star formation rate in a collapsing protogalaxy in determining the structure of the resulting system: if star formation occurs very rapidly, a spheroidal system of stars is formed, whereas if star formation is very slow, most of the gas may settle into a disc before forming stars. The disc/bulge ratio of a galaxy thus depends on the rate of star formation during early stages of the collapse, and in order to understand the Hubble sequence it is necessary to understand the factors which govern the star formation rate in protogalaxies. Likewise, since element synthesis occurs predominantly in massive, short-lived stars,the rate of heavy element production and chemical evolution in a galaxy depend on the star formation rate and how it varies with time and position. The chemical evolution and many of the observed properties of a galaxy, such as the mass/light ratio, color, and supernova rate, also depend on the initial mass spectrum with which stars are formed. Since properties such as the mass/light ratio seem to differ between different galaxies or stellar populations, it is also important to understand the factors which influence the initial stellar mass spectrum.

Possible Processes of Star Formation

In order for an element of gas with a mass of,say, 1 $M_{\odot}$ to collapse into a star, it must first be compressed to a sufficiently high density that its self gravity can overcome the forces tending to disperse it. If thermal pressure is the dominant dispersive force, the mass, density, and temperature of a marginally unstable condensation are related by the Jeans criterion (equation 1). For a temperature of the order of 5 °K, as commonly encountered in dark molecular clouds, a condensation of mass 1 $M_\odot$ must have a minimum density of the order of 10^4 cm^{-3} to collapse; this density is higher than that of typical interstellar (or protogalactic) clouds, and somewhat higher even than the typical density of dark molecular clouds. To understand how stars form, we must therefore understand how the gas is compressed to such high densities.

The traditional view is that star formation begins with the collapse of large interstellar clouds of mass $\gtrsim 10^3$ $M_\odot$, and that smaller condensations and protostars form by a hierarchial fragmentation process as the density in such a collapsing cloud rises and the Jeans mass decreases. Such a hierarchial fragmentation process may provide a basis for understanding the stellar mass spectrum (Larson 1973) if fragmentation is sufficiently efficient. However, it is not clear that successive fragmentation can proceed as effectively as postulated by Larson (1973), and the numerical calculations that have been done for the non-spherical collapse of isothermal gas clouds suggest that not much fragmentation is likely to occur after the early stages of the collapse, except perhaps for the formation of binary or small multiple systems (Larson 1972b, Black & Bodenheimer 1976).

Instead, it may be that the masses of collapsing cloud fragments or protostars are largely determined by the value of the Jeans mass at the time when a collapsing region begins to collapse gravitationally and its evolution is no longer dominated by external forces. If so, it is important to understand the way in which dense molecular clouds are formed and the conditions existing when they begin to collapse gravitationally.

If a protostar of 1 $M_\odot$ is to be compressed to a density of $\gtrsim 10^4$ cm^{-3} before it can begin to collapse gravitationally, the required external pressure can be expressed as $P/k = nT \gtrsim 10^5$ cm^{-3} °K. This pressure is much higher than that of the general interstellar medium or typical interstellar clouds in our galaxy, but pressures of this order are easily generated by hydrodynamic phenomena such as cloud collisions; for example, if clouds of density 10 cm^{-3} collide at a velocity of 10 km s^{-1}, the dynamical pressure ρu^2 is equivalent to $nT \sim 10^5$ cm^{-3} °K, which is sufficient to compress the gas to protostellar densities. Likewise, the observed internal motions in molecular clouds, where the densities are of the order of 10^3 cm^{-3} and the velocities are typically $\approx$ 1 km s^{-1}, can also create dynamical pressures of the order of 10^5 cm^{-3} °K and hence produce locally compressed regions with densities $\gtrsim 10^4$ cm^{-3}. Thus it may be that hydrodynamical processes produce directly the conditions required for the collapse of protostars of mass $\lesssim 1$ $M_\odot$, and that hierarchial fragmentation is less important than has usually been thought.

If the above speculations are correct, star formation is closely related to the larger scale gas dynamics of a galaxy and to processes such as collisions between gas streams or clouds which cause strong shock compression of the gas. Observationally, a relation between star formation and large scale dynamics is indicated by the occurrence of star formation in spiral arms which, whatever their detailed nature, evidently represent regions of compression of the gas in spiral galaxies. Another possible indication of a connection between star formation and large scale dynamics is the evidence for rapid star formation in a number of peculiar galaxies which are interpreted as systems in collision. On small scales, the relation between star formation and dynamics is less clear, mainly because of the difficulty in understanding the complicated motions in molecular clouds, but in the case of NGC 1333 it has been found that a region of active star formation is closely associated with a collision between two molecular clouds (Loren 1976).

If a galaxy forms by the violent collapse or coalescence of a system of gas clouds, the collisions of these gas clouds with each other at velocities of some tens or hundreds of km s^{-1} may trigger a rapid burst of star formation which converts a considerable fraction of the gas into stars at an early stage of the collapse. As will be seen, such an early phase of rapid star formation is required in conventional models for the formation of spheroidal systems, i.e. elliptical galaxies or the bulge-halo components of spirals. However, the gas that eventually forms a disc system must experience much less efficient star formation, so that it has time to settle into a thin disc before turning into stars. Disc systems may therefore form from gas which never participates in strong shock compression but remains relatively homogeneous and quiescent, and settles gradually into a disc after the initial rapid formation of the spheroidal component. The processes which eventually form stars in the disc may then differ considerably from those which occurred in the halo; for example, because the velocity dispersion of gas clouds in the disc is much smaller, one might expect less effective compression of the gas and less efficient star formation. Even if large scale dynamical phenomena like spiral shock fronts are important, the velocities involved are only of the order of 10 km s^{-1} and may not be able to generate very efficient compression of the gas. A general prediction would then be that star formation proceeds much more slowly in the discs of spiral galaxies than in the rapidly collapsing protogalaxies or colliding galaxies which eventually form spheroidal stellar systems. Clearly these remarks are still rather speculative, and much more study, both theoretical and observational, of the detailed processes by which stars form is required before the modeling of galaxy formation can be put on a firm predictive footing.

The Stellar Mass Spectrum

It has usually been assumed that the initial mass function (IMF) with which stars are formed is determined by the detailed way in which a cloud fragments into protostars of different masses as it collapses. Since the dynamics of fragmentation is very complicated and poorly understood, no detailed predictions of the outcome of this process can be made, although it may be possible to adopt a stochastic model of hierarchial fragmentation to make probabilistic predictions (Larson 1973).

Another type of process which may turn out to be more important for understanding stellar masses and the IMF involves interactions that can occur between an accreting protostellar core and the surrounding gas; these processes can determine how much mass the accreting core or "embryo star" can acquire before its growth is stopped. Larson and Starrfield (1971) studied a number of effects which become important for massive and luminous protostellar cores and which may set an upper limit on stellar masses; they concluded that the most important effect is the ionization of the remnant protostellar envelope when the core becomes sufficiently massive and hot, and that this sets a limit of $\sim$ 30 - 60 $M_\odot$ on stellar masses. Kahn (1974), on the basis of a similar model but a more detailed consideration of the radiative transfer, concluded that radiation pressure is a more important effect and sets a mass limit of $\sim$ 40 $M_\odot$ with normal dust opacities. A third effect which has not yet been studied quantitatively in this context but which may turn out to be the most important one is the effect of stellar winds from hot stars in dispersing surrounding gas. Winds may also be important for less massive stars, since a number of T Tauri stars show evidence for mass outflow even at very early stages of evolution, and the T Tauri winds may be responsible for dispersing remnant protostellar envelopes and stopping the growth in mass of protostellar cores (Strom et al.

1975, Larson 1975a).

If the mass to which an accreting protostellar core can grow is limited because the core blows off its infalling envelope at some stage in its growth, then one would expect that, regardless of the details of the mechanism, there will be a correlation between the limiting mass of the core and the density of the surrounding cloud. This is because any process which stops accretion must counteract the dynamical pressure ρu^2 of the infalling gas, which is proportional to the ambient density; thus the higher the cloud density, the stronger is the force required to stop the infall, and this presumably requires a more massive and luminous star. In dark clouds of moderate density, a T Tauri star with a mass of only $\approx 1\ M_\odot$ may generate a strong enough wind to disperse its remnant envelope and cut off further growth in mass, whereas in a much denser cloud, more massive stars may form before stellar winds or other effects become strong enough to disperse the remnant gas. Observational evidence for such a correlation is provided by the existence of large numbers of low mass T Tauri stars widely distributed in associations containing dark clouds of moderate size and density, while more massive young objects are found near the centers of massive, dense, centrally concentrated clouds. Another example is provided by the two molecular clouds OMC1 and OMC2 in Orion which contain clusters of infrared sources thought to be protostars or newly formed stars; the more massive, dense, and well studied cloud contains very luminous infrared sources, i.e. massive young stellar objects, whereas the less dense cloud contains much fainter infrared sources, i.e. less massive stars (eg. Werner et al. 1974); this suggests that massive stars form only in the densest clouds, and less massive stars form in less dense clouds.

We cannot yet predict how much such variations may contribute to the observed stellar mass spectrum, which may represent an average over different regions of star formation which produce stars of different masses. However, we might expect some broad correlations between the average stellar mass and the ambient density in various regions of galaxies; for example, it may be that low density regions in the outer parts of galaxies produce many low mass stars and therefore have high mass/light ratios, whereas the dense inner parts of galaxies produce more massive stars and therefore have smaller mass/light ratios, in qualitative agreement with present evidence. Again, however, the need for a more detailed understanding of star formation processes is evident.

4. FORMATION OF ELLIPTICAL GALAXIES

Because elliptical galaxies appear to have the simplest and most regular structures, and have radial surface brightness profiles I(r) that are quite similar to each other, it is natural that the most theoretical attention has been devoted to understanding the structure of E galaxies. A successful theory of galaxy formation should provide some basis for understanding why E galaxies have such similar surface brightness profiles (Liller 1960, 1966; van Houten 1961; King 1976; Oemler 1976), while allowing for the fact that they also vary considerably in their ellipticity profiles $\varepsilon(r)$. In addition, E galaxies apparently all possess radial composition gradients in at least their nuclear regions, which must be accounted for by a model of the formation process.

Several types of models for the formation of E galaxies have been proposed. In some of these it is assumed that stars or small galaxies are somehow first formed at a very early time, and that E galaxies then form by processes involving only stellar dynamics, such as the collapse of a system of stars or the collision and coalescence of smaller stellar systems. In other models it is assumed that the material is initially gaseous, and gas dynamics and star formation are the most important processes. We shall consider these two classes of models in turn.

Stellar Dynamical Models

Because of the relative simplicity of stellar dynamics as compared with gas dynamics, and because of the possibility that violent relaxation processes may be able to establish a centrally condensed, "relaxed" distribution of stars without involving the complexities of gas dynamics and gaseous dissipation, purely stellar collapse models have received a fair amount of attention. Detailed numerical calculations of the collapse of a rotating axisymmetric system of stars have been made by Gott (1973, 1975), who has represented the system by 2000 discrete mass rings whose orbits are computed individually.

In his first paper, Gott (1973) considers the collapse of a system with a total mass of 10^{11} $M_\odot$, whose initial configuration is a uniform sphere of radius 20 kpc with uniform (solid body) rotation and a small initial velocity dispersion; the initial angular velocity is treated as a free parameter, and several values are considered. The system of rings is allowed to collapse and its dynamics is followed for several free-fall times until it has settled into a steady equilibrium state. The final equilibrium configurations obtained by Gott are centrally condensed and moderately flattened, and qualitatively resemble elliptical galaxies in structure. The ellipticity $\varepsilon = 1 - b/a$ increases with increasing initial angular velocity up to a maximum value of $\varepsilon \sim 0.5$, and the radial density distribution $\rho(r)$ is fairly similar in models of different ellipticity; again, these characteristics resemble those of elliptical galaxies. A point of disagreement with observations is that the predicted density distributions in the outer parts of the models are approximately of the form $\rho(r) \propto r^{-4}$, considerably steeper than the density distributions observed in E galaxies, which are approximately $\rho \propto r^{-2.7}$ to $\rho \propto r^{-3.0}$ (assuming a constant M/L). Also, these models do not reproduce the small sharp cores observed in E galaxies, although this could be partly because of numerical limitations. Thus the stellar collapse models are not as centrally condensed as real E galaxies and do not yield quantitative agreement with observed radial density distributions.

In a second paper, Gott (1975) considers the effect of "cosmological infall" and shows that when this effect is included in the collapse models, better agreement between predicted and observed density distributions is obtained. If a galaxy begins as a density fluctuation in an otherwise uniform expanding medium, then some of the material surrounding this density perturbation is gravitationally bound to it and will eventually stop expanding and fall back to join it; as shells farther and farther out from the center of the perturbation stop expanding and recollapse, a continuing "cosmological infall" of matter into the forming galaxy is created, and this may add significantly to its mass. This infalling matter is assumed by Gott to be in the form of stars, so that it forms an extended envelope around the inner dense part which is formed from the initial density perturbation. For these "infall" models the radial density distribution is approximately $\rho \propto r^{-2.8}$, in good agreement with that of elliptical galaxies. A detailed comparison is made with the E4 galaxy NGC 4697, and the model is shown to fit both the radial surface brightness profile $I(r)$ and the ellipticity profile $\varepsilon(r)$ of the envelope of this galaxy. The sharp cores of E galaxies are, however, still not reproduced by these stellar infall models, and it appears that additional processes such as gas infall and dissipation are required to produce very dense nuclei.

Although the infall models of Gott (1975) successfully reproduce the density profiles of the envelopes of E galaxies, the resulting density distribution is determined primarily by the infall effect rather than by "violent relaxation", which plays a relatively less important role in the infall models. Since the infall process depends on the assumed initial conditions and boundary conditions for a collapsing protogalaxy, the resulting density distribution might be expected to differ for systems formed in different circumstances, although Gott (1975) argues that such differences should not be great since inhomogeneities in the structure and expansion rate of the universe were relatively small at the time when protogalaxies began to separate out from the expanding background. At any rate, it does not seem that the amount of violent relaxation in the Gott models is sufficient to produce a nearly "universal" structure for E galaxies which is independent of the initial conditions; some degree of similarity of the initial conditions must still be assumed.

A second purely stellar model for the formation of E galaxies which again involves a form of violent relaxation has been suggested by Toomre (1974, 1976). There is strong evidence that some close pairs of galaxies are interacting tidally, and such interactions must dissipate a certain amount of orbital energy, causing at least bound pairs of galaxies to spiral together and eventually merge. Toomre has estimated that the number of mergers between galaxies is large enough that a significant fraction of galaxies must have formed in this way; in particular, a significant fraction of E galaxies may have formed by collisions and coalescence of smaller galaxies. Toomre has simulated such a coalescence process by calculating a face-on collision between two axisymmetric disc galaxies represented as systems of rings. The colliding systems pass through each other but lose enough energy by dynamical friction that they fall together again and after only a few oscillations merge into a single elliptical-like system. The resulting system is centrally condensed and posseses an extended envelope of stars thrown out during the collision; the density distribution is found to be approximately $\rho \propto r^{-3}$, satisfactorily matching that of typical E galaxies. This final density distribution is apparently the result of a violent relaxation effect which produces sufficient energy redistribution to provide both a dense core and an extended envelope, although the model still does not reproduce the observed degree of central concentration. Possibly the collision process studied by Toomre involves more "violent relaxation" than the simple collapse model studied by Gott, and hence produces a closer approach to an asymptotic "relaxed" distribution. If so it may, after all, be possible for violent relaxation processes to produce an approach to some universal limiting distribution if galaxies form with a sufficient amount of chaos and violence; however, this question will require many more calculations to clarify.

Although it is not clear whether purely stellar dynamical processes can account for the density profiles of normal E galaxies, stellar dynamics may be more important for the formation of the cD galaxies, which occur only in dense clusters and possess extended envelopes which probably had a different mode of origin from those of normal E galaxies, possibly involving a tidal stripping or disruption of galaxies in the cluster (Oemler 1976). Since nearly all known cD galaxies are the central brightest members of giant clusters and their envelopes extend throughout a considerable part of the cluster, it is possible that the envelopes represent an intergalactic sea of debris which has been lost from galaxies in the cluster but remains bound in the gravitational potential well of the cluster. Dynamical friction and stripping effects are relatively small for high velocity encounters, but are much more important for lower velocity encounters such as might occur between a large galaxy and a smaller orbiting companion. Such encounters can eject a significant fraction of the mass of the galaxy, (eg. Gutowski and Larson 1976), thereby helping to populate very extended halos around galaxies.

Gas Dynamical Models

Gas dynamical models for the formation of E galaxies based on the concepts discussed in Sections 2(b) and 3 have been computed by Larson (1969, 1974b, 1975c), and the results of these calculations will be summarized here. We consider first the results obtained with spherical non-rotating models, since they successfully represent many of the basic properties of E galaxies and also allow a more detailed study of the effects of various parameters and assumptions than do the rotating models, which are otherwise more realistic but require more computation.

Larson (1974b) found that a simple spherical collapse model in which the star formation time scale and the gaseous dissipation time are both assumed to be proportional to the local free-fall time scale was able, with reasonable values of the parameters, to reproduce quite closely the radial surface brightness profile of the typical E1 galaxy NGC 3379, which is the galaxy for which the most accurate and extensive photoelectric surface photometry is available. Although this model is not a unique fit to the observations and therefore not a strong test of the validity of the underlying assumptions, it is noteworthy that, unlike the available purely stellar models, it reproduces the entire density profile, including the sharp core, to the limit of resolution. In fact, the density near the center of the model increases toward smaller radii according to a power law $\rho \propto r^{-0.75}$, so that the density distribution is actually singular at the center, within the numerical resolution. This result provides the possibility of qualitatively explaining the extremely small, dense nuclei observed in M31 and M32.

A second important result of the spherical gas dynamical models is the prediction of a radial gradient of metal abundance, at least in the inner regions of the models. This abundance gradient results because the protogalactic gas becomes more and more enriched in heavy elements as it falls toward the center through a background of already formed stars, and later generations of stars formed from this gas reflect the abundance gradient thus established in the gas. Composition gradients are similarly established in any system in which gas flows transport heavy elements from one part of the system to another before star formation is complete. Thus the presence of composition gradients in galaxies is an indication of the importance of gas flows during the formation process. Many, if not all, large E galaxies appear to possess radial composition gradients in at least the innermost few kpc, in qualitative agreement with the predictions of the gaseous collapse models. It is not yet clear whether these abundance gradients extend out into the halo regions of E galaxies, but there is some evidence for gradients extending far into the halos of at least some galaxies. Such measurements can provide important information on the efficiency of star formation during early stages of the collapse process; for example, models such as Model D of Larson (1974b) which have rapid star formation during the early stages of the collapse predict fairly high metal abundances at all radii and only a weak abundance gradient in the halo, whereas models such as Model F which have less efficient star formation during the earliest stages predict steeper abundance gradients and lower metal abundances at large radii.

Although no galaxy formation models have yet properly calculated the gas flows that occur when supernova heating causes some of the gas to be expelled from a young galaxy (at least in the case of low mass systems), the effect of supernova-driven gas loss can be roughly simulated by suddenly removing all of the residual gas from a protogalaxy when the estimated total energy input from supernovae is sufficient to eject all of the remaining gas (Larson 1974c). Since galaxies of smaller mass have smaller binding energies, the fraction of the initial mass which is thereby lost increases with decreasing mass. This affects both the structure and the chemical composition of the resulting system, since if a major fraction of the mass is lost, the gas which would otherwise have formed a centrally condensed, metal rich core is instead lost and the resulting galaxy is less centrally condensed and has a lower metal abundance. These predictions are in qualitative agreement with the observed trends, which are that the metal abundances of E galaxies increase systematically with mass, and that dwarf E galaxies lack the sharp condensed cores characteristic of large E galaxies. These results again indicate that gas flows probably play an important role in the formation of galaxies.

Axisymmetric rotating models for the formation of elliptical galaxies have been computed by Larson (1975c). A principal conclusion of these calculations is the importance of turbulent viscosity in redistributing angular momentum during the collapse and making it possible for the inner regions to become highly condensed without becoming excessively flattened. Without turbulent viscosity, no dense nucleus is formed and the inner part of the system becomes highly flattened, unlike E galaxies which, more often than not, become rounder toward the center; however, when a plausible amount of turbulent viscosity is assumed, a density distribution closely resembling that of typical elliptical galaxies is obtained, including iso-density contours that become rounder toward the center. The existence of a viscosity effect due to random motions is not restricted to gas dynamical models, and in fact some viscosity due to random stellar motions is present even in the purely stellar models of Gott and Toomre. From the results described above, it appears that both dissipation, which is characteristic of a system containing gas, and viscosity, which is characteristic of a system with random motions, are required to account for the highly centrally condensed yet not highly flattened structure of elliptical galaxies.

It is observed that E galaxies, while having similar radial density profiles, have a great variety of ellipticity profiles $\varepsilon(r)$. The rotating collapse models of Larson (1975c) show that the ellipticity of the outer regions depends mainly on the initial angular velocity, whereas the ellipticity of the inner part depends mainly on the assumed effectiveness of the turbulent viscosity; by varying these two parameters, it is possible to generate a variety of ellipticity profiles. The observed differences in the $\varepsilon(r)$ profiles could therefore reflect differences in the details of viscous transfer of angular momentum during the collapse, which could arise on a statistical basis if a protogalaxy contains only a small number of independently moving mass elements; in this case, large differences in both the initial distribution and in the subsequent redistribution of angular momentum during the collapse might be expected. Thus, although E galaxies appear smooth and regular, their ellipticity profiles may betray some vestiges of the initial chaos which have not been completely obliterated by any relaxation processes.

The metal abundance distributions in the rotating models are not radially symmetrical but reflect the fact that in a rotating protogalaxy the gas infall occurs predominantly in the vertical direction as the gas settles toward a flattened disc-like distribution in the equatorial plane. Because the gas flow is predominantly vertical, the metal abundance gradient is also predominantly in the vertical direction and the metallicity distribution is strongly flattened toward the equatorial plane. In fact, the metallicity distribution is predicted to be more strongly flattened than the density distribution, and an observational confirmation of this prediction would provide strong evidence that the composition gradient is produced by gas flows rather than, for example, by a dependence of the stellar IMF on the local density. Strom et al. (1976) have recently obtained evidence for flattened metallicity distributions in E galaxies which are in qualitative agreement with the predicitions of the gas dynamical collapse models.

The predicted distributions of mean rotational velocity and velocity dispersion may be described as reflecting a composite structure of the models, with two main components: a nearly spherical metal-poor "halo" component with slow rotation and a large velocity dispersion, and a relatively flattened, metal-rich, disc-like component with faster rotation and a smaller velocity dispersion. Thus, a sequence of stellar population characteristics resembling that found in our galaxy is present even in these much less flattened models of E galaxies. Evidence for these predicted trends in the kinematics of E galaxies would again argue for a gaseous collapse model for the formation of elliptical galaxies, because the existence of subsystems with different flattening and different rotational velocity would not be expected on the basis of a purely stellar violent relaxation model. For example, the stellar dynamical equilibrium models of Wilson (1975) do not possess the strong gradient of rotational velocity perpendicular to the equatorial plane which is characteristic of the gaseous collapse models of Larson (1975c). However, observations to test this prediction are lacking, and will be difficult to obtain.

Summary

The similar radial light profiles I(r) of E galaxies suggest that some rather general mechanism operates during the formation of E galaxies to produce an approach to a standard radial density distribution $\rho(r)$. However, theorists have produced several possible models, all of which claim to be able to reproduce observed light profiles, and it is not yet possible to decide which processes are most important in determining $\rho(r)$. The mechanisms which may play a role are: (1) "cosmological infall", as in the infall models of Gott (1975) and the expanding boundary models of Larson (1974b); (2) a variety of relaxation-type effects produced by large and rapid fluctuations in the gravitational potential field, as in the collision and coalescence model of Toomre (1976); (3) a more complicated interplay between gaseous dissipation, infall, and star formation, as in the gas dynamical models of Larson (1974b, 1975c). The models involving only stellar dynamics are appealing because of their simplicity, and it cannot be ruled out that infall or relaxation-type processes play an important role in establishing the characteristic radial structure of E galaxies. However, purely stellar models have not yet been shown to be capable of producing the full density range from highly condensed cores to very extended envelopes that is observed in galaxies; the sharp dense cores of giant E galaxies, in particular, are reproduced only by models in which gaseous dissipation and turbulent viscosity are both present. The composition gradients observed in galactic nuclei also provide evidence for the importance of gas infall in the formation of at least the central parts of galaxies. If composition gradients also extend throughout galactic halos, gas flows must play a role even in the formation of galactic halos. However, this evidence for the importance of gas dynamics does not rule out a picture in which stellar relaxation-type effects are also important, as in the collision model of Toomre; if the colliding galaxies are at an early stage of evolution and still contain a large amount of gas, the subsequent dynamics of this gas and the formation of later generations of stars from it could produce the basic features of the gas dynamical models.

In view of the many theoretical uncertainties, considerable importance attaches to any attempts to identify and study observationally systems in which the processes discussed above may be still operating. Such systems would not yet have attained the very smooth and regular structure charactistic of normal E galaxies, but instead would presumably be classed as peculiar objects. The possibility that some peculiar galaxies may in some sense be young or still forming systems will be considered further below.

5. FORMATION OF SPIRAL GALAXIES

Spiral and S0 galaxies are generally considered to be made up of two more or less distinct components: a centrally condensed, spheroidal "bulge" component, and a less centrally condensed flat disc component. The relative prominence of the disc and bulge components varies greatly in different galaxies, and is one of the criteria of the Hubble classification scheme for spiral galaxies, the other criteria being the openness and degree of resolution of the spiral arms. The optical appearance of spiral galaxies is presumably related through the detailed mechanics of spiral structure and star formation to fundamental structural properties, such as the underlying mass distribution and the relative amount and distribution of gas. According to a recent application of density wave theory (Roberts et al. 1975), the form of the spiral pattern depends on the degree of central concentration of the mass distribution; a galaxy with a more centrally concentrated mass distribution, i.e. a more dominant bulge component, should have a more tightly wound spiral pattern, as is generally observed. On this basis, it might be concluded that the Hubble sequence is fundamentally a sequence in disc/bulge ratio. However, Sandage et al. (1970) have concluded that the Hubble sequence is fundamentally a sequence in fractional gas content, and that the disc/bulge ratio correlates only weakly with gas content or Hubble type. Probably both disc/bulge ratio and gas content are important, and a more adequate classification scheme should treat them as independent parameters (van den Bergh 1976). From the point of view of formation theories, however, the disc/bulge ratio is probably the more fundamental parameter, since the gas content can be altered in a number of ways during the evolution of the galaxy. Thus we must be prepared to explain the formation of two-component systems ranging all the way from those containing little or no disc (E galaxies) through those in which both bulge and disc are prominent (eg. Sa galaxies) to those which are nearly all disc, with little visible bulge (eg. late-type Sc galaxies).

In contrast to the situation for E galaxies, it is clear that the formation of a flat disc by the collapse of an extended distribution of matter requires a large amount of dissipation, and therefore must involve gas rather than stars. Star formation cannot be too rapid in this gas, or it will be turned into stars before it has dissipated enough energy to settle into a thin disc. This is the opposite of what is required to form a spheroidal stellar system, in which the stars must form while the gas is still in an extended distribution and has not yet collapsed to a disc. Thus the fundamental distinction between the formation of spheroidal and disc systems is in the relative time scales for star formation and collapse of the gas to a disc, and the disc/bulge ratio in a composite system depends on the fraction of the initial gas which is turned into stars before the gas settles to a disc. This is illustrated by the simple model of Gott & Thuan (1976) in which gas is turned into stars at a prescribed rate in a collapsing protogalaxy, and the material which is still in the form of gas when it reaches the equatorial plane is assumed to remain in the plane and form a disc.

Some of the models of Larson (1975c) for the formation of E galaxies contain an incipient disc component which forms from the small amount of residual gas (including recycled gas) which is not consumed by star formation during early stages of the collapse. Models of this type are thus capable of representing the formation of systems with both a spheroidal and a disc component, qualitatively resembling the spiral and SO galaxies. However, in none of these models does the disc component contain more than a very small fraction ($\sim$ 5 - 10 percent) of the total mass. Several ways in which the assumptions of these models might be modified to yield more massive disc components have been considered in the calculations of Larson (1976) of models for the formation of spiral galaxies.

It has often been proposed, for example, that the Hubble sequence represents a sequence of increasing angular momentum per unit mass, the elliptical galaxies having the least angular momentum per unit mass and the late-type spirals the largest amount. However, the model calculations do not support this hypothesis; increasing the initial angular momentum does not significantly increase the fractional mass in a disc component but merely results in a more flattened spheroidal component and a more extended disc. Another suggestion (eg. Gott & Thuan 1976) is that the disc/bulge ratio is a function of the initial proto-galactic density: if the star formation rate is approximately proportional to the square of the gas density ρ, as is often assumed, then the ratio of the star formation time scale ($\propto \rho^{-1}$) to the free-fall time ($\propto \rho^{-1/2}$) increases as the gas density is decreased, and one might then expect a greater fraction of the initial matter to remain in the form of gas until it has settled into a thin disc. However, model calculations again show that this effect is too small to explain the entire Hubble sequence on this basis alone, at least with simple collapse models, and the predicted trend may even be negated by other counteractiing effects, such as an increase in the gaseous dissipation time scale along with the star formation time as the initial density is decreased. Thus it is clear that more basic changes in the model assumptions, particularly with regard to the rate of star formation, are necessary to account for the formation of galaxies with massive disc components.

Star Formation Rates in Proto-Spiral Galaxies

What appears to be necessary to explain the formation of spiral galaxies with distinct bulge and disc components is basically a two-stage formation process, with an early phase of rapid star formation which forms a spheroidal component, followed by a later stage during which star formation is much slower and the residual gas condenses into a thin disc before forming stars. Thus, relative to the rapid early star formation that forms the bulge, the star formation rate (SFR) must be greatly reduced during the later stages of the collapse, by an amount greater than is predicted by a conventional power-law function of gas density. This means either that star formation must be suppressed during the later stages of the collapse by mechanisms not yet considered, assuming that star formation is normally a rapid process, or that the star formation rate must somehow be strongly enhanced during the initial phases of the collapse, if star formation is normally a much slower process.

A possible mechanism tending to suppress star formation during the later stages of the collapse is the action of tidal forces in inhibiting small scale gravitational instabilities and collapse processes in residual gas of low density. If the density of the residual gas at any radius in a protogalaxy is less than the mean density of all matter inside that radius, tidal disruptive forces will predominate over local self gravity in the gas and prevent collapse into stars, unless sufficiently large density enhancements are already present. Since the gas probably does contain large inhomogeneities, it is difficult to predict quantitatively the effect of tidal forces on the SFR without knowing in detail the small scale density distribution in the gas. Larson (1976) adopted the following simple but arbitrary formula to represent the star formation rate in a situation where tidal forces are important:

$$\frac{d\rho_s}{dt} = A\,\rho_g^{\ n}\ [\ 1 + B\,\bar{\rho}(r)\ /\rho_g\ ]^{-1} \tag{11}$$

where $\bar{\rho}(r)$ is the mean density of all matter inside a sphere of radius r and B is a constant of order unity. Equation (11) implies that the SFR follows the power law $\rho_g^{\ n}$ during early stages of the collapse, but is reduced by approximately a factor of $\rho_g/B\bar{\rho}(r)$ during later stages when the gas density becomes small compared with $B\bar{\rho}(r)$; such a reduction factor might be justified physically if, for example, a fraction $\rho_g/B\bar{\rho}(r)$ of the gas is in clumps of density greater than $\bar{\rho}(r)$ which can still collapse to form stars despite the tidal forces.

Models calculated using equation (11) for the star formation rate have the desired property that during the early stages of the collapse, rapid star formation produces a spheroidal bulge-halo component resembling a small elliptical galaxy, whereas during the later stages of the collapse, star formation is suppressed until the residual gas has settled into a thin dense disc. This results in the formation of a distinct disc component which may contain as much as roughly half of the total mass of the system, depending on the values of A and B in equation (11). Since the effectiveness of tidal forces in inhibiting star formation clearly depends on the degree of clumpiness of the gas, being less important for a very clumpy gas, the parameters A and B should perhaps be considered as depending on the clumpiness of the gas, and they may therefore have different values for different protogalaxies or may vary with time in a protogalaxy as the dense clouds collapse into stars. Model calculations in which the parameters A and B are varied show, as might be expected, that the fraction of the total mass which ends up in a disc is increased if the SFR is reduced during the early stages of the collapse, causing a smaller fraction of the gas to form stars in a spheroidal component; this might occur, for example, in a protogalaxy in which the degree of clumping of the gas is relatively small. Also, the distinction between bulge and disc components is made sharper if the SFR is further reduced after a centrally condensed spheroidal component has formed, and if no significant further star formation occurs until the gas has settled into a thin dense disc; this might occur, for example, if all of the dense cloud gas is turned into stars early in the collapse, leaving only relatively diffuse gas whose density is too low to form stars until it has settled to a disc.

The model results referred to above clearly demonstrate the importance of the star formation rate and its variation with time in determining morphological properties such as the disc/bulge ratio of the resulting system, and they also suggest the importance of the clumpiness of the gas as a physical factor influencing the star formation rate. However, the modifications which were tried in the assumed SFR were fairly ad hoc and it would be desirable to have some theoretical basis for these assumptions. A possibility which was discussed in Section 3 is that strong turbulence or fast collisions between gas clouds or streams in an irregularly collapsing protogalaxy may compress parts of the gas to very high densities and hence cause rapid star formation. If the gas in a protogalaxy is distributed in discrete clouds, the SFR may then depend on the rate and perhaps also on the velocity with which the clouds collide.

A number of models in which the SFR depends on a cloud collision rate have been calculated by Larson (1976). In the colliding cloud model of Larson (1969), the cloud collision time is

$$\tau_{coll} \sim 3.1 \times 10^2 \, (\rho_g \alpha_g)^{-1/2} \text{ Myr} \quad , \tag{12}$$

and the ratio of cloud collision time to free-fall time τ_{ff} (equation 6) is

$$\tau_{coll}/\tau_{ff} \sim 40 \, (\bar{\rho}/\rho_g \alpha_g)^{1/2} \quad . \tag{13}$$

Larson (1976) assumed that the SFR depends on the ratio τ_{coll}/τ_{ff} , being high when the collision time is short comparted with τ_{ff}, but reduced by a factor proportional to τ_{ff}/τ_{coll} during later stages of the collapse when τ_{coll} becomes longer than τ_{ff}; more specifically, a star formation law of the form

$$\frac{d\rho_s}{dt} = A \, \rho_g^{\,n} \, [\, 1 + C \, \bar{\rho}/\rho_g \alpha_g \,]^{-1} \tag{14}$$

was assumed, analogous to equation (11). The results calculated using equation (14) are qualitatively similar to those obtained with equation (11), but the important effects are more marked because the dependence of the SFR on the velocity dispersion as well as the density of the gas tends to amplify the variations in the SFR. For example, the decrease in velocity dispersion of the residual gas during the later stages of the collapse as it settles toward a flattened disc-like distribution causes a decrease in the SFR, with the result that more of the gas settles into a thinner layer before forming stars. Models of this type may also offer a greater possibility of explaining the Hubble sequence on the basis of different initial conditions without additional ad hoc assumptions about different clumping factors (see below).

The Hubble Sequence

The disc/bulge ratio of a galaxy must depend on the initial conditions and boundary conditions for the collapsing protocloud from which it formed. The model calculations mentioned above suggest two ways in which the disc/bulge ratio could be influenced by the initial conditions. First, if the SFR varies as a power of the gas density ρ_g^n where $n > 1.5$, then the ratio of star formation time to free-fall time increases as the initial density is decreased, so one might expect that models with larger boundary radii and lower initial densities would develop less bulge and more disc. This expectation is confirmed by the model results; for example, in two models calculated by Larson (1976) using equation (11) for the SFR, the fractional mass in the disc component increases from ~ 0.25 to ~ 0.45 when the boundary radius is increased from 30 to 100 kpc. Secondly, if the SFR increases with increasing velocity dispersion of the gas, the disc/bulge ratio will depend on the initial velocity dispersion of the protogalaxy, increasing with decreasing initial velocity dispersion. Since a protogalaxy of larger radius and lower density has a smaller velocity dispersion, this effect acts in the same direction as the dependence on initial density, and again suggests that diffuse protogalaxies should form galaxies with large disc/bulge ratios. In two models calculated using equation (14) for the SFR, the fractional disc mass increases from $\sim$ 0.40 to $\sim$ 0.60 when the radius is increased from 30 to 100 kpc.

These model results still do not reproduce the complete range of apparent disc/bulge ratios in galaxies, ranging from pure spheroidal to almost pure disc systems, but are nevertheless suggestive that two of the principal factors determining the Hubble type of a galaxy are the density and the velocity dispersion of the initial protogalaxy. We therefore expect galaxies in which the spheroidal component is dominant to be formed in regions of unusually high density and/or high velocity dispersion, and galaxies in which the disc component is dominant to form in regions with lower density and/or less turbulence. These predictions are in general agreement with the observed location of elliptical galaxies mainly in dense clusters with large velocity dispersions, and spiral galaxies mostly in the field or in small loose groups where the density and velocity dispersion are low. Among clusters of galaxies, there are also correlations between galaxy type and cluster properties (Oemler 1974): the proportion of elliptical galaxies is smallest in clusters which are relatively loose and irregular, and largest in the massive, dense, centrally concentrated clusters; spiral galaxies, on the other hand, are relatively deficient in the densest clusters and are lacking entirely in their dense cores. Although these correlations could be due in part to the stripping of gas from spirals in dense clusters, which might turn them into S0 galaxies, it is notable that even among the gas-free E and S0 galaxies there is a greater proportion of E galaxies in the denser clusters (Oemler 1974). These correlations are as would be expected if the star formation rates and final disc/bulge ratios in galaxies depend on the initial density and velocity dispersion in the way discussed above. (It is true that the present conditions in clusters of galaxies are probably not the same as the conditions existing at the time of galaxy formation, because in many cases the clusters have probably collapsed after the galaxies formed; nevertheless, it is likely that the dense, massive clusters originated as regions with higher than average density and internal motion, so the same qualitative correlations would still be expected.)

The above qualitative interpretation of the Hubble sequence does not directly involve the angular momentum of the collapsing protogalaxies. It is possible that a correlation between angular momentum per unit mass and Hubble type could arise indirectly from the proposed dependence of Hubble type on initial density, since a more diffuse and extended protogalaxy is likely to possess more angular momentum; this would be the case, for example, if characteristic rotational velocities scale approximately with the virial velocity. However, there does not appear observationally to be a close correlation between Hubble type and specific angular momentum (Brosche 1973).

Continuing Infall of Gas in Spiral Galaxies

If the above general picture relating morphological type to initial density is correct, proto-spiral galaxies are more diffuse and have longer collapse times than proto-elliptical galaxies; in other words, proto-spirals represent smaller density perturbations on the expanding cosmological background. We might therefore expect the evolution of proto-spiral galaxies to depend more strongly on interactions with the ambient medium, i.e. on the boundary conditions for the collapsing protogalaxy. In particular, we might expect that for a protogalaxy which forms in relative isolation in a low density region, the outer boundary will expand with time, and material farther and farther from the center will continue to expand for longer and longer periods of time before finally turning around and falling back into the galaxy, which may be long after the inner part of the system has collapsed. Such continuing infall of matter occurs in the infall models of Gott (1975) and the expanding boundary models of Larson (1969, 1974b), and was also predicted by Oort (1970) in a discussion of the formation of spiral galaxies. Thus, in addition to differences in the initial conditions, another fundamental distinction between spiral and elliptical galaxies may be the continuing infall of outlying matter for a relatively long period of time in spiral galaxies.

We recall that simple collapse models with fixed boundaries yield disc/bulge ratios which increase with decreasing initial density, but still do not span the entire observed range from pure bulge systems (E galaxies) to almost pure disc systems. However, if proto-spiral galaxies are surrounded by diffuse material which does not participate in the initial rapid collapse but falls in later over a more extended period of time, and if star formation does not occur in this infalling gas until it has condensed into a disc, then this could contribute to the formation of massive disc components in spiral galaxies, and thus help to explain the variation in disc/bulge ratio along the Hubble sequence. The conclusion that the formation of disc systems must involve some continuing infall of outlying gas seems almost inevitable, in view of the fact that even the idealized collapse models of Larson (1976) which start with a uniform density protogalaxy show that the disc is built up gradually over several orbital periods by the continuing infall of gas from the outer parts of the protocloud.

In some of the "expanding boundary" models of Larson (1969, 1974b), infall of gas continues at a quite significant rate of the order of 1 or 2 $M_{\odot}$ yr^{-1} even after 10^{10} years. The rates of star formation in these models are essentially governed by the infall rate, and are comparable with those in many spiral galaxies, as inferred from their UBV colors (Larson & Tinsley 1974). This suggests the possibility that the observed gas contents and star formation rates in spiral galaxies are maintained by the continuing infall of gas. The possibility that some spiral galaxies, including our own, may still be accreting gas from an ambient intergalactic medium or remnant protocloud at a rate of the order of $\sim$ 1-2 $M_{\odot}$ yr^{-1} has been discussed by Oort (1970) and Larson (1972a). The question of extragalactic infall in our galaxy is still debated, but the two main competing interpretations of the "high velocity hydrogen clouds" (Hulsbosch 1975, Verschuur 1975) share the basic feature that some hydrogen outside the conventional boundaries of the galactic disc possesses a component of velocity directed into the disc; this is most clearly seen near the galactic poles and in the anticenter direction, where galactic rotation produces no radial component of velocity but negative velocities are seen. Thus there is at least some evidence that the infall process which formed our galaxy is still going on, although it is difficult to estimate the infall rate or determine its importance for present conditions in the galactic disc.

6. EVOLUTION OF GALAXIES

Evolution of Elliptical Galaxies

Although the mass distribution in elliptical galaxies is determined only by the initial collapse process, the interpretation of other observed properties such as the gas content (if any) and the photometric properties of E galaxies requires a knowledge of their subsequent evolution. After the initial collapse, the gas shed by evolving stars becomes a major contributor to the gas content of the galaxy, so to study the evolution of the gas content it is necessary to follow the evolution of stars of different masses and calculate the amount of gas lost when they die. Also, to predict the distribution of metal abundances in the stars that form during the initial collapse and subsequent evolution, it is necessary to calculate the amount of heavy elements returned to the interstellar gas by stars of different masses when they die. Finally, a prediction of the evolution of the photometric properties of a galaxy requires a detailed knowledge of the stellar content and its evolution with time. The problems of stellar and chemical evolution in galaxies have been extensively reviewed by Tinsley in this volume; here we consider only some results that relate to the presently observed properties of E galaxies.

In the dynamical models of Larson (1974b) for the evolution of spherical galaxies, most of the star formation takes place during the first $\sim 10^9$ years, and the gas content and star formation rate subsequently decline rapidly. These models assume that all of the gas released by evolving stars is retained in the galaxy and forms new stars; after 10^{10} years, the star formation rate is essentially equal to the gas production rate, which is about 0.1 $M_\odot$ yr^{-1} for a system of mass 10^{11} $M_\odot$. The gas content at this time depends on the assumed relation between the star formation rate and the gas density, but is less than $\sim 10^8$ $M_\odot$ in all of the models calculated, small enough to escape detection in most surveys of neutral hydrogen in galaxies.

The photometric properties of these models have been calculated by Larson & Tinsley (1974). If a conventional initial stellar mass spectrum similar to that inferred for the solar neighborhood is assumed, the predicted integrated colors and mass/light ratios after 10^{10} years agree satisfactorily with those of typical E galaxies; also, the predicted average stellar metal abundance is about 0.02, a value thought to be typical for E galaxies. This agreement shows that conventional assumptions about star formation and stellar evolution based on our knowledge of the solar neighborhood in our galaxy are consistent with the observed properties of E galaxies. The calculations also show that the predicted small residual star formation rate of $\sim 0.1\ M_\odot\ yr^{-1}$ after 10^{10}yr has an insignificant effect on the integrated colors, and therefore cannot be ruled out on this basis. However, the dynamical models predict that at late times the gas and young stars are concentrated toward the center of the galaxy, causing the nucleus to look bluer than the outer parts of the galaxy, in contradiction to the observations for most E galaxies. This suggests that star formation is not going on at the predicted rate in most E galaxies, and hence that the gas produced by stars is not retained by most E galaxies.

The mechanisms by which most E galaxies are apparently kept clean of gas include supernova-driven hot galactic winds (Mathews & Baker 1971, Faber & Gallagher 1976) and sweeping by passage through an ambient intergalactic medium (Gisler 1976). It seems likely that supernovae can expel the residual gas from at least the smaller E galaxies, beginning at an early stage of evolution (Larson 1974c). However, whether supernova-driven winds are effective in massive ellipticals depends on the uncertain supernova rate and the uncertain efficiency of supernova heating of the gas; according to Gisler (1976), the most massive ellipticals cannot support a hot wind. If other removal mechanisms do not operate, the recycled gas must then presumably accumulate near the center and form new stars or other condensed objects.

Sweeping by intergalactic gas can be effective for galaxies located in clusters which contain a sufficient amount of diffuse hot gas, and which have a large enough velocity dispersion ($\gtrsim$ 1000 km s^{-1}; Gisler 1976). Since many E galaxies are in fact located in dense clusters where sweeping seems likely to occur, this may be at least as important a mechanism as supernova-driven winds in explaining the observed absence of gas and young stars from most E galaxies. One would then expect to see little or no evidence for gas or young stars in E galaxies in clusters, whereas it is possible that E galaxies outside clusters might sometimes contain gas and young stars; this prediction is consistent with the observed fact (van den Bergh 1976b) that very few cluster ellipticals show any evidence for gas, while there are some ellipticals in the field or in loose groups which show evidence for gas, dust, and young stars (eg., NGC 5128). Also, if a cluster contains a central dominant E galaxy, it is presumably moving slowly, if at all, with respect to the cluster gas and so cannot experience sweeping, but might instead accumulate gas and form new stars.

This prediction is again consistent with the evidence for gas, young stars, or nuclear activity in many central cluster E galaxies (eg., NGC 1275). It is noteworthy that NGC 1275 is a Seyfert galaxy, and that many central cluster galaxies are radio sources; this suggests that nuclear activity is associated with the accumulation of gas in galactic nuclei (see below).

Evolution of Spiral Galaxies

We have seen that, in comparison with elliptical galaxies, proto-spiral galaxies are likely to have relatively low initial densities and relatively long time scales for infall of the outermost parts of the initial protocloud. In some of the spherical models with expanding boundaries, which may be more appropriate for the formation of spiral galaxies than elliptical galaxies, infall of primordial gas continues at a significant rate of the order of 1 $M_\odot$ yr^{-1} even after 10^{10} years (Larson 1974b). Thus, for spiral galaxies the formation process probably takes longer than for ellipticals, and the star formation rate remains appreciable for a longer period of time because of continued infall of gas. In addition, even after most of the primordial gas has fallen into a disc, the time scale for turning this gas into stars may still be very long, especially in the outermost parts of the disc where the low gas density, small velocity dispersion, and long dynamical time scale all lead to a low star formation rate. For example, in the collapse models of Larson (1976), which do not have expanding boundaries or a particularly extended time scale for infall of the outermost regions, the time scale for star formation in the outer parts of the disc is still quite long, and typically exceeds 10^{10} yr at radii r $\gtrsim$ 15 kpc. An increase with radius in the time scale for star formation in the disc is expected in any reasonable disc model, and qualitatively similar results are obtained, for example, in the disc models of Talbot & Arnett (1975), which relate the star formation time scale to a local dynamical time scale for the gas.

Because of the slow rate of gas consumption in the outer parts of the disc, the spiral galaxy models of Larson (1976) still have a significant gas content after 10^{10} years (several percent of the total mass), and a star formation rate at this time that is of the order of 1 $M_{\odot}$ yr^{-1}. This predicted star formation rate is comparable with that of a typical early-type spiral galaxy, as inferred from UBV colors; according to Larson & Tinsley (1974), for example, a typical Sa galaxy of mass 10^{11} $M_{\odot}$ has a star formation rate of $\sim$ 1 $M_{\odot}$ yr^{-1}. Compared with other spiral galaxies, the Sa galaxies have less gas and lower star formation rates; in the classification scheme of van den Bergh (1976b), many of them are classified as "anaemic" or gas-poor spirals. Thus models without expanding boundaries or continuing gas infall seem to be able to reproduce the properties of gas-poor spirals with low rates of star formation, but not those of gas-rich systems such as typical Sc spirals which have star formation rates of $\sim$ 5 $M_{\odot}$ yr^{-1} (Larson & Tinsley 1974). For the latter systems, it is appealing to speculate that continuing gas infall from an extended protocloud or accretion from an ambient medium (see below) may contribute to explaining the high gas content and high star formation rate. In this regard, it seems significant that some particularly gas-rich galaxies with high rates of star formation, such as NGC 1068 (M77), NGC 5236 (M83), and NGC 3034 (M82) (see below) are embedded in extensive hydrogen envelopes which extend to several times the optical radius of the galaxy.

At the other extreme, it is also necessary to understand the SO galaxies, which overlap with normal spirals in their basic structural properties, yet in most cases possess little or no gas (Gallagher et al. 1975) and little or no visible indication of a young stellar population. None of the collapse models of Larson (1976) with respectable disc components predict as low a gas content or star formation rate after 10^{10} yr as are found in typical SO galaxies; star formation in the outer parts of the models is always so slow that a significant content of gas and young stars remains even after 10^{10} yr. Even apart from the possible inadequacies of the models, there is the more basic problem that the amount of gas shed by evolving stars is enough to violate observational limits in many cases; thus a gas removal mechanism is called for, just as in the case of elliptical galaxies.

Whether a supernova-driven wind can expel the gas from a spiral galaxy is not known; according to Faber & Gallagher (1976), hot winds should operate effectively in the bulge components of spirals, just as in elliptical galaxies, but should be less effective in discs because of the smaller velocity dispersion. Thus it is possible that some SO galaxies, at least those with small disc/bulge ratios, may lose their gas through hot galactic winds. However, the fact that S0 galaxies are more abundant in clusters than in the field suggests the importance also of environmental effects, such as sweeping by intergalactic gas; presumably this mechanism is at least as effective for S0 galaxies as for E galaxies. According to Biermann & Tinsley (1975), the observed colors of typical S0 galaxies are consistent with star formation having proceeded at a "normal" rate until any time up to only $\sim$ 1-2 x 10^9 yr ago, so that in some cases the loss of gas from S0 galaxies may have been a relatively recent event in their evolution.

Interactions between Galaxies and Intergalactic Gas

While there may be some correlation between the gas content of galaxies and the disc/bulge ratio (Faber & Gallagher 1976), this correlation is not perfect since pairs of galaxies can be found which have the same disc/bulge ratio but quite different gas contents. Van den Bergh (1976b) has proposed a two-dimensional classification scheme in which disc/bulge ratio and gas content are treated as independent parameters; as van den Bergh points out, these parameters need not be closely correlated, since the disc/bulge ratio of a galaxy depends on its "genetic heritage", whereas its gas content may be strongly affected by "environmental factors". We have seen that galaxies may exchange gas with their surroundings through several mechanisms: the gas content of a galaxy may be maintained or increased by inflow of gas from a remnant protocloud or intergalactic medium, and galaxies may lose gas by supernova-driven hot winds or through sweeping by rapid motion through an intergalactic medium.

Thus the gas content of a galaxy may depend in some detail on interactions between the galaxy and its surroundings, and on which processes dominate.

We consider first the conditions under which a galaxy may accrete a significant amount of gas from an ambient intergalactic or circumgalactic medium. The accretion rate for a galaxy of mass M moving with velocity V through a uniform intergalactic medium in which the sound speed is c is given approximately by

$$\frac{dM}{dt} \cong \frac{4\pi G^2 M^2 \rho}{(V^2+c^2)^{3/2}} \tag{15}$$

(eg. Hunt 1971). Thus a high accretion rate is favored by a high ambient density ρ and by a low relative velocity V and a low temperature of the ambient gas. For illustration, a possible set of parameters is $\rho \sim 10^{-4}$ atoms cm^{-3}, $V^2 + c^2 \sim (200 \text{ km s}^{-1})^2$ (Larson 1972a); this gives an accretion rate of $\sim 2\ M_\odot\ yr^{-1}$ for a galaxy of $1.5 \times 10^{11}\ M_\odot$, approximately the same as the infall rate estimated by Oort (1970) for our galaxy. Thus accretion may be important if a galaxy is surrounded by gas of density substantially higher than the mean cosmological density, and if the velocity of the gas particles with respect to the galaxy is not too great. We note that a density of $10^{-4}\ cm^{-3}$ corresponds to a free-fall time of 5×10^9 yr, and that a velocity of 200 km s^{-1} is a typical internal or orbital velocity for a galaxy in a small group, so that the parameters in the above example would not be unreasonable for any leftover protogalactic gas in the vicinity of a galaxy that may have stopped expanding and begun to recollapse only recently.

The effectiveness of sweeping or ablation by motion through an intergalactic medium depends on the dynamical pressure ρV^2 of the intergalactic gas, and also on the density of gas inside the galaxy. Gisler (1976) finds for spherical galaxy models that ablation is not important if V is smaller than the escape velocity of the galaxy (typically $\sim$ 500 km s^{-1}), but that it can sweep out most or all of the gas if V $\gtrsim$ 1000 km s^{-1}, provided that the gas production rate is not much higher than that thought to be typical for old stellar systems; otherwise, accretion may occur instead of sweeping.

In summary, we expect that for galaxies in the field or in small groups where velocities are less than $\sim$ 500 km s^{-1}, sweeping will not be important and accretion may be the dominant process, particularly for galaxies with velocities $\lesssim$ 200 km s^{-1}. On the other hand, in giant clusters where the velocity dispersion is $\gtrsim$ 1000 km s^{-1} and the gas temperature is also very high, accretion cannot occur, and most galaxies are likely to be swept clean by ablation. These predictions are in qualitative agreement with the fact that the most gas-rich galaxies seem to be located in the field or in small loose groups, whereas the galaxies in dense clusters are mostly gasless systems. Van den Bergh (1976b) also points out that the spirals in the Virgo cluster are somewhat gas-poor compared to those in the field, and suggests that this may be a result of partial removal of their gas by sweeping. An alternative possible explanation of this observation is that many field spirals may have diffuse extended gas envelopes which have not yet collapsed, and from which they continue to slowly accrete gas.

If galaxies exchange matter with their surroundings, it is evident that studies of the evolution of galaxies cannot treat them as closed, isolated systems. In particular, the chemical evolution of a galaxy can be strongly affected by such gas exchanges. If a galaxy loses gas at an early stage of evolution, this will limit the amount of heavy element enrichment that can take place, whereas if a galaxy continues to accrete gas, this tends to make the metal abundance rise to an asymptotic value which is nearly independent of time, and reduces the proportion of relatively metal-poor stars (Tinsley, this volume). Additional complications for chemical evolution are possible if galaxies lose metal-rich gas through winds or sweeping and thereby enrich the surrounding intergalactic medium, and if some of this metal-enriched intergalactic gas later falls back again into galaxies (Larson & Dinerstein 1975). The close relation between chemical evolution and gas flows in galaxies means that a complete understanding of the chemical evolution of a galaxy requires a knowledge of its dynamical evolution as well; however, it also means that observed chemical abundances, if properly interpreted, can potentially provide much useful information about the dynamics of the formation and evolution of galaxies.

7. YOUNG GALAXIES AND ACTIVE GALAXIES

Finally, we consider the possible properties of galaxies which are still in the process of formation, and the question of whether such systems are or can be observed. Because of the many uncertainties in the theory of galaxy formation, it will be very important to try to identify and study observationally possible examples of galaxies still in the process of formation. Such galaxies should be distinguishable from normal galaxies in a number of ways: they should have a large and perhaps chaotically distributed content of gas and dust, intense star formation activity, and probably an irregular or peculiar morphological structure indicating a system which has not yet settled into a regular equilibrium configuration. If rapid star formation and supernova production take place under conditions of high density in the nucleus of a galaxy, this may also generate non-thermal nuclear activity (Arons et al. 1975). We shall consider as "active galaxies" those galaxies which display particularly intense star formation activity, or any other indications of galaxy-forming processes still actively taking place.

Primeval Galaxies

From all indications, most galaxies have not formed very recently but have ages that are a substantial fraction of the age of the universe. If quasars are located in galaxies, the existence of quasars with redshifts of at least 3.5 indicates that some galaxies must have formed by the time the universe was only about 1/5 of its present age. Galaxies in the process of formation at significant redshifts are often called "primeval galaxies".

The appearance and observability of a primeval galaxy depend on its redshift as well as its luminosity and spatial structure. The luminosity of a primeval galaxy is produced mainly by massive short-lived stars, and therefore is approximately proportional to the star formation rate; thus the most luminous phase in the evolution of a primeval galaxy occurs when the star formation rate reaches a maximum. In the galaxy formation models of Gott (1973, 1975) and Gott & Thuan (1976), most of the star formation is assumed to take place in a brief initial burst before the system as a whole has collapsed. The models of Larson (1974b, 1975c, 1976) also have rapid early star formation, but the star formation rate continues to increase as the system collapses and the gas density rises, reaching a maximum when a dense core forms with a burst of fast star formation. This peak in the star formation rate typically occurs about 1.5 free-fall times after the beginning of the collapse, or about 2.5 free-fall times after the beginning of cosmological expansion, assuming that the expansion of the protogalaxy to its maximum radius takes one free-fall time.

The collapse times of protogalaxies cannot be predicted with any quantitative certainty, but some rough limits can be set. If a protogalaxy of typical mass 10^{11} $M_\odot$ can be idealized as a uniform sphere with radius between 30 and 100 kpc, the corresponding range of free-fall times τ_{ff} is from 2.7×10^8 to 1.7×10^9 yr, and if we assume that the maximum star formation rate occurs after a time $t_c \sim 2.5\tau_{ff}$, the corresponding range of values of t_c is $7 \times 10^8 \lesssim t_c \lesssim 4 \times 10^9$ yr. With currently popular values of the cosmological parameters ($H = 50$ km s^{-1} Mpc^{-1}, $\Omega = 0.1$) the corresponding range of redshifts is $3 \lesssim z_c \lesssim 20$. These estimated values of z_c should be considered as upper limits, since the radii and collapse times of protogalaxies could have been larger than the minimum values assumed above. On the other hand, shorter collapse times could occur if the masses of galaxies are greater than the conventional values. Thus, while typical collapse redshifts may lie in the range $3 \lesssim z_c \lesssim 20$, values outside this range are entirely possible. In particular, since no real lower limit can be set on z_c, it is possible that some galaxies are still forming now (i.e., are currently experiencing the phase of maximum luminosity).

The maximum luminosity attained by a primeval galaxy depends on the assumed model, but can be roughly estimated if the free-fall time τ_{ff} is known since, at least in proto-elliptical galaxies, a major fraction of the star formation and of the total nuclear energy production must occur within roughly one free-fall time. If a galaxy of mass 10^{11} $M_\odot$ forms in 10^9 yr, the average star formation rate during this period is of the order of 100 $M_\odot$ yr^{-1}, and it can be estimated that if stars form at this rate with a normal mass spectrum, the luminosity produced by massive stars is of the order of $3 \times 10^{11} L_\odot$ (Larson 1974b). A similar estimate is obtained by assuming that the observed heavy element abundance of $Z \sim 0.02$ is produced by nuclear burning of hydrogen over a period of 10^9 yr. The results of model calculations are consistent with these simple estimates: various models (Larson 1974b, 1975c, 1976) all have peak star formation rates in the range of $\sim$ 100 - 300 $M_\odot$ yr^{-1}, corresponding to peak luminosities of $\sim 3 \times 10^{11} - 10^{12}$ $L_\odot$. These luminosities are of the order of 30 to 100 times the typical present luminosity of a galaxy of mass 10^{11} $M_\odot$, and they extend into the luminosity range of the quasars. Thus it may be that some primeval galaxies, particularly the more massive and luminous ones, are observable in the magnitude and redshift range of quasars.

The observable properties of primeval galaxies have been discussed in more detail by Meier (1976b). He concludes that typical primeval galaxies of mass $\sim 10^{11}$ $M_{\odot}$ may have redshifts in the range $2 \lesssim z \lesssim 16$, apparent magnitudes in the range $21 \lesssim m \lesssim 29$, nearly stellar images with half-light diameters between 0.5 and 2 arc sec, and UBV colors which are a sensitive function of redshift. Such objects would have escaped detection in the searches which have been made for primeval galaxies but might possibly appear as very faint semi-stellar objects near the limit of detectability in deep surveys. Better prospects for detection would, of course, be offered by the atypical primeval galaxies which have larger masses and luminosities, or by those which form at later times and smaller redshifts.

The fact that primeval galaxies probably overlap with quasars in magnitude and redshift and probably have stellar or near stellar images raises the question of whether some (or all) of the quasars may in fact be very young or primeval galaxies. As was noted by Schmidt (1972), the space density of quasars at redshift 2.5 is comparable with the present space density of giant E galaxies, so that the quasars could represent an early stage in the evolution of giant E galaxies. The non-thermal emission and rapid variability characteristic of many quasars are clearly not explained by a population of young stars alone, but if star formation and supernova production occur at a high rate in a small dense nucleus, the effects produced by the rapid formation of pulsars in close proximity may be able to explain the principal properties of the non-thermal emission of quasars (Arons et al. 1975). Thus quasars may represent an early stage of galaxy evolution when intense star formation is occurring or has recently occurred in the nucleus of a galaxy. Since the nuclear energy which is radiated thermally by the massive stars is of the same order as the gravitational energy emitted non-thermally by the pulsars (Larson 1974b), one would expect that at least some quasars should show evidence for thermal emission from hot stars and HII regions. In fact, some quasars have spectra which resemble those of hot stars, rather than the typical non-thermal quasar spectrum; these thermal-type spectra become more predominant with increasing redshift, as would be expected if the "thermal" quasars represent earlier stages of evolution. Two such objects at redshifts of 2.9 and 3.4 have spectra which match closely the predicted spectra of primeval galaxies (Meier 1976a), so it may well be that some primeval galaxies are observed as high redshift quasars.

Possible Nearby Young Galaxies

Finally, we consider the possibility that young galaxies or galaxies which are still forming may exist nearby at small redshifts. (By "young galaxy" we shall mean a galaxy in which a major fraction of the star formation has taken place relatively recently, since it is never possible to rule out the presence of a few old stars in a galaxy.) Presumably such young galaxies, if they exist, will not yet have achieved the regular structure characteristic of old galaxies but will instead be among the more irregular or peculiar galaxies. Exactly what they might look like is difficult to predict in detail, but a common feature should be intense star formation activity, probably taking place nonuniformly in localized regions or dense knots.

We consider first some possible examples of elliptical galaxies in which star formation is still actively going on. According to the collapse models, the outer parts of an E galaxy form first, and the residual gas then settles toward the center or toward the equatorial plane where active star formation continues for a longer period of time. Thus we might expect that during the later stages of the formation process, an elliptical galaxy already has an outer envelope of approximately spheroidal shape, while star formation continues in the inner regions, perhaps in an irregular distribution of clumps of gas and young stars. A galaxy which matches this description is the small peculiar elliptical galaxy NGC 5253 which, in addition to containing knots of gas and young stars in its interior, has produced two supernovae. Other similar objects may include NGC 185 and NGC 205 in the local group, both of which contain patches of dust and young stars, and NGC 3077 in the M81 group, which has a large gas content, a chaotic dust pattern, and concentrations of young stars.

Also interesting as examples of E galaxies with active star formation are the two large and very peculiar systems NGC 5128 and NGC 1275 (van den Bergh 1976b). NGC 5128 appears to be an E0 galaxy containing a broad, chaotic band or layer of dust, gas, and young stars orbiting around its center; possibly this represents the early stage of a disc component, seen just now in the process of formation from infalling gas. There is also evidence for patches of dust, young stars, and HII regions far outside the equatorial plane, some of which may represent material falling into the galaxy. NGC 1275 also shows extensive dust lanes, and photographic as well as spectroscopic evidence for young stars distributed in a filamentary pattern. Both NGC 5128 and NGC 1275 have active nuclei, NGC 5128 being a radio galaxy (Cen A) and NGC 1275 a Seyfert galaxy as well as a radio source (Per A).

A question of interest is whether the gas in these galaxies is of internal or external origin, and in some cases circumstantial evidence suggests an external origin. NGC 3077, along with M82 which it resembles, is embedded in an extensive hydrogen cloud surrounding the large spiral M81, and it therefore seems most likely that the peculiarities of both NGC 3077 and M82 are due to interactions with the surrounding gas cloud, such as recent accretion of matter. Both NGC 5253 and NGC 5128 are located in the same loose group as M83, a particularly gas-rich spiral with an extended hydrogen distribution; thus we might again suspect that interactions with the surroundings are involved, and we may be seeing the later stages of the formation of galaxies by the infall of residual gas from an ambient medium or remnant protocloud.

Earlier stages in the formation of E galaxies may be more difficult to identify, since we have less idea what to expect. The gas-dynamical collapse models suggest that at a very early stage a galaxy may consist of a chaotic and clumpy distribution of gas, dust, and young stars, while the collision models of Toomre (1974, 1976) suggest that E galaxies may begin as pairs or small groups of galaxies, probably spirals, which collide and merge. Many systems with a very chaotic or distorted appearance are seen in the Atlas of Peculiar Galaxies (Arp 1966), and it is possible that some of these will eventually evolve into normal elliptical galaxies and can therefore be considered proto-elliptical galaxies. However, some may also eventually evolve into normal spiral galaxies, especially if a large amount of gas is present, so it is difficult to associate such systems with any particular Hubble type. A more careful study of the possible evolutionary relations between various peculiar and "normal" types of galaxies might prove enlightening in this regard. For example, Toomre (1976) has put together a sequence of photographs of objects in the Arp atlas which may illustrate successive stages in the merging of two spiral-like galaxies into a single elliptical-like galaxy. Detailed studies of the kinematics and stellar and gas content of such systems should help to clarify such possible evolutionary sequences.

According to the collapse models of Larson (1976), the disc components of spirals take much longer to form than the spheroidal components, and it is possible that the outer parts of spiral galaxies may still be slowly settling down into a thin disc even after 10^{10} yr. Also, infall of gas from an extended remnant proto-cloud may continue for a long period of time in many spiral galaxies. Thus it is possible that many spiral galaxies are dynamically young in the sense that their outermost parts are still slowly forming. In this regard it is perhaps noteworthy that spiral galaxies frequently have extended hydrogen distributions, warps, and non-circular motions in their outer parts; all of these characteristics seem to indicate gas which has not yet settled into a plane layer with circular motion. A particularly interesting example of a gas-rich and possibly young spiral is M83, which has a hydrogen distribution extending out to several optical radii, a strongly warped hydrogen layer (Rogstad et al. 1974), and anomalously blue colors and a high supernova rate which indicate a high recent rate of star formation (Tinsley 1975). M83 belongs to the same loose group as NGC 5253 and NGC 5128; also in this group is NGC 5102, an anomalous SO galaxy with blue colors and a high gas content (Gallagher et al 1975). All this evidence suggests that M83 is young in the sense that a large amount of primordial gas is still settling into a disc and forming stars at a high rate.

Perhaps the most striking and most intensively studied peculiar galaxy is M82, which shares a common hydrogen envelope with M81 and NGC 3077. According to recent interpretations of this object, the filaments that appear to radiate out from the center contain dust which reflects light from luminous HII regions located near the center of the galaxy but obscured from view by dust; in this interpretation, the velocities of the filaments are modest, and the system is not exploding violently as once thought. Instead, much evidence now points to the occurrence of an intense burst of star formation and supernova production near the center of this galaxy, most of this activity being optically obscured by dust (eg., Hargrave 1974). This may have been caused by recent infall of new gas from the surrounding hydrogen cloud. (The very dusty appearance of M82 is not necessarily inconsistent with this possibility, since the gas around M82 may have been enriched in heavy elements by the loss of enriched gas from galaxies in the M81 group (Larson & Dinerstein 1975).) The nature of the dusty filaments is still not understood, but a number of recent attempts have been made to explain them in terms of interactions between M82 and the surrounding medium. A speculative possibility is that they represent relatively dense, cool infalling gas which is concentrated into filaments by interaction with an outflowing hot wind generated by supernovae near the center of M82. All of the properties of M82 could then be understood as a result of recent or ongoing infall of gas and the occurrence of intense star formation activity near the center. Whether or not M82 is presently accreting matter, the present star formation activity in M82 is probably similar to the activity which takes place in galaxies at early stages of evolution, so it may be hoped that by studying the activity in M82 and similar galaxies we can learn much about the processes by which galaxies form.

REFERENCES

Arons, J., Kulsrud, R. M., & Ostriker, J. P. 1975. Astrophys. J., 198, 687.

Arp, H. C. 1966. Atlas of Peculiar Galaxies (California Institute of Technology); also Astrophys. J. Suppl., 14, 1.

Biermann, P., & Tinsley, B. M., 1975. Astr. Astrophys., 41, 441.

Black, D. C., & Bodenheimer, P., 1976. Astrophys. J., 206, in press.

Brahic, A., 1975. Dynamics of Stellar Systems, IAU Symp. No. 69, ed. A. Hayli, p. 287. (D. Reidel, Dordrecht).

Brosche, P., 1970. Astr. Astrophys., 6, 240.

Brosche, P., 1973. Astr. Astrophys., 23, 259.

Cox, D. P., & Smith, B. W., 1974. Astrophys. J. Lett., 189, L105.

Faber, S. M., & Gallagher, J. S., 1976. Astrophys. J., 204, 365.

Gallagher, J. S., Faber, S. M., & Balick, B. 1975. Astrophys. J., 202, 7.

Gott, J. R., 1973. Astrophys. J., 186, 481.

Gott, J. R., 1975. Astrophys. J., 201, 296.

Gott, J. R., & Thuan, T. X., 1976. Astrophys. J., 204, 649.

Gutowski, W. J., & Larson, R. B., 1976. Publ. Astr. Soc. Pacific, 88, in press.

Hargrave, P. J., 1974. Mon. Not. R. Astr. Soc., 168, 491.

Hénon, M., 1973. Dynamical Structure and Evolution of Stellar Systems, Third Advanced Course of the Swiss Society for Astronomy and Astrophysics, p. 183. (Geneva Observatory).

Hulsbosch, A. N. M., 1975. Astr. Astrophys., 40, 1.

Hunt, R., 1971. Mon. Not. R. Astr. Soc., 154, 141.

Jones, B. J. T., 1976. Rev. Mod. Phys., 48, 107.

Kahn, F. D., 1974. Astr. Astrophys., 37, 149.

King, I. R., 1966. Astr. J., 71, 64.

King, I.R., 1976. In preparation.

Larson, R. B., 1969. Mon Not. R. Astr. Soc., 145, 405.

Larson, R. B. 1970. Mon Not. R. Astr. Soc., 147, 323.

Larson, R. B., 1972a. Nature, 236, 21.

Larson, R. B., 1972b. Mon. Not. R. Astr. Soc. 156, 437.

Larson, R. B., 1973. Mon Not. R. Astr. Soc., 161, 133.

Larson, R. B., 1974a. Fund. Cos. Phys., 1, 1.

Larson, R. B., 1974b. Mon. Not. R. Astr. Soc., 166, 585.

Larson, R. B., 1974c. Mon. Not. R. Astr. Soc., 169, 229.

Larson, R. B., 1975a. Problèmes d'Hydrodynamique Stellaire, 19th Liège Astrophysical Symposium, p. 451. (Société Royale des Sciences de Liège).

Larson, R. B., 1975b. Dynamics of Stellair Systems, IAU Symp. No. 69, ed. A. Hayli, p. 247. (D. Reidel, Dordrecht).

Larson, R. B., 1975c. Mon. Not. R. Astr. Soc. 173, 671.

Larson, R. B., 1976. Mon. Not. R. Astr. Soc. 176, in press.

Larson, R. B., & Dinerstein, H. L., 1975. Pub. Astr. Soc. Pacific, 87, 911.

Larson, R. B., & Starrfield, S., 1971. Astr. Astrophys., 13, 190.

Larson, R. B., & Tinsley, B. M., 1974. Astrophys. J., 192, 293.

Liller, M. H., 1960. Astrophys. J., 132, 306.

Liller, M. H., 1966. Astrophys. J., 146, 28.

Loren, R. B., 1976. Astrophys. J., in press.

Lynden-Bell, D., 1967. Mon. Not. R. Astr. Soc., 136, 101.

Mathews, W. G., & Baker, J. C., 1971. Astrophys. J., 170, 241.

McCrea, W. H., 1960. Proc. Roy. Soc. A, 256, 245.

Meier, D. L., 1976a. Astrophys. J. Lett., 203, L103.

Meier, D. L., 1976b. Astrophys. J., in press.

Oemler, A., 1974. Astrophys. J., 194, 1.

Oemler, A., 1976. Astrophys. J., in press.

Oort, J. H., 1965. Galactic Structure, ed. A. Blaauw & M. Schmidt, p. 455. (University of Chicago Press).

Oort, J. H., 1970. Astr. Astrophys., 7, 381.

Richstone, D. O., 1976. Astrophys. J., 204, 642.

Rogstad, D. H., Lockhart, I. A., & Wright, M. C. H., 1974. Astrophys. J., 193, 309.

Schmidt, M., 1972. Astrophys. J., 176, 303.

Spitzer, L., 1971. Nuclei of Galaxies, ed. D. J. K. O'Connell, p. 443. (North-Holland, Amsterdam).

Strom, S. E., Strom, K. M., & Grasdalen, G. L., 1975. Ann. Rev. Astr. Astrophys., 13, 187.

Strom, S. E., Strom, K. M., Goad, J. W., Vrba, F. J., & Rice, W., 1976. Astrophys. J., 204, 684.

Talbot, R. J., & Arnett, W. D., 1975. Astrophys. J., 197, 551.

Tinsley, B. M., 1975. Publ. Astr. Soc. Pacific, 87, 837.

Toomre, A., 1974. The Formation and Dynamics of Galaxies, IAU Symp. No. 58, ed. J. R. Shakeshaft, p. 347. (D. Reidel, Dordrecht).

Toomre, A., 1976. In preparation.

Tremaine, S. D., 1976. Astrophys. J., 203, 345.

Tremaine, S. D., Ostriker, J, P., & Spitzer, L., 1975. Astrophys. J., 196, 407.

van den Bergh, S., 1976a. Astrophys. J., 203, 764.

van den Bergh, S., 1976b. Tercentenary Symposium of the Royal Greenwich Observatory, in press.

van Houten, C. J., 1961. Bull. Astr. Inst. Netherlands, 16, 1.

Verschuur, G. L., 1975. Ann. Rev. Astr. Astrophys., 13, 257.

Werner, M. W., Elias, J. H., Gezari, D. Y., & Westerbrook, W. E., 1974. Astrophys. J. Lett., 192, L31.

Wilson, C. P., 1975. Astr. J., 80, 175.