LARGE-SCALE ASPECTS OF STAR FORMATION AND GALACTIC EVOLUTION

Richard B. Larson
Yale Astronomy Department
Box 6666
New Haven, Connecticut 06511, USA

1. INTRODUCTION

Star formation in spiral and irregular galaxies often appears to occur in large gas complexes or spiral arm segments with characteristic sizes of the order of a kiloparsec and masses of the order of 10^7 solar masses, typically containing one or more giant H II regions (Elmegreen and Elmegreen 1983, 1987; Elmegreen 1987a). Thus, to understand such basic features of galaxies as the spatial distribution of star formation and the overall star formation rate (SFR), it is necessary to understand the origin and development of these large star-forming complexes. A quantity of particular interest for the evolution of galaxies is the timescale τ(gas) for turning the interstellar gas into stars. Estimates of this quantity for many galaxies (Kennicutt 1983; Donas et al. 1987) show a well-defined range of values with a median of several Gyr and with a small but apparently real increase toward later Hubble types; such an increase in τ(gas) with Hubble type is also implied by the fact that galaxies of later Hubble type have consumed less of their gas than galaxies of earlier Hubble type. Many observations indicate that the SFR per unit area in galactic disks also increases with the local surface density of gas, but the form of this dependence is not yet well established (Freedman 1986; Kennicutt 1987; Donas et al. 1987; Young 1988).

It seems clear from present knowledge that star formation occurs as part of a complex cycle of processes whereby dispersed gas is assembled into massive star-forming clouds and complexes, and after a small fraction of the gas has condensed into stars these clouds and complexes are disrupted again and the gas is recycled into a more dispersed form. A variety of forms of interstellar matter and many conversion processes are probably involved in this cycle of cloud formation and destruction: atomic, molecular, ionized, and shock-heated gas may all take part, and many thermal and dynamical processes may act to change their state. A complete understanding of star formation will ultimately require an understanding of how the whole system of interstellar gas and stars works, but this is not yet available. All of the mechanisms that have been proposed for driving star formation – density waves, gravitational instabilities, explosions, cloud collisions – may well play a role in the evolution of the system, but it is unlikely that any single mechanism can by itself provide an adequate basis for understanding star formation.

One can nevertheless hope that a few dominant processes can be identified that govern the large-scale aspects of star formation. For example, while various shock

R. E. Pudritz and M. Fich (eds.), Galactic and Extragalactic Star Formation, 459–474.

compression processes may generate filamentary or shell-like structures on smaller scales, the only evident way to account for the existence of star-forming complexes with sizes up to a kiloparsec is by the effect of large-scale self-gravity in the interstellar medium (Larson 1977, 1983; Elmegreen and Elmegreen 1983; Elmegreen 1985). Gravity probably plays an important role in the formation of smaller clouds as well (Section 3). The most important destruction mechanism for large star-forming clouds is probably ionization (Whitworth 1979; Larson 1987b). There has been considerable progress within the past decade in understanding both how gravity generates structure in galactic disks and how ionization destroys star-forming clouds; therefore the present review will focus on the effects of gravity and ionization in driving the cycle of cloud formation and destruction, and on the implications of these processes for star formation rates and the evolution of galaxies.

2. STAR FORMATION AS AN AGGREGATION PROCESS

Star formation evidently involves the condensation of interstellar gas to higher densities on a hierarchy of scales ranging from spiral arms down to individual protostars. However, star formation requires more than just compression of the gas, since compression by itself is of little help unless it produces gravitationally bound structures. Many proposed theories of star formation have been based on compression mechanisms such as density-wave shocks, cloud collisions, and various explosive phenomena (for reviews, see Larson 1977; Woodward 1978; Lada, Blitz, and Elmegreen 1978). However, all of these mechanisms involve violent and disruptive events that are more likely to destroy clouds than to create new gravitationally bound objects. For example, the shock-compressed layer produced by a collision between two clouds will generally not be self-gravitating and will quickly disperse once the collision is over (Stone 1970; Smith 1980). Also, while such explosive consequences of star formation as ionization, winds, and supernovae may compress the ambient gas into dense layers or shells, on the whole the effect will be to disperse the gas rather than to form new bound clouds; moreover, the shock-compressed layers may be disrupted by hydrodynamical instabilities before gravity can act to fragment them into stars (Giuliani 1980). Thus, the compression mechanisms that have been proposed do not seem capable by themselves of forming new giant molecular clouds, much less the even larger complexes that are apparently the basic units of star formation; instead, star formation seems to be more fundamentally an aggregation process, probably driven by large-scale gravitational effects that can collect the interstellar gas into massive star-forming clouds and complexes.

A related question concerns the role of spiral structure in star formation. If a large-scale disturbance such as a spiral density wave compresses the interstellar medium in a galaxy, such a density wave might help to trigger star formation. However, Elmegreen and Elmegreen (1986; see also Elmegreen 1987a) find no correlation between the SFR in spiral galaxies and the type of spiral structure that they contain; in particular, they find no evidence that the SFR is influenced by the presence of a density wave. Star formation can in any case evidently proceed quite vigorously even in irregular galaxies that have no spiral pattern at all (Hunter and Gallagher 1986), and there is no obvious difference in the way star formation proceeds on small scales in galaxies with and without spiral structure. Thus it appears that spiral structure, rather than being of fundamental importance for star formation, is instead only incidental or of secondary importance. In fact, spiral structure may often be just a consequence in a differentially rotating system of the same large-scale gas aggregation processes that lead to star formation (Section 3). Thus, we again infer that large-scale gas aggregation processes driven or aided by the self-gravity of the gas are the most fundamental requirement for star formation, rather than the external compression of interstellar gas to higher densities.

3. GRAVITATIONAL AGGREGATION PROCESSES IN GAS DISKS

Gravitational aggregation processes in the gas layers of galaxies can take several forms: (1) Gravitational instability, leading to the unbounded growth of small density perturbations, can cause a rotating gas disk to fragment into clumps (for a review, see Larson 1985). (2) In a basically stable but differentially rotating gas layer, "swing amplification" can produce a large but finite increase in the amplitudes of shearing, spiral-shaped density fluctuations (Goldreich and Lynden-Bell 1965; Toomre 1981). (3) A seed condensation in a shearing gas layer can grow by gravitationally accreting gas, thereby building up a much more massive cloud or complex (Julian and Toomre 1966; Larson 1983, 1987b; Toomre 1987). (4) Random cloud collisions can lead to coalescence if gravity acts as "glue" to hold the colliding clouds together and prevent them from being disrupted by the collisions. The rate of star formation in galaxies depends on how fast all of these processes can collect interstellar matter into massive star-forming clouds and complexes. The conditions under which each of these processes can be effective, and the associated timescales for gas aggregation, are summarized below.

3.1. Gravitational Instability

The gas layers of most disk galaxies are thin enough that they are governed to a good approximation by the stability criterion for an infinitely thin disk. Here we neglect the significant contribution that the stellar component may make to destabilizing the gas layer (Jog and Solomon 1984). Assuming that the interstellar medium behaves as a perfect isothermal gas with a constant velocity dispersion c, the gas layer is gravitationally unstable if

$$Q = c\kappa/\pi G\mu < 1$$

where κ is the epicyclic frequency and μ is the surface density of the gas layer. The growth time for the instability is

$$\tau \sim c/\pi G\mu \sim 5 \times 10^7 \text{ yr},$$

where the local values $c \sim 6$ km/s and $\mu \sim 8$ M(sun)/pc^2 have been substituted. For these values, the critical length scale or "Jeans length" for gravitational instability is about one kiloparsec and the associated mass is about 10^7 solar masses, similar to the observed masses of large star-forming complexes. In most spiral galaxies the value of Q is actually somewhat larger than unity (for example, $Q \sim 2$ locally), so that true gravitational instability and fragmentation would not be expected to occur, although strong growth of perturbations may still be possible even for $Q > 1$ if magnetic fields are sufficiently important on large scales (Elmegreen 1987b). However, values of Q of the order of unity or less are found in irregular galaxies and in the outermost parts of some spiral galaxies; thus gravitational instability may occur in these systems and may account for the extremely clumpy gas distributions typically observed in them.

3.2. Swing Amplification

Closely related to gravitational instability is the phenomenon of swing amplification, which is a sort of "incipient instability" that can occur in the presence of differential rotation. In a differentially rotating gas disk, if Q is not much larger than unity, shearing density fluctuations are amplified by finite but possibly large factors as they are wound up by differential rotation. Swing amplification is important if Q lies in the approximate range

$$1 < Q <\sim 2.$$

The amplification factor depends sensitively on Q, being very large for Q near 1 but relatively small for Q larger than about 2 (Toomre 1981; Larson 1984). Extensive numerical simulations of the dynamics of galactic disks containing gas have shown that at least the less regular spiral patterns in galaxies can be understood to a large extent as resulting from swing amplification effects (Miller, Prendergast, and Quirk 1970; Sellwood and Carlberg 1984; Carlberg and Freedman 1985; Sellwood 1987; Toomre 1987). The amplification factors found in the simulations typically range from about 5 to 20. The growth time τ is similar to that for gravitational instability, but the requirement that Q must be in the range ~ 1 – 2 implies in addition that τ must be closely related to the epicyclic frequency κ:

$$\tau \sim c/\pi G\mu \sim (1-2)\,\kappa^{-1}.$$

Since swing amplification produces only a finite increase in the amplitudes of density fluctuations, perhaps typically by a factor of 10, a further requirement for swing amplification to be effective for star formation is that there must already be density fluctuations of the order of 10% on kiloparsec scales that can be amplified into fluctuations of order unity. N-body simulations by Toomre (1987) of part of a shearing gas disk show the recurring appearance of trailing spiral features that can be understood as amplified noise of purely statistical origin arising from the graininess of the numerical simulation. The interstellar medium in spiral galaxies is in fact quite clumpy or "grainy", and this might provide sufficient input noise for swing amplification. Random density fluctuations of order 10% on kiloparsec scales could, for example, occur if the interstellar medium is strongly clumped on a scale of 100 pc or 10^5 solar masses. Clumping of the dust on such scales is in fact seen in photographs of nearby spiral galaxies (e.g. Malin 1987). This clumping could result in turn from various effects of star formation; the question of the nucleation of gravitational amplification processes will be considered again in Section 4.

3.3. Accretion from a Shearing Gas Layer

If a bound condensation is present or is somehow formed in a differentially rotating gas disk, it will perturb the orbital motions of the stars and gas passing near it and will induce a trailing spiral-shaped wake of enhanced density around itself (Julian and Toomre 1966; Icke 1982). Within this wake, free particle motions are strongly deflected from circular orbits and particles may make one or more loops, resulting in many orbit crossings. In a gas disk, dissipative cloud collisions will therefore occur in this wake and may cause much of the gas to be captured and accreted by the seed object, resulting in the runaway growth of a much more massive cloud or complex (Larson 1983, 1987b; Toomre 1987). In Toomre's numerical simulations, runaway clump growth sometimes occurs to dramatic effect, the clump eventually accreting nearly all of the gas in its annulus. Evidently, if such a runaway accretion process can somehow be nucleated (see Section 4) it can be very effective in building very massive star-forming clouds and complexes. It may also be an effective way of inducing spiral structure in galaxies, closely related to the swing amplifier mechanism (Julian and Toomre 1966).

In the idealized case of accretion from a uniform shearing gas layer of surface density μ with no random motions, a seed object of mass M grows at a rate given approximately by

$$dM/dt \sim (2GM)^{2/3} A^{-1/3} \mu$$

(Larson 1987b), where A is the Oort constant specifying the local shear rate. Numerical simulations by Sekiya, Miyama, and Hayashi (1987) of the accretion of gas by Jupiter from the primordial solar nebula yield accretion rates that agree with this prediction within 10% in the limit where gas pressure effects are unimportant; even when pressure effects are significant, the calculated accretion rates agree with this formula within a factor of 2. If the accretion rate is proportional to $M^{2/3}$ as predicted, the mass of the accreting object grows rapidly as t^3 and the time required for a mass M to be accumulated is

$$\tau \sim 3(2G)^{-2/3}(AM)^{1/3}\mu^{-1} \sim 10^8 \text{ yr};$$

here a mass of 10^5 M(sun) has been assumed and local values of A and μ have been used. A mass of 10^5 M(sun) has been adopted because this is about the smallest mass for which gravitationally induced motions are not dominated by random motions, and for which the interstellar medium can be even crudely approximated as a uniform thin layer. In any case, the accretion time is only weakly dependent on both A and M, and even weaker dependences are suggested by the numerical results of Sekiya et al. (1987); therefore the timescale for this process depends primarily on the gas surface density μ. With a modest enhancement in the local surface density such as might occur in spiral arms, accretion would proceed as fast as the gravitational amplification effects discussed above, and these processes could work together to build massive star-forming clouds and complexes in spiral arms.

3.4. Cloud Collisions with Gravity as "Glue"

Even without help from gravity to bring clouds together, random collisions can lead to cloud growth if the colliding clouds stick together and are not destroyed by the collisions. The requirement that they not be destroyed by the collisions is important because diffuse clouds that are not strongly gravitationally bound are very fragile and easily disrupted. Numerical simulations of cloud collisions by Hausman (1981), Gilden (1984), and Lattanzio el al. (1985) have shown that "standard clouds" colliding under typical conditions are usually shredded and dispersed by the collisions. In cases where a head-on collision creates a gravitationally bound cloud, coalescence can occur hat is assisted by the self-gravity of the coalesced object (Nagasawa and Miyama 1987). In more general types of collisions, even bound clouds may be disrupted if the gravitational forces holding them together are not strong enough to withstand the destructive impact of the collision.

If the destruction of clouds by collisions is mainly due to ablation (e.g. Woodward 1978) caused by the ram pressure ρV^2 of the impacting clouds or gas streams, then the condition required for a cloud to survive and accrete gas rather than be ablated away is that its gravitational binding pressure should exceed the ram pressure tending to disrupt it. For a gravitationally bound cloud of surface density μ_c, the pressure produced by gravity is approximately $G\mu_c^2$, independently of whether the cloud is spherical, flattened, or filamentary. Thus the condition for a cloud to survive an interaction producing a ram pressure ρV^2 is approximately

$$G\mu_c^2 > \rho V^2.$$

If a typical diffuse cloud density ρ is 20 atoms per cm^3 and a typical collision velocity V is 10 km/s, the minimum surface density μ_c implied by this condition is

$$\mu_c > 120 \text{ M(sun)/pc}^2.$$

Surface densities of this order are in fact quite typical for gravitationally bound molecular clouds with a wide range of sizes and masses (Larson1981; Myers 1983, 1985; Myers and Goodman1987; Falgarone 1988). The observed correlation between internal velocity dispersion and cloud size can be explained if all molecular clouds have approximately the same surface density and if they are all in virial equilibrium, as appears to be the case. A surface density of the order of 120 M(sun)/pc^2 might be expected to be typical because clouds with smaller surface densities are easily ablated and destroyed and therefore are presumably short-lived, while clouds with much higher surface densities probably form stars very rapidly and therefore are soon destroyed by the effects of star formation (Section 5).

Myers and Goodman (1987) have proposed that the motions in molecular clouds are magnetically dominated and that the correlation between velocity dispersion and size can be understood if all clouds have similar magnetic field strengths of the order of 30 μG. If the destructive impact of collisions is resisted primarily by magnetic pressure, and if magnetic pressure is assumed to balance ram pressure in the above argument, then the required magnetic field strength is about 30 μG, in agreement with the typical value proposed by Myers and Goodman.

If accreting clouds maintain an approximately constant surface density μ_c as they grow in mass, and if the average density of the interstellar medium is ρ and a typical cloud velocity is V, then the accretion rate for a cloud of mass M is

$$dM/dt = \rho V M/\mu_c.$$

This implies that the cloud mass grows exponentially in time with a growth timescale

$$\tau = \mu_c/\rho V \sim 3 \times 10^8 \text{ yr},$$

assuming $\mu_c \sim 100$ M(sun)/pc^2, $\rho \sim 1$ atom/cm^3, and $V \sim 10$ km/s. Thus cloud growth due to random collisions has a somewhat longer timescale than gravitational accretion in a shear flow, but not by a large factor. If the average interstellar density were 3 atoms/cm^3, as might be appropriate in spiral arms, the timescale for accretion due to random collisions would be $\sim 10^8$ years, comparable to the timescales for the gravitationally driven accumulation processes discussed above. Thus random collisions can also contribute importantly to building up massive star-forming clouds in spiral arms.

Summarizing the various accumulation processes that have been discussed in this section, it appears that a combination of large-scale gravitational instability or amplification effects and smaller-scale cloud accretion processes, both gravitationally driven and random, can all play a role in collecting interstellar matter into massive star-forming clouds and complexes, and that the timescales for all of these processes are of the order of 10^8 years. This timescale can be estimated directly from the observations if it is accepted that the basic units of star formation are complexes with sizes of the order of a kiloparsec and that a typical gas velocity is of the order of 10 km/s; then the time required to collect the gas into complexes of this size is of the order of 10^8 years. The timescale for turning the gas into stars will of course be much longer than this because star formation is a very inefficient process; the efficiency of star formation will be considered below in Section 5.

4. NUCLEATION OF GAS AGGREGATION

With the exception of a gravitational instability which can in principle grow from an infinitesimal perturbation, all of the aggregation mechanisms discussed above require a finite initial density fluctuation or seed condensation to nucleate the growth process. Both swing amplification and gravitational accretion processes are probably effective only if the interstellar medium is already clumpy on a scale of perhaps 100 pc or 10^5 M(sun). Cloud coalescence due to random collisions can occur for smaller clouds, but it still requires a gravitationally bound seed cloud of sufficiently high surface density. A possible mechanism for producing structure on a scale of 100 pc is thermal instability in a medium heated by star formation; this can lead to the accumulation of dense cool gas into an extensive network of filaments with sizes of ~ 100 – 200 pc (Chiang and Bregman 1988). A related possibility is that explosive phenomena such as expanding H II regions, stellar winds, and supernovae can produce compressed layers or shells on similar scales (Lada, Blitz, and Elmegreen 1978; Elmegreen 1985, 1987a; McKee 1986). Fragments of molecular clouds disrupted by star formation may also provide suitable seeds from which new star-forming clouds can grow. If such processes are important, star formation may help to sustain itself by generating the smaller-scale grainy structure required to nucleate new growth processes.

Another way to nucleate large-scale aggregation processes may be through the hierarchical development of progressively larger gravitationally generated structures starting from much smaller length and mass scales than are usually assumed to be required for gravitational instability. This might be possible because the velocity dispersion c of interstellar matter is generally observed to be smaller in regions of smaller size; in fact, the observed motions in the denser parts of the interstellar medium have a spatial structure that resembles at least superficially a turbulent cascade (Larson 1979, 1981). Thus, even though the stability parameter $Q = c\kappa/\pi G\mu$ may be larger than unity on kiloparsec scales, it may be smaller than unity in smaller regions where c is smaller, allowing gravitational instability to occur on smaller scales. Such an effect is seen in the simulations by Carlberg and Freedman (1985) of disks containing very dissipative gas, in which localized dissipation leads to regions with very small gas velocity dispersions and consequently to small-scale fragmentation of the gas into clumps. For example, if c were as small as 1.5 km/s in some regions, gravitational instability could occur on scales as small as 100 pc or 10^5 M(sun) for $\mu \sim 8$ M(sun)/pc^2, or on even smaller scales for larger values of μ. Such small-scale structures could then nucleate the growth of larger-scale structures. Indeed, the effect of self-gravity in a highly dissipative medium may generally be to produce a hierarchy of self-gravitating structures over a wide range of scales, as is in fact observed in nearby molecular clouds (Larson 1981).

5. CLOUD DESTRUCTION

Massive star-forming clouds and complexes must evidently be dispersed before most of their gas has condensed into stars, since the fractional mass of young stars contained in them is generally inferred to be very small; values in the range of 1 to 10% are typically obtained, assuming a standard initial mass function. Searches for molecular gas associated with young star clusters (Bash, Green, and Peters 1977; Leisawitz 1985) have found no molecular gas associated with clusters older than 10 Myr, and this implies that clouds that form star clusters are dispersed within 10 Myr after a cluster has formed. The extensive results of Leisawitz (1985) show in addition that the clouds associated with clusters younger than 5 Myr are already beginning to be dispersed, apparently by being driven away from the cluster; only small remnant cloud fragments can be detected around clusters older than 5 Myr. Thus molecular clouds are largely destroyed within only 5 Myr after a cluster of stars has formed. Cloud destruction at some point is needed to halt the runaway growth in cloud mass that would otherwise occur as a result of the accretion processes discussed in Section 3.

Some mechanisms that might contribute to cloud destruction are:

(1) *Tidal disruption.* Some large gas complexes may be only weakly bound and easily sheared apart by galactic differential rotation. Swing amplification, for example, produces only transient spiral features that are eventually sheared apart. However, the timescale for this process is of the order of 40 Myr, so it cannot account for the rapid cloud destruction in only 5 Myr that is inferred from the observations discussed above.

(2) *Bipolar outflows.* Bipolar outflows from newly formed stars, as discussed extensively at this conference, could contribute importantly to the destruction of dark clouds. From the data presented by Lada at this conference, it appears that bipolar outflows might sweep up an order of magnitude more gas than is condensed into stars, possibly removing it from dark clouds and contributing to a low efficiency of star formation in them.

(3) *Stellar winds and supernovae.* These two phenomena associated with massive stars can have similar consequences, namely the acceleration, compression, and heating of residual cloud gas by strong shock fronts. While these effects may contribute to the destruction of star-forming clouds, they generally become significant only after ionization has already begun to restructure the clouds and disperse their gas (Yorke, Bodenheimer, and Tenorio-Tagle 1982; McKee 1986). Moreover, most of the supernovae produced by massive stars do not begin to explode until after 10 Myr have passed, too late to contribute to cloud destruction in the short period of only 5 Myr indicated by the observations.

(4) *Ionization.* This is probably the most important destruction mechanism for clouds that form O stars (Larson 1987b). Much of the gas in such clouds will be ionized, and the rest will be accelerated and probably largely dispersed by ionization-driven shock fronts; as a result, by the time stellar winds and supernovae begin to act, ionization will already have done most of the damage that will be done to a cloud.

The amount of gas that can be ionized by a cluster of newly formed stars was estimated by Whitworth (1979), who found that if only 4% of the mass of a star-forming cloud condenses into stars with a standard IMF, enough ionizing photons are produced by these stars to completely ionize the rest of the cloud. The cloud ionization process may often proceed via the formation of "champagne flows" (Tenorio-Tagle 1979), and numerical simulations of this phenomenon confirm the importance of ionization as a cloud destruction mechanism (Bodenheimer, Tenorio-Tagle, and Yorke 1979; Tenorio-Tagle, Yorke, and Bodenheimer 1979). The detailed calculations show that typically about 1% of the ionizing photons emitted by the O stars create new electron-proton pairs that are evaporated from the cloud. For a conventional IMF this implies that the mass of gas ionized is about 20 times the mass that condenses into stars, in good agreement with the estimate of Whitworth (1979).

Therefore it appears likely that clouds that form O stars are destroyed primarily by the effects of ionization. Most of the ionizing photons that are ever produced are emitted within the first 5 Myr following the formation of a cluster of stars, and this is consistent with the evidence that cloud destruction is largely completed within 5 Myr. In fact, all of the clusters younger than 5 Myr studied by Leisawitz (1985) are associated with H II regions which are almost certainly playing a major role in dispersing the remnant molecular clouds. If ionization is the main cloud destruction mechanism, and if clouds are completely ionized after only 5% of their mass has condensed into stars, the predicted efficiency of star formation is

$$\text{efficiency } \varepsilon = \text{stellar mass / cloud mass} \sim 0.05.$$

The efficiency of star formation is defined here as the fraction of the gas in a star-forming complex that condenses into stars before the remaining gas is dispersed and star formation ceases or is strongly suppressed. If the interstellar medium is continually being cycled through star-forming clouds, the efficiency of star formation is the fraction of the interstellar gas that goes into stars during each passage through the cycle.

The efficiency predicted above is comparable to the efficiencies inferred from observations of regions of star formation, which are typically of the order of a few percent. The predicted efficiency would be too small if not all of the ionizing photons emitted by O stars contribute to cloud destruction, which would be the case if many O stars end their lifetimes no longer closely associated with their birth clouds. On the other hand, the predicted efficiency could also be too high because the above estimate neglects the role of ionization-driven shocks in disrupting clouds; in addition to the gas ionized, a comparable or even larger amount of cloud material can be accelerated and dispersed in the form of expanding neutral shells (Mazurek 1980; Beltrametti, Tenorio-Tagle, and Yorke 1982; McKee 1986). Such shells are actually observed around some H II regions. If more material is dispersed in neutral than in ionized form, a star formation efficiency smaller than 5% would be predicted, perhaps in better agreement with observational estimates like that of Myers et al. (1986), who find a typical efficiency of star formation of only about 2%.

The total rate at which gas is presently being ionized in our Galaxy may be estimated from the rate of production of ionizing photons deduced from observations of thermal radio emission (Güsten and Mezger 1983); if 1% of these photons create new ions, the resulting ionization rate is sufficient to completely ionize the entire interstellar medium once every 5×10^7 years. At the solar distance from the galactic center, the timescale for processing all of the interstellar gas through H II regions is about 10^8 years. These timescales are comparable to those for gas accumulation discussed in Section 3, so there is apparently an approximate balance between cloud formation by gravity and cloud destruction by ionization. The existence of such a balance would imply that star formation is normally a self-regulated process whose efficiency, and therefore whose rate, are strongly limited by the negative feedback effect associated with ionization (Larson 1987b).

If the efficiency of star formation is known, we can estimate the expected timescale τ_{SF} for converting all of the interstellar gas into stars. If the timescale for cycling the gas through star-forming clouds is $\sim 10^8$ years, and if 5% of the gas condenses into stars during each cycle, then the timescale for converting all of the gas into stars is about 2 Gyr; if the efficiency is 2%, the timescale τ_{SF} is about 5 Gyr. Allowing for gas recycling, the timescale τ(gas) for gas depletion is about 1.5 times longer than τ_{SF} for a standard IMF. The resulting predicted values of τ(gas) are comparable to the values that have been inferred from observational estimates of the SFR in galaxies; several different studies have yielded median values for τ(gas) that are of the order of 3 – 4 Gyr (Larson, Tinsley, and Caldwell 1980; Kennicutt 1983; Donas et al. 1987), again assuming a standard IMF

6. STAR FORMATION RATES

If the efficiency of star formation remains constant, the star formation rate is proportional to the rate at which interstellar gas is assembled into star-forming clouds. The timescales for the gas accumulation processes discussed in Section 3 all depend on the surface density μ of the gas in a galactic disk, so μ is probably the most important parameter governing the SFR in galaxies. For gravitational instability or swing amplification effects, the growth time τ is approximately $c/\pi G\mu$; if the large-scale gas velocity dispersion c is the same everywhere in galactic disks, as appears to be true observationally to a first approximation, then τ depends only on μ and is proportional to μ^{-1}. For gravitational accretion in a shearing gas layer, the cloud growth time is approximately $G^{-2/3}(AM)^{1/3}\mu^{-1}$, which apart from weak dependences on A and M is again proportional to μ^{-1}. For cloud growth due to random collisions, the growth time is proportional to $c^{-1}\rho^{-1}$, where ρ is the average gas density; thus if c is constant and if variations in the scale height of the gas are not important, so that ρ is proportional to μ, the cloud growth time is once again proportional to μ^{-1}. If we therefore assume that the timescale for star formation is at least approximately proportional to μ^{-1}, i.e. that

$$\tau_{SF} \propto \mu^{-1},$$

then the star formation rate per unit area in a galactic gas layer of surface density μ follows

$$\mathrm{SFR/area} = \mu/\tau_{SF} \propto \mu^2.$$

(Larson 1987b). This is essentially the relation first proposed by Schmidt (1959), expressed in terms of the surface density of gas in a galactic disk. It is important to note that the timescales and star formation rates being considered here apply only to quantities averaged over scales of the order of a kiloparsec or more, since the large-scale gas aggregation processes discussed in Section 3 mostly operate on scales of this order.

The above predicted relation is roughly consistent with the results of many of the observational studies that have been carried out to test the validity of the Schmidt law. The results of observational comparisons between the SFR and μ are found to depend on the spatial resolution with which these quantities are measured, since a better correlation and a stronger dependence of the SFR on μ are found when the data are averaged over larger regions (Freedman 1986). For example, in M31 and M33 the best correlation between the SFR per unit area and the gas surface density is obtained when the data are binned into resolution elements at least 500 pc across; the observations then follow closely a Schmidt law with an exponent near 2 (Freedman 1986; Nakai and Sofue 1982). In our own galaxy, the data on the SFR per unit area and on the gas surface density μ as functions of galactocentric distance assembled by Lacey and Fall (1985) also follow a similar relationship (Fall, private communication), although in M51 and NGC 6946 the corresponding azimuthally averaged quantities appear to show only a linear dependence of the SFR per unit area on μ (Young 1988). On the scale of entire galaxies, the data of

Donas et al. (1987) for many galaxies suggest a strong dependence of the SFR on the average gas surface density that follows approximately a Schmidt law with an exponent near 2, although with considerable scatter. A particularly well studied galaxy is M31 (Walterbos 1988); its total SFR is nearly an order of magnitude smaller than that of our Galaxy although its average gas surface density is only about a factor of 2 lower, and this again suggests a strong dependence of the SFR on the average surface density of gas in a galactic disk.

The above Schmidt-type relation between the SFR and μ does not, however, provide any basis for understanding why the timescale for gas depletion in galaxies increases significantly toward later Hubble types even though the average gas surface density does not vary much with Hubble type. For example, the median gas lifetimes implied by the results of Donas et al. (1987) increase systematically from about 2 Gyr for Sbc galaxies to about 7 Gyr for Irr galaxies. A similar increase of evolutionary timescale with Hubble type is also required to account for the variation in the colors of galaxies along the Hubble sequence. Despite this, the average gas surface density is only slightly smaller in irregular than in spiral galaxies (Hunter and Gallagher 1986).

Numerical simulations of the dynamics of galactic disks containing gas and exhibiting strong swing-amplified spiral structure (Sellwood and Carlberg 1984; Carlberg and Freedman 1985) show that the value of the stability parameter Q tends to be regulated at a value of about 2: dissipation tends to reduce the velocity dispersion c and hence the value of Q, but the resulting stronger swing amplification activity tends to heat the disk again and maintain Q at a steady-state value of about 2. If such self-regulation effects are generally important in the gas layers of spiral galaxies, they might tend to maintain a nearly constant value of Q in these systems. If Q rather than c is assumed to be constant, the growth time $c/\pi G\mu = Q/\kappa$ for gravitational disturbances in disks becomes proportional to κ^{-1} instead of μ^{-1}. To a good approximation, the epicyclic frequency κ is proportional to the angular velocity of rotation Ω, so that we can then write

$$\tau_{SF} \propto \kappa^{-1} \propto \Omega^{-1}.$$

The star formation rate per unit area then follows

$$\mathrm{SFR/area} = \mu/\tau_{SF} \propto \Omega\mu,$$

a form also suggested at this conference by Silk. In practice, it may often be difficult to distinguish this type of law from a Schmidt law with an exponent near 2, since regions with high gas surface densities also tend to be regions of high Ω. However, when applied to galaxies of different Hubble types, the second relation implying a dependence of the SFR on Ω would predict that the more slowly rotating later-type galaxies should convert their gas into stars more slowly than galaxies of earlier Hubble type, as is indeed suggested by their higher gas fractions. Thus the rate of evolution of disk galaxies could be determined largely by their rotation rates (Larson 1983; see also Section 7).

If Q were exactly the same in galaxies of all Hubble types, and if the gas surface density μ were also the same, the gas velocity dispersion c would vary as κ^{-1} and so would be considerably larger in the later-type galaxies, in contradiction to the observations. However, for two reasons the value of Q required for significant self-gravitational effects to occur is smaller in later-type galaxies, and this reduces the implied increase of c along the Hubble sequence. Later-type galaxies are more gas-rich, and therefore their gas is confined more by its own self-gravity and less by the gravity of the stars; if the gas component becomes dominant, the thickness of the gas layer can no longer be neglected and the value of Q required for gravitational amplification effects to be significant is reduced by about a factor of 1.5 (Larson 1984, 1985). In addition, differential rotation becomes less important in later-type galaxies, and is weak or absent in irregular galaxies (Hunter and Gallagher 1986); therefore swing amplification also becomes less effective, and it may not operate at all in irregular galaxies. If swing amplification does not occur and a true gravitational instability is required to initiate star formation, then Q must be less than 1 for a thin layer and less than about 2/3 for a self-gravitating gas disk. Thus the value of Q required for a galactic gas layer to be in a dynamically steady or marginally stable state may decrease by as much as a factor of 3 along the Hubble sequence from ~ 2 for typical spirals to ~ 2/3 for the most gas-rich irregulars. Little or no variation of c along the Hubble sequence would then be implied by the assumption that galaxies of all types are in a dynamically steady state.

7. TOWARD AN UNDERSTANDING OF THE HUBBLE SEQUENCE

The above discussion leaves unanswered the question of whether the gas velocity dispersion c, which plays a key role in controlling the rates of star formation processes, is itself determined mainly by large-scale gravitational amplification effects or by smaller-scale acceleration processes such as ionization powered by stellar energy sources (Larson 1987b). If small-scale processes such as ionization determine the value of c, then c might be expected always to be of the order of 5 – 10 km/s, whereas if large-scale gravitational effects regulate c, it might be more valid to assume that Q is always of order unity. Probably both types of effects operate at a significant level, and an intermediate situation prevails in which neither c nor Q is strictly constant. This would imply that for much larger gas surface densities, as might exist particularly during the early evolution of galaxies, c would be somewhat larger and Q would be somewhat smaller than is now typical, but neither quantity would differ as much as if the other were assumed to remain constant. The gas layer would then be more gravitationally active because of the smaller value of Q, and the timescale $\tau \sim c/\pi G\mu$ for gravitational aggregation and hence for star formation would be shorter. Thus galaxies or parts of galaxies that begin with higher gas surface densities should convert their gas into stars faster than galaxies or regions of lower initial surface density. To the extent that the angular velocity Ω is an independent parameter determining the SFR, galaxies with higher Ω should also evolve faster than galaxies with lower Ω.

We can therefore understand why galaxies of earlier Hubble type, which generally have both higher surface densities and higher angular velocities than galaxies of later type, have evidently evolved faster than galaxies of later type and presently contain a smaller proportion of gas and young stars. If it is assumed that Q is presently of order unity in all spiral and irregular galaxies, then it can be shown from the definition of Q that the present fractional gas content $\mu(\text{gas})/\mu(\text{total})$ is approximately equal to the ratio $c/V(\text{max})$ of the gas velocity dispersion to the maximum galactic rotation speed $V(\text{max})$ (see also Quirk 1972); since this ratio varies from about 1/50 for Sa galaxies to 1/5 or more for Irr galaxies, the present gas content is predicted to vary from ~ 2% to > 20% along this sequence, in reasonable agreement with the observations (Hunter and Gallagher 1986). A larger range of variation and even better agreement with the observations are obtained if Q decreases somewhat along the Hubble sequence, as discussed in Section 6. Thus the gas contents of galaxies of different Hubble types can be understood if their relative rates of evolution and present gas contents depend on the ratio $c/V(\text{max})$ and if this ratio is a fundamental underlying parameter that varies systematically along the Hubble sequence. Disks with smaller $c/V(\text{max})$ should generally evolve faster than disks with larger $c/V(\text{max})$ because they are relatively colder and thinner, and subject to stronger gravitational amplification effects (Larson 1984). Dynamically, the Hubble sequence does appear to be primarily a sequence of decreasing maximum rotation speed, since this is the dynamical parameter that correlates best with Hubble type (Rubin et al. 1985). Earlier-type galaxies also tend to have stronger differential rotation, and this may further accelerate their evolution (Larson 1983) because the effectiveness of swing amplification depends on the amount of shear present.

In addition to rapid star formation, another consequence of vigorous gravitational activity is that gravitational torques associated with trailing spiral density enhancements will transfer angular momentum outward, causing the mass distribution to become more centrally condensed (Julian and Toomre 1966; Lynden-Bell and Kalnajs 1972; Larson 1984; Carlberg 1987; Lin and Pringle 1987). For a typical present-day spiral galaxy, the timescale for redistribution of angular momentum by this effect may be estimated to be of the same order as the Hubble time, and similar timescales have also been found in numerical simulations of galactic disks (e.g. Carlberg and Freedman 1985). However, earlier in the evolution of galaxies when they were presumably more gas-rich, their level of gravitational activity would have been higher and the timescale for transport of angular momentum would have been shorter; thus significant radial redistribution of matter in galactic disks may have resulted. Such effects would have been most important in the galaxies of earliest Hubble type, since they would have experienced the most vigorous gravitational activity; therefore these galaxies should now be the most centrally condensed ones, as is indeed observed. Thus the variation in the radial structure of disk galaxies along the Hubble sequence may be a result of the same large-scale gravitational processes that are involved in star formation. Lin and Pringle (1987) have also suggested that the characteristic exponential structure of galactic disks can be produced by the action of gravitational torques if the timescale for transfer of angular momentum is comparable to the timescale for star formation, as would indeed be plausible if both angular momentum transfer and star formation result from the same gravitational amplification processes in galactic disks.

Another property of the Hubble sequence that may be explainable by the processes discussed above is that, while earlier-type spiral galaxies form stars in a well-regulated fashion and show only a limited range of τ_{SF} within each type, later-type galaxies show much more variability in their level of star formation; irregular galaxies, in particular, can range from very active to very inactive, there being no obvious correlation between the SFR and other galactic properties (Hunter and Gallagher 1986). These characteristics may result if star formation in early-type galaxies is mainly driven by swing amplification and is regulated by negative feedback effects that tend to increase c and Q and thus reduce the amplification factor, whereas in the latest-type galaxies swing amplification does not occur and therefore such "fine control" of the SFR is not possible. In the latter case, star formation may require a large-scale gravitational instability and the system may fluctuate between stable and unstable states with only small differences in Q, leading to large fluctuations in the SFR. An explanation of the highly variable properties of irregular galaxies based on "stochastic self-propagating star formation" was previously suggested by Gerola, Seiden, and Schulman (1980), but they proposed no physical origin for the random fluctuations of the SFR that dominate in these systems; here a specific mechanism is proposed, namely global gravitational instabilities producing large changes in the SFR for only slight changes in the stability parameter Q.

While some features of the structure and evolution of galaxies may thus be at least qualitatively understandable on the basis of the processes that have been discussed, a complete theoretical understanding of the evolution of galaxies (Tinsley 1980) will not be possible until a much more detailed understanding of these and many other processes is available. Indeed, it has so far not even been possible to resolve convincingly the discrepancy between observational inferences that the SFR in galactic disks has varied little with time and theoretical expectations that the SFR should decline strongly with time as the gas supply is depleted (Kennicutt 1987). Both large-scale gas flows (Tinsley and Larson 1978; Lacey and Fall 1985) and variations in the stellar IMF (Scalo 1986; Larson 1987a) may well play a significant role in galactic evolution, but neither of these effects is presently at all well understood. Star formation processes are also undoubtedly far more complex than the simple effects that have been discussed here; for example, our understanding of the internal evolution of large star-forming clouds is still only embryonic, and it may be that magnetic fields play a major role in determining how they evolve (Myers and Goodman 1987). Fortunately, since the study of the evolution of galaxies is becoming increasingly an observational subject, we can hope that direct lookback observations will soon begin to provide significant constraints on our understanding of the star-forming history of galaxies.

REFERENCES

Bash, F. N., Green, E., and Peters, W. L., 1977, *Astrophys. J.*, **217**, 464.

Beltrametti, M., Tenorio-Tagle, G., and Yorke, H. W., 1982, *Astron. Astrophys.*, **112**, 1.

Bodenheimer, P., Tenorio-Tagle, G., and Yorke, H. W., 1979, *Astrophys. J.*, **233**, 85.

Carlberg, R. G., 1987, in *Nearly Normal Galaxies: From the Planck Time to the Present*, ed. S. M. Faber, p. 129. Springer-Verlag, New York.

Carlberg, R. G., and Freedman, W. L., 1985, *Astrophys. J.*, **298**, 486.

Chiang, W.-H, and Bregman, J. N., 1987, *Astrophys. J.*, in press.

Donas, J., Deharveng, J. M., Laget, M., Milliard, B., and Huguenin, D., 1987, *Astron. Astrophys.*, **180**, 12.

Elmegreen, B. G., 1985, in *Birth and Infancy of Stars*, eds. R. Lucas, A. Omont, and R. Stora, p. 215. North-Holland, Amsterdam.

Elmegreen, B. G., 1987a, in *Star Forming Regions*, IAU Symposium No. 115, eds. M. Peimbert and J. Jugaku, p. 457. D. Reidel, Dordrecht.

Elmegreen, B. G., 1987b, *Astrophys. J.*, **312**, 626.

Elmegreen, B. G., and Elmegreen, D. M., 1983, *Mon. Not. Roy. Astron. Soc.*, **203**, 31.

Elmegreen, B. G., and Elmegreen, D. M., 1986, *Astrophys. J.*, **311**, 554.

Elmegreen, B. G., and Elmegreen, D. M., 1987, *Astrophys. J.*, **320**, in press.

Falgarone, E., 1988, this volume.

Freedman, W. L., 1986, in *Luminous Stars and Associations in Galaxies*, IAU Symposium No. 116, eds. C. W. H. de Loore, A. J. Willis, and P. Laskarides, p. 61. D. Reidel, Dordrecht.

Gerola, H., Seiden, P. E., and Schulman, L. S., 1980, *Astrophys. J.*, **242**, 517.

Gilden, D. L., 1984, *Astrophys. J.*, **279**, 335.

Giuliani, J. L., 1980, *Astrophys. J.*, **242**, 219.

Goldreich, P., and Lynden-Bell, D., 1965, *Mon. Not. Roy. Astron. Soc.*, **130**, 125.

Güsten, R., and Mezger, P. G., 1983, *Vistas in Astron.*, **26**, 159.

Hausman, M. A., 1981, *Astrophys. J.*, **245**, 72.

Hunter, D. A., and Gallagher, J. S., 1986, *Publ. Astron. Soc. Pacific*, **98**, 5.

Icke, V., 1982, *Astrophys. J.*, **254**, 517.

Jog, C. J., and Solomon, P. M., 1984, *Astrophys. J.*, **276**, 127.
Julian, W. H., and Toomre, A., 1966, *Astrophys. J.*, **146**, 810.
Kennicutt, R. C., 1983, *Astrophys. J.*, **272**, 54.
Kennicutt, R. C., 1987, in *Stellar Populations*, eds. C. A. Norman, A. Renzini, and M. Tosi, p. 125. Cambridge University Press, Cambridge.
Lacey, C. G., and Fall, S. M., 1985, *Astrophys. J.*, **290**, 154.
Lada, C. J., Blitz, L., and Elmegreen, B. G., 1978, in *Protostars and Planets*, ed. T. Gehrels, p. 341. University of Arizona Press, Tucson.
Larson, R. B., 1977, in *The Evolution of Galaxies and Stellar Populations*, eds. B. M. Tinsley and R. B. Larson, p. 97. Yale University Observatory, New Haven.
Larson, R. B., 1979, *Mon. Not. Roy. Astron. Soc.*, **186**, 479.
Larson, R. B., 1981, *Mon. Not. Roy. Astron. Soc.*, **194**, 809.
Larson, R. B., 1983, *Highlights of Astronomy*, **6**, 191.
Larson, R. B., 1984, *Mon. Not. Roy. Aston. Soc.*, **206**, 197.
Larson, R. B., 1985, *Mon. Not. Roy. Astron. Soc.*, **214**, 379.
Larson, R. B., 1986, *Mon. Not. Roy. Astron. Soc.*, **218**, 409.
Larson, R. B., 1987a, in *Stellar Populations*, eds. C. A. Norman, A. Renzini, and M. Tosi, p. 101. Cambridge University Press, Cambridge.
Larson, R. B., 1987b, in *Starbursts and Galaxy Evolution*, 22nd Rencontre de Moriond, eds. T. Montmerle and J. T. T. Van, in press. Editions Frontieres, Gif sur Yvette.
Larson, R. B., Tinsley, B. M., and Caldwell, C. N., 1980, *Astrophys. J.*, **237**, 692.
Lattanzio, J. C., Monaghan, J. J., Pongracic, H., and Schwarz, M. P., 1985, *Mon. Not. Roy. Astron. Soc.*, **215**, 125.
Leisawitz, D. T., 1985, Ph.D. thesis, University of Texas, Austin; Millimeter Wave Observatory Technical Report 85-2.
Lin, D. N. C., and Pringle, J. E., 1987, *Astrophys. J.*, in press.
Lynden-Bell, D., and Kalnajs, A. J., 1972, *Mon. Not. Roy. Astron. Soc.*, **157**, 1.
Malin, D. F., 1987, photographs published in *Mercury*, **16**, 48.
Mazurek, T. J., 1980, *Astron. Astrophys.*, **90**, 65.
McKee, C. F., 1986, *Astrophys. Space Sci.*, **118**, 383.
Miller, R. H., Prendergast, K. H., and Quirk, W. J., 1970, *Astrophys. J.*, **161**, 903.
Myers, P. C., 1983, *Astrophys. J.*, **270**, 105.
Myers, P. C., 1985, in *Protostars and Planets II*, eds. D. C. Black and M. S. Matthews, p. 81. University of Arizona Press, Tucson.
Myers, P. C., Dame, T. M., Thaddeus, P., Cohen, R. S., Silverberg, R. F., Dwek, E., and Hauser, M. G., 1986, *Astrophys. J.*, **301**, 398.
Myers, P. C., and Goodman, A. A., 1987, *Astrophys. J.*, in press.
Nagasawa, M., and Miyama, S. M., 1987, *Prog. Theor. Phys.*, in press.
Nakai, N., and Sofue, Y., 1982, *Publ. Astron. Soc. Japan*, **34**, 199.
Quirk, W. J., 1972, *Astrophys. J. (Letters)*, **176**, L9.
Rubin, V. C., Burstein, D., Ford, W. K., and Thonnard, N., 1985, *Astrophys. J.*, **289**, 81.
Scalo, J. M., 1986, *Fundamentals of Cosmic Physics*, **11**, 1.
Schmidt, M., 1959, *Astrophys. J.*, **129**, 243.
Sekiya, M., Miyama, S. M., and Hayashi, C., 1987, *Earth, Moon, and Planets*, in press.
Sellwood, J. A., 1987, talk presented at the Aspen workshop on The Evolution of Galaxies, June 1987.
Sellwood, J. A., and Carlberg, R. G., 1984, *Astrophys. J.*, **282**, 61.
Smith, J. A., 1980, *Astrophys. J.*, **238**, 842.

Stone, M. E., 1970, *Astrophys. J.*, **159**, 293.
Tenorio-Tagle, G., 1979, *Astron. Astrophys.*, **71**, 59.
Tenorio-Tagle, G., Yorke, H. W., and Bodenheimer, P., 1979, *Astron. Atrophys.*, **80**, 110.
Tinsley, B. M., 1980, *Fundamentals of Cosmic Physics*, **5**, 287.
Tinsley, B. M., and Larson, R. B., 1978, *Astrophys. J.*, **221**, 554.
Toomre, A., 1981, in *The Structure and Evolution of Normal Galaxies*, eds. S. M. Fall and D. Lynden-Bell, p. 111. Cambridge University Press, Cambridge.
Toomre, A., 1987, talk presented at the Aspen workshop on The Evolution of Galaxies, June 1987.
Walterbos, R. A. M., 1988, this volume.
Whitworth, A., 1979. *Mon. Not. Roy. Astron. Soc.*, **186**, 59.
Woodward, P. R., 1978. *Ann. Rev. Astron. Astrophys.*, **16**, 555.
Yorke, H. W., Bodenheimer, P., and Tenorio-Tagle, G., 1982, *Astron. Astrophys.*, **108**, 25.
Young, J. S., 1988, this volume.