PART I

FORMATION AND EVOLUTION
OF
DISC GALAXIES

FORMATION OF SPIRAL GALAXIES

Richard B. Larson

Yale University Observatory

1. INTRODUCTION

Conventionally, two major components of spiral galaxies are dis-
tinguished: (1) a "spheroidal component", which includes the central
bulge and the surrounding halo, and (2) a "disc component", which is
basically flat and circular but is often warped at the edges. Other
structures, such as bars or rings, may also be present, but they are
probably of less fundamental significance for the processes of galaxy
formation, and therefore they will not be considered here. The dis-
tinction between bulge and disc components corresponds approximately to
the classical distinction between Population II, as exemplified by the
globular clusters which have a nearly spherical distribution, and
Population I, which is confined to the galactic plane and is relatively
young and metal-rich. However, unlike the "classical" Population II
objects which are old and metal-poor, the inner part of the bulge of our
galaxy and of other galaxies is in fact metal-rich, and may even contain
some young stars.

It is widely accepted, at least as a working hypothesis, that the
spheroidal components of spiral galaxies are similar in all respects to
elliptical galaxies of comparable size and mass; present data do not
contradict this hypothesis. A spiral galaxy may then be described as
an elliptical galaxy plus a disc component, whose properties and origin
are not necessarily closely related to those of the elliptical component.
We can therefore conveniently discuss separately the formation of the
spheroidal and disc components of spiral galaxies, keeping open for the
time being the question of the relation between them.

2. GRAVITATIONAL COLLAPSE OF PROTOGALAXIES

Since it is widely believed that galaxies form by some sort of
gravitational collapse or accumulation process, it is useful to consider
first some general properties of gravitational collapse. Numerical

calculations always result in a very nonuniform collapse characterized
by the runaway growth of any density fluctuations, whether they are
present initially or whether they propagate inward from the boundary.
Thus a collapsing cloud may be expected to develop a very lumpy
structure, with dense condensations embedded in a more diffuse medium.
This is indicated in some recent calculations of 3-dimensional iso-
thermal collapse made with a finite-particle scheme (Larson 1978a,b);
the number of condensations formed is found to be approximately equal
to the number of Jeans masses in the initial cloud, as might be expected.
The condensations rapidly collapse and develop dense cores which then
proceed to accrete the gas around them; in general, because of the
angular momentum of the accreted material, this leads to the formation
of accretion discs around the dense cores.

     In a collapsing protogalaxy, the temperature may be near 10^4K for
a wide range of conditions, implying Jeans masses in the range
$\approx 10^7 - 10^9$ $M_\odot$. Condensations with masses of this order may then be
expected to form and to begin collapsing on their own. After a
sufficient increase in density, rapid cooling will eventually occur and
reduce the temperature in the collapsing condensations to far below
10^4K, greatly reducing the Jeans mass and allowing the formation of
large numbers of stars. Thus the first structures to appear in a
collapsing protogalaxy may be subsystems of gas and stars having the
size of small galaxies. If these subsystems have a sufficient density
contrast, they may develop into well-formed separate galaxies. However,
tidal interactions between them will eventually become important and
cause them to merge into a single larger system, as emphasized in
different contexts by Toomre (1977) and Ostriker (1977). Dynamical
friction will cause the densest parts of the subsystems to spiral
together to form a dense core, while the less dense parts are stripped
off to form a diffuse envelope, producing ultimately a smooth, centrally
condensed, more or less spheroidal system of stars. If the gas which
has not yet turned into stars is retained in the system, it will
eventually dissipate its random motions and settle into a disc.

     Thus one can understand in a general way how galaxies with both
spheroidal and disc components can form: the spheroidal component
forms from stars made in dense subsystems before interactions and
gaseous dissipation processes have become very important, while the
disc forms from whatever gas remains at later times to experience
further dissipation. The processes involved in the formation of
spheroidal and disc systems will be considered in more detail below.
It can be noted here, however, that the processes of tidal disruption
and coalescence mentioned above are predicted to occur eventually for
any gravitationally bound system of subunits, such as a cluster of
galaxies; we still see separate galaxies in bound clusters only because
the timescale for coalescence is longer than the present age of the
universe, but eventually all of the galaxies in a system like the
Local Group (if it is bound) will merge into a single large galaxy.
Many peculiar and interacting galaxies may represent various stages of

such tidal interaction and coalescence processes (Toomre 1977).

3. FORMATION OF SPHEROIDAL SYSTEMS

All of the models that have been proposed for the formation of elliptical galaxies share the common requirement that most of the stars must form early, i.e. before there has been too much interaction between different parts of the system (Larson 1976b, Gott 1977). Rapid early star formation is plausible only in dense condensations or subsystems, so we are led again to the picture of a very nonuniform protogalaxy. An extreme case is represented by the suggestion of Toomre (1977) that elliptical galaxies may often originate from mergers between a small number of disc galaxies in which most of the gas has already been turned into stars. However, since purely stellar models do not directly explain the highly condensed cores or the metallicity gradients frequently observed in elliptical galaxies, it is probably necessary to assume that some gas remains after the initial coalescence of subsystems and forms metal-enriched stars with a more centrally condensed spatial distribution, as in the models of Larson (1975).

Eventually, stellar dynamical effects will smooth out most of the initial lumpiness in the distribution of stars; the larger and more massive subsystems will suffer tidal disruption and coalescence, while the smallest ones will have short relaxation times and thus will soon evaporate. After 10^{10} years only dense clusters of intermediate mass, such as the presently observed globular clusters, can survive (Rees 1977), and even these will eventually disappear. Thus the globular clusters may represent the last surviving remnants of the initially very nonuniform structure of the galactic halo.

It would be of interest to know how many subsystems were initially present in our own galaxy, how much mass they contained, and how star formation and metal enrichment processes occurred in them. The fact that the globular clusters in the outer halo show a great dispersion in metal abundances which is not correlated with position suggests that they formed in many separate "fragments" that evolved independently and achieved different degrees of metal enrichment (Searle 1977). On this picture the spread in metallicities of the globular clusters can be explained explained as a result of statistical fluctuations if star formation occurred in bursts and if each fragment experienced only a few such bursts (Searle 1977). These bursts of star formation and metal enrichment might even be identifiable with the formation of individual globular clusters. Only a small fraction of the total galactic mass may have been involved in the formation of the initial fragments, since only a small fraction of the present visible mass of our galaxy is in the halo.

4. FORMATION OF DISCS

     Disc systems can form only from material that remains in gaseous
form long enough for nearly all of its random motion to be dissipated.
This might include both gas swept out of the original subsystems by
collisions between them, and diffuse outlying gas that falls in later;
a further source of gas may be mass loss from halo stars. Most traces
of the initial nonuniform distribution of the gas are probably destroyed
by the strong dissipative processes associated with the formation of the
disc, so it may not be possible to infer from the present properties of
the galactic disc much about the origin or spatial distribution of the
gas from which it formed. However, the timescale for disc formation,
i.e. the time dependence for the infall of gas into the galactic plane,
is of great importance for galactic evolution, and may bear directly on
the chemical properties of disc stars. In discussing the timescale of
disc formation, it will be useful to consider first two possible limit-
ing models for the formation of discs.

     One limiting case is the rapid collapse of a uniform protogalactic
gas cloud, for example a uniform spheroid of gas that collapses in free
fall to a disc. This is essentially the model proposed by Eggen,
Lynden-Bell, and Sandage (1962, hereafter ELS). In this case all of the
gas arrives in the disc at almost the same time, and the disc thereafter
evolves as a closed system without further addition of mass. If star
formation and metal production occur during the collapse, the progressive
enrichment of the collapsing cloud leads to a spatial gradient in the
average metallicity of the stars formed, such that the metallicity
decreases with increasing height above the galactic plane. Moreover,
since the formation of the disc must be accompanied by the dissipation
of random motions in the gas, the velocity dispersion of newly formed
stars decreases with time, leading to an inverse correlation between
velocity dispersion and metallicity that is in qualitative agreement
with that observed for solar neighborhood stars. The observed correla-
tion between velocity dispersion and metallicity was taken by ELS as
evidence that our galaxy did in fact collapse during its early history.
It may be noted, however, that this interpretation of the observations
can be applied only to stars that actually formed during the initial
rapid collapse, and not to the majority of disc stars which formed later.

     ELS supposed that the initial protocloud was basically uniform and
quiescent, each part having only a circular motion of rotation, and they
argued that the present large orbital eccentricities of some stars can
be explained only if their initially circular orbits were altered by a
rapid change in the galactic potential field caused by a rapid collapse
lasting no longer than $\sim 2 \times 10^8$ yr. If, however, stars were initially
formed with large non-circular velocities, as seems much more likely in
an irregular protogalaxy, this argument no longer applies and the
collapse time could in principle be much longer. Slower collapse and
disc formation could result if dissipation processes are slow, as in
some of the models of Larson (1976a), or if gas continues to fall in

from great distances over a long period of time, as in the "cosmological
infall" picture of galaxy formation (Oort 1970, Larson 1972, Gott 1977).

Thus, as an opposite limiting case, we can consider the possibility
that the disc forms by gradual infall of primordial gas, the infall
continuing at a steady rate throughout the lifetime of the galaxy. In
this case a vertical metallicity gradient is again established because
of the progressive enrichment of the infalling gas as it flows into the
galactic plane. Likewise a correlation between velocity dispersion and
metallicity is predicted because stars formed farther away from the
galactic plane have a higher velocity dispersion. It is important to
note that in the infall picture these correlations result largely or
entirely from spatial gradients in a system with a steady gas flow,
rather than from time variations in the properties of a homogeneous
collapsing cloud, as before. These spatial gradients can even be
independent of time, as is illustrated by some of the results obtained
for spherical collapse models by Larson (1974). Thus it is important
to determine to what extent the observed trends reflect time variations
and to what extent they result from spatial gradients present at a
given time.

Infall or gas flow models possess some advantages over simple
closed models of galactic chemical evolution, since they characteris-
tically predict a slow variation of metallicity with time; this might
explain the fact that the solar neighborhood has fewer metal-poor stars
and a narrower metallicity distribution than is predicted by the
simplest closed models (Audouze and Tinsley 1976). Also, an inward
flow of gas in the galactic disc could contribute to explaining the
radial abundance gradient in the disc (Tinsley, this colloquium).

A variant of this type of model is the halo infall model proposed
by Ostriker and Thuan (1975), in which part of the disc gas is supplied
by mass loss from halo stars. In this case, most of the infall occurs
at early times and is strongly enriched by the nucleosynthesis products
of massive halo stars; later, both the infall rate and the metallicity
of the infalling gas decline as mass loss from less massive stars
becomes the dominant contributor. This type of model has the additional
advantage that the early infall of already enriched gas enhances the
rapid initial rise of metallicity in the disc, and thus helps to
account for the observed paucity of metal-poor stars.

The effects discussed above are all present to varying degrees in
the hydrodynamic collapse models of Larson (1976a), in which star
formation and dissipation take place concurrently in a contracting
protogalaxy, forming first a spheroidal component and later a disc. In
these models the disc is formed by infall at a decreasing rate of both
primordial gas and gas lost from halo stars; halo infall is most
important in the inner part of the disc, while the outer disc forms
mostly from primordial gas which falls in from greater distances and
settles into a disc over a longer period of time. With appropriate

choices of the assumptions and model parameters, particularly those
governing the star formation rate, these models produce a two-component
structure like that of many spiral galaxies, with most of the mass in
the disc component. In addition, they predict a radial metallicity
gradient in at least the inner part of the halo, and a steep metallicity
gradient perpendicular to the plane of the disc, both in qualitative
agreement with observations for our galaxy.

Improved calculations of the chemical evolution of the disc com-
ponents of some of these models have recently been made by Tinsley and
Larson (1978). At all radii, the metallicity Z_g of the gas rises
rapidly at first because of the infall of gas previously enriched by
massive halo stars, as in the Ostriker-Thuan (1975) models. This
initial rapid increase in Z_g is most pronounced at small radii, where
there is a brief "super-metal-rich" peak in metallicity before halo
star formation subsides and Z_g drops again. Later, as disc star
formation becomes important, Z_g rises more slowly and eventually
approaches a nearly constant value at most radii, owing to the continuing
infall of metal-poor gas. The resulting radial gradients in metallicity
for both gas and stars are fairly shallow in the inner disc where the
gas fraction is small, but become steeper in the outer part of the disc
where the gas fraction increases considerably due to slower star forma-
tion. The predicted steepening of the stellar metallicity gradient at
large radii is similar to the observational results of Janes (this
colloquium) for disc stars in our galaxy. The predicted metallicity
gradient of the gas is smaller than that derived by Peimbert (this
colloquium), but is very sensitive to the amount of continuing gas
infall; if infall at late times is suppressed, for example by a galactic
wind, a steeper gradient of Z_g is predicted.

The dynamical models have also been used to predict the distribution
of stellar metallicities and the relation between velocity dispersion
and metallicity in a region comparable with the solar neighborhood
(Tinsley and Larson 1978). The slow decline in the velocity dispersion
of the gas over the lifetime of the disc leads to a relation between
velocity dispersion and metallicity that is in good agreement with the
data of Janes (1975), if a plausible dispersion in metallicity at a
given time is assumed. Also, because of the rapid initial rise in
metallicity followed by a relatively slow continuing increase, the
models predict relatively few metal-poor disc stars, in agreement with
observations.

5. THE TIMESCALE FOR DISC FORMATION

An important feature of those dynamical models that yield the best
agreement with observations is the relatively long timescales involved:
halo formation takes a few times 10^9 yr, and disc formation takes an
even longer time which approaches $\approx 10^{10}$ yr in the outer parts because
of the long time required for the gas to settle into a layer that is

thin and dense enough for significant star formation to occur. For
example, in Model 9 of Larson (1976a), the star formation rate at 15 kpc
from the center does not reach a peak until nearly 10^{10} yr after the
beginning of the collapse. Although the exact timescales depend on
model parameters, these results illustrate the possibility that many of
the properties of the disc of our galaxy could result from slow disc
formation by gradual infall and settling of gas into the galactic plane
over a period of many times 10^9 years. If this is the case, most of
the stars in the outer part of the galactic disc could be considerably
younger than the halo.

     To distinguish between the different possibilities that have been
discussed, and to reconstruct the past history of our galaxy, it is
clearly necessary to have accurate information about the ages of stars
and star clusters. With such information it may eventually be possible
to establish the timescales for halo and disc formation and to decide
to what extent various observed correlations result from secular evolu-
tion and to what extent they result from spatial variations present at
a given time.

     At present it is difficult to determine accurate and reliable ages
for individual stars, and the most reliable age determinations are
those made for clusters. New age determinations for many clusters have
been made by Demarque and McClure (1977) and by McClure (this colloquium).
They find that the ages of the metal-poor globular clusters M92 and M15
are about 15×10^9 yr, the ages of the intermediate metallicity
globulars M13 and M3 are about 14×10^9 yr, and the age of the metal-
rich "nuclear" globular cluster 47 Tuc is about 12×10^9 yr, with
uncertainties of around 10^9 yr. These results suggest a timescale of
$\sim 3 \times 10^9$ yr for the collapse and enrichment of the galactic halo. A
more surprising result is the age of the oldest known disc cluster
NGC 188, which apparently provides an upper limit for the ages of most
(solar neighborhood) disc stars; this age is found to be only (5 ± 1)
$\times 10^9$ yr. If there are any disc stars older than this, they must be
significantly more metal-poor than NGC 188, but there are not many such
metal-poor stars in solar-neighborhood samples. For clusters younger
than NGC 188, there does not appear to have been much (if any) increase
of mean metallicity with time; instead, it appears that at least part
of the variation in metallicity is due to spatial gradients (McClure,
this colloquium).

     These results suggest that the outer disc of our galaxy may be
considerably younger than the halo, at least in the sense that most of
the star formation did not occur until long after the formation of the
halo stars. Possibly the outer disc has formed from gas that fell in
relatively recently from great distances, or possibly inefficient
dissipation has caused the gas in the outer disc to take a long time to
settle into a layer that is sufficiently thin and dense for star forma-
tion to become important. It may be relevant to note that the galactic
gas layer does in fact become thicker at larger radii, and that many

R.B. Larson

spiral galaxies have gas layers that extend beyond their optical images
and become broad and warped in the outer regions (Sancisi 1976).
Whether the warps are due to the recent addition of gas or to some
other disturbing influence, the gas in these regions has evidently not
yet been able to settle into a thin, dense, coplanar layer, as is
probably required for star formation to occur. Thus the outer parts
of many spiral discs may indeed be relatively young in terms of their
star formation history and stellar content, a possibility that is con-
sistent with the fact that the colors of spiral galaxies generally
become bluer with increasing radius.

REFERENCES

Audouze, J., and Tinsley, B. M., 1976. Ann. Rev. Astron. Astrophys.
    14, 43.
Demarque, P., and McClure, R. D., 1977. The Evolution of Galaxies and
    Stellar Populations, ed. B. M. Tinsley and R. B. Larson, p. 199.
    Yale University Observatory, New Haven.
Eggen, O. J., Lynden-Bell, D., and Sandage, A. R., 1962; Astrophys. J.
    136, 748.
Gott, J. R., 1977. Ann. Rev. Astron. Astrophys. 15, in press.
Janes, K. A., 1975. Astrophys. J. Suppl. 29, 161.
Larson, R. B., 1972. Nature 235, 21.
Larson, R. B., 1974. Mon. Not. Roy. Astr. Soc. 166, 585.
Larson, R. B., 1975. Mon. Not. Roy. Astr. Soc. 173, 671.
Larson, R. B., 1976a. Mon. Not. Roy. Astr. Soc. 176, 31.
Larson, R. B., 1976b. Galaxies, Sixth Advanced Course of the Swiss
    Society of Astronomy and Astrophysics, ed. L. Martinet and M. Mayor,
    p. 67. Geneva Observatory.
Larson, R. B., 1978a. J. Computational Physics, in press.
Larson, R. B., 1978b. In preparation.
Oort, J. H., 1970. Astron. Astrophys. 7, 381.
Ostriker, J. P., 1977. The Evolution of Galaxies and Stellar
    Populations, ed. B. M. Tinsley and R. B. Larson, p. 369. Yale
    University Observatory, New Haven.
Ostriker, J. P., and Thuan, T. X., 1975. Astrophys. J. 202, 353.
Rees, M. J., 1977. The Evolution of Galaxies and Stellar Populations,
    ed. B. M. Tinsley and R. B. Larson, p. 339. Yale University
    Observatory, New Haven.
Sancisi, R., 1976. Astron. Astrophys. 53, 159.
Searle, L., 1977. The Evolution of Galaxies and Stellar Populations,
    ed. B. M. Tinsley and R. B. Larson, p. 219. Yale University
    Observatory, New Haven.
Tinsley, B. M., and Larson, R. B., 1978. Astrophys. J., submitted.
Toomre, A., 1977. The Evolution of Galaxies and Stellar Populations,
    ed. B. M. Tinsley and R. B. Larson, p. 401. Yale University
    Observatory, New Haven.