STAR FORMATION AND GALACTIC EVOLUTION: FROM PROTOGALAXIES TO STARBURSTS

Joseph Silk
Astronomy Department
University of California
Berkeley, California 94720 U.S.A.

ABSTRACT. Three topics in star formation theory are reviewed: the initial mass function, the star formation efficiency, and the star formation rate. A physical mechanism for bimodal star formation is developed. Applications are made to the solar neighborhood, to the inner galaxy, to starburst galaxies, and to past star formation in protodisks and in protoellipticals. Implications are drawn for galactic morphology, for chemical evolution, and for the present density of stellar remnants.

1. INTRODUCTION

Our knowledge of star formation in nearby molecular clouds is largely phenomenological. Hence, it may seem presumptuous to attempt to extrapolate this knowledge to early phases of galactic evolution for which there are no direct observations. Such obstacles have rarely deterred theoreticians, however, and in this review I will show how far indirect evidence, both observational and theoretical, has been able to aid the theory of galaxy formation.

Three of the most essential ingredients needed to understand star formation are the initial mass function, the star formation rate, and the star formation efficiency. In this review, I shall argue that it is likely that all three of these are similar both in starbursts today and in the initial starbursts that occurred in protogalaxies when elliptical galaxies and population II stars formed, as well as when the old disk (population I) stars formed. Moreover, theory has developed to the point that one now has a glimmer of understanding about the star formation rate, and we understand at least the extrema of the initial mass function, although insight into star formation efficiency is still elusive.

The following topics will be considered: the initial mass function (IMF), the efficiency of star formation (SFE), and the rate of star formation (SFR), as

M. Peimbert and J. Jugaku (eds.), Star Forming Regions, 663–689.

appropriate to the early stages of galactic evolution. I will then describe how early star formation is likely to control the morphology and chemical history of mature galaxies. Similar processes may be at work today in starburst regions.

2. INITIAL MASS FUNCTION

2.1. The Local IMF

Consider first the locally measured IMF, which is shown in figure 1 (from Scalo 1985). This represents the total number of stars as a function of mass ever born per unit pc^2 perpendicular to the plane of the galaxy in the solar vicinity. In converting the luminosity function, an observed quantity, to mass function, some assumption must be made about the past birth rate of massive stars for stellar masses above 1 $M_\odot$. For stars of mass below 1 $M_\odot$, this is no problem: they are all still on the main sequence. Two features stand out in the IMF. There is a peak near 0.2 $M_\odot$, corresponding to the mass of the most frequently occurring stars. There is also a secondary minimum at about 1.2 $M_\odot$, unless a finely tuned adjustment is made to the past birth rate relative to the present birth rate. This seems too *ad hoc* to be acceptable: one concludes that massive (> 1 $M_\odot$) and low mass (< 1 $M_\odot$) stars have different past birth rates, that is to say, star formation is bimodal. Indeed there is considerable evidence that the sites of low and massive star formation differ in both space and time, although when averaged over, say, giant molecular cloud scales and lifetimes, a more or less universal IMF results.

There is some indirect evidence that the IMF differs between spiral arms and the interarm regions of our galaxy. To avoid excessive gas consumption, Gusten and Mezger (1983) argue that the presence of OB stars observed in both arm and interarm regions requires them to adopt an IMF deficient in low mass stars within the spiral arms (section 3.2 below).

2.2. The IMF In Starburst Regions

Starbursts are regions of greatly enhanced star formation rate, often but not always in the central regions of a galaxy. There is accumulating evidence that starbursts are exclusively forming massive stars. This argument was forcefully presented for the starburst nuclei M82 and NGC 253 by Rieke *et al.* (1980), who showed that the observed 2.2 μm luminosity, due to red giants and supergiants in the adopted galaxy evolution models, could only be explained together with the other observations if the low mass IMF cut-off were at least 3 $M_\odot$. Other constraints include the total luminosity from the starburst region, the total mass in this region, and the Bα infrared line intensity, which measures the flux of ionizing photons. As reviewed by Scalo (1985), possible loopholes remain in this result. However several recent analyses of starburst galaxies strongly support the conclusion that very few low mass stars are formed. These analyses include studies of Arp 220 and NGC 6240 (Rieke *et al.* 1985), IC 2153 (Olaffsson*et al.* 1984),

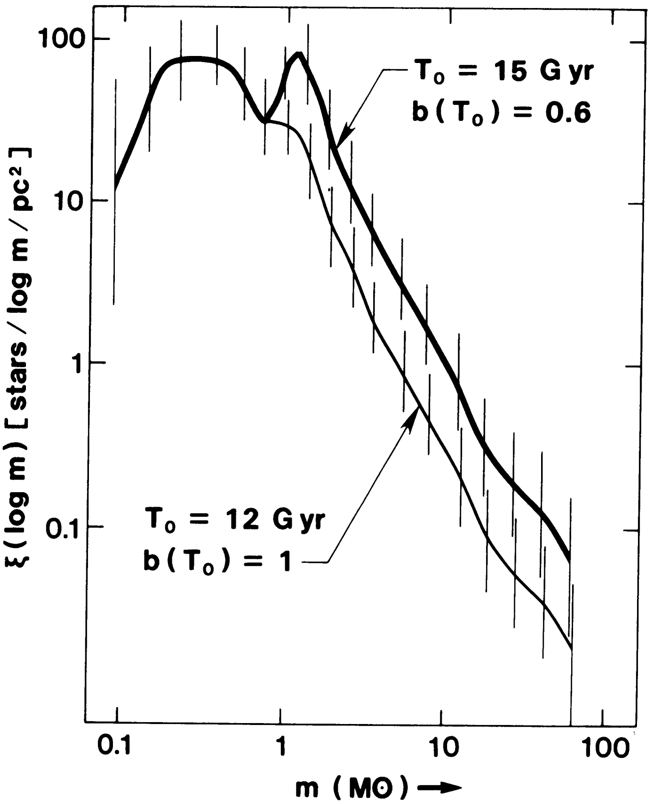

Figure 1

The initial mass function in the solar neighborhood (adapted from Scalo 1985). Uncertainties allow for scatter between different luminosity function determinations, errors in the mass–luminosity relation, and uncertainties in the bolometric corrections, effective temperature scales, and scale heights of the sampled stars. T_o is the age of the disk in the solar neighborhood, and b (T_o) is the ratio of present to past average birthrate.

and Mk 171 (Augarde and Lequeux 1985). A less direct approach utilizes the low mass–to–light ratios in starburst regions. Extreme values as low as 0.01 reported for "superstarbursts", often found in closely interacting galaxies (Joseph and Wright 1985), can only be explained if the IMF is truncated at low masses, according to models constructed by Telesco *et al.* (1985) and star formation has been initiated recently.

2.3. The Primordial IMF

It is often presumed that the first stellar generation had zero metals. This is misleading. Star formation is not an instantaneous process, and many massive stars will be born and die during the period while lower mass stars are forming. Star formation occurs in a burst that may have a time–scale of $\sim 10^8$ yr or longer. This means that the only surviving, genuinely pristine, zero metal stars are those that were born within the first 3×10^6 yr or so of the burst, before any supernova or wind–driven enrichment could have occurred. If the rate of star formation were constant during the $\sim 10^9$ yr over which the halo, with mean metallicity $<[Z]> \approx -1.5$, formed, then one might expect a fraction of order one percent of the halo to be in zero metallicity stars. One would need a sample of hundreds of halo stars in order to expect to find a single zero metallicity star, always assuming that the primordial IMF was similar to the present IMF.

The observational situation is that the number of metal–poor halo stars as a function of Z behaves very roughly as (Beers *et al.* 1985) $dN/dZ \approx$ *constant*, below $[Z] \approx -1.5$. The most metal–poor halo star known is CD -38^o 245 with $[Z] \approx -4.5$ (Bessel and Norris 1984). The formation of such a star would be indistinguishable from that of a star with zero metals. The observed frequency of very low metallicity stars (several are known with $[Z] \lesssim -3$) is not inconsistent with a primordial IMF similar to that in the solar neighborhood. As noted originally by Hartwick (1976), in a closed box model of chemical evolution applied to the halo, the fraction of stars below a given metallicity increases proportionately to Z until a value $<[Z]> \approx -1.5$. If recycling via star formation terminates at this metallicity, due to infall to the disk, then one obtains a reasonable fit to the observed halo metallicity distribution.

The observations of metal–poor halo stars do not of course require the primordial IMF to be similar to that today. Two arguments have been given which imply that the population III IMF primarily consisted of massive stars. One is that certain abundance anomalies seen in extreme Population II stars can be explained by a precursor population enhanced in massive stars. A well–known difference between extreme population II and less metal–poor population II stars is that the former are enhanced in oxygen relative to iron. This has been cited as evidence for supermassive stars ($> 100\ M_\odot$); however, an alternative (Tinsley 1979; Matteuci and Greggio 1985) interpretation is that iron is mostly produced by stars of intermediate and low mass(type I supernovae) on a much longer time-scale than oxygen (produced by massive stars). Hence the oxygen overabundance seen in the oldest stars merely reflects the fact that chemical evolution averaged over a short time-scale ($< 10^9$yr) cannot incorporate the contribution from the less massive stars, that contribute over $\sim 10^9$–10^{10} yr. A similar effect could conceivably explain other abundance ratio differences between the oldest and more recently formed stars.

The second argument that favors massive stars in population III is far more speculative: the dark halo matter could be the black hole remnants of a generation of supermassive objects (Truran and Cameron 1971). The only argument against such a speculation is based on its implausibility: the initial mass function would have to be very strongly suppressed below $\sim 10^3\ M_\odot$, so that the intergalactic gas and later generations of stars can be almost metal-free. A possible unique signature of such a hypothesis would be the generation of spectral distortions near the peak of the cosmic microwave background radiation spectrum (Bond *et al.* 1985; Negroponte 1986); however the existence of such distortions has not been confirmed.

2.4. Theory of the IMF Today

Most simply put, there is no acceptable theory of the IMF. It generally is believed that accretion and coagulation processes play important roles, but precisely how these processes are mediated by magnetic fields, accretion disk formation, and protostellar outflows remains very uncertain. One suspects that energy feedback by massive stars may be responsible for the predominance of low mass stars, and for the slope of the IMF, but it is exceedingly difficult to quantify this in any reasonably unique fashion.

A very oversimplified argument goes as follows: if protostars of mass m and luminosity L(m) form in sufficient number to heat their surroundings and raise the critical fragmentation mass in the ambient cloud, then with typical grain opacity averaged over frequency for a Planck spectrum $\kappa \propto T$, one expects (Silk 1977a)

$$\int \frac{dN}{dm} L dm \sim T^5$$

while the fragmentation criterion for opaque fragments of surface density μ is

$$m \gtrsim m_{\rm crit} \propto T^2/\mu \overset{\propto}{\sim} T^3.$$

Thus neglecting any geometrical dilution or inefficiency in tapping the protostellar energy, one finds, if L $\overset{\propto}{\sim}$ m^3, that

$$dN/dm \propto m^{-7/3}. \qquad (1)$$

One may have greater confidence in theoretical attempts to understand the extrema of the IMF. The upper limit is undoubtedly due to the interaction of radiation from a massive protostellar core with infalling accreting matter (Larson and Starrfield 1971; Kahn 1974; Yorke and Krugel 1977; Wolfire and Cassinelli 1985). Radiation pressure acting on the dust yields a mass limit that in Kahn's model may be taken to be

$$m_u = 60(Z_\odot/Z)^{\frac{1}{2}} M_\odot. \qquad (2)$$

To understand the lower limit on protostellar masses, one has to delve into theories of how the initial cores form by fragmentation. The first question to be asked is: what is the mass of the smallest unstable gas clump formed in a molecular cloud? The second question is: what is the minimum stellar mass likely to form in this clump?

2.4.1. Minimum Clump Mass

Observations suggest that massive stars preferentially form in massive, warm molecular clouds (Larson 1982). Typical clump sizes in giant molecular clouds are observed to be $\sim 100 M_\odot$ (Blitz and Stark 1986), while in the smaller, cold molecular clouds, clump masses are $\sim 1 M_\odot - 10 M_\odot$ (Myers 1985). The minimum clump mass is extremely sensitive to the gas temperature; for example, a general discussion (Larson 1985a) of fragmentation of pressure-supported sheets or filaments yields a scale for the most rapidly growing unstable mode of

$$M = 2.4 T^2/\mu(M_\odot pc^{-2}) \qquad M_\odot, \qquad (3)$$

where T is the temperature and μ is the surface density of a sheet or filament of molecular gas. The Jeans criterion applied to a collapsing, uniform, spherical cloud also yields a minimum unstable fragment mass that is similar to (3), but three–dimensional simulations suggest that generic fragmentation is more likely to occur after anisotropic collapse to transient sheet–like or filamentary configurations that are supported by some combination of thermal pressure, rotation and magnetic stresses (Tohline 1980). To apply (3) to molecular clouds, one may note that over a wide range of densities and length scales, molecular clouds appear to satisfy (Larson 1981; Myers 1983).

$$\mu \approx 150 M_\odot pc^{-2}. \qquad (4)$$

With T ranging from $\sim$ 3K in cold clouds to $\sim$100 K in warm clouds, one infers fragment masses in the range 0.3 $M_\odot$ to $\sim$ 300 $M_\odot$. Even if the correlation (4) is disregarded, (3) still yields low mass fragments in cold clouds and massive fragments in warm clouds, consistent with observations.

2.4.2. Minimum Stellar Mass

The efficiency at which a fragment forms a star or stars is unknown. For massive stars, feedback is important in limiting accretion; for low mass stars, the mass of the fragment is doubtless a critical factor. Multiple subfragmentation is likely to occur, both as inferred from three-dimensional simulations, and from studies of dark globules such as B5 (Goldsmith *et al.* 1985). The minimum mass of a protostellar fragment cannot be inferred from the numerical simulations which lack sufficient dynamic range, and recourse must be made to analytic arguments.

Foremost among these was the idea put forward originally by Hoyle (1953) of opacity–limited fragmentation. The instability criterion, if given by the Jeans length, yields a minimum mass proportional to $T^{3/2}\rho^{-1/2}$. As long as radiative cooling is effective, collapse remains nearly isothermal and the Jeans mass decreases. The minimum Jeans mass, attained as the optical depth across a fragment becomes large (Hayashi and Nakano 1965) is about 0.01 $M_\odot$ if the collapse is spherically symmetric. Low and Lynden–Bell (1976) and Silk (1977b) argued that this scale, derived for dust opacities, could be identified with the minimum protostellar mass. Application to clouds of primordial composition was pioneered by Hirasawa (1967), Matsuda *et al.* (1969) and Yoneyama (1972).

Molecular hydrogen forms via H^- that is produced by H atoms reacting with residual electrons left over after the recombination epoch. Rotational excitations of trace amounts of H_2 maintain the temperature at about 1000K in a collapsing cloud until a density of about 10^9 cm^{-3} is attained. Palla *et al.* (1983) realized that at this density, the three–body process

$$3H \rightarrow H_2 + H \tag{5}$$

became important, and despite the absence of dust grains, primordial clouds become completely molecular at very high density. The conversion to H_2 means that H_2 cooling persists until the H_2 is collisionally dissociated. The H_2 transitions become optically thick as the temperature gradually rises, the H_2 being dissociated above ~3000K, but only when the minimum Jeans mass is well below 0.1 $M_\odot$, much as in the non–primordial cloud.

However the minimum Jeans mass encountered in opacity–limited fragmentation is probably too naive a concept to have much relevance to star formation. Several processes are likely to augment the minimum fragment mass. These include anisotropic collapse, rotation, magnetic fields, inhibition of fragmentation as gradients develop, and fragment coagulation and accretion (Silk 1980). It is more useful to ask what is the mass of the minimum quasi–statically contracting protostellar core that forms by fragmentation and these other processes. A necessary condition for such a core to form is that the rate of increase of binding energy in free–fall collapse be sufficient to dissociate either the grains that dominate the opacity in a conventional cloud, or the H_2 in a primordial cloud. Coincidentally, both are destroyed (albeit by very different processes) at about the same temperature, namely $\sim$ 2000 K-3000 K. Consequently, the minimum core mass is nearly the same, namely $\sim$ 0.1-0.2$M_{\odot}$. The condition that yields this mass is simply (Gaustad 1963)

$$\frac{1}{2}\frac{GM^2}{Rt_{ff}} > 4\pi\frac{R\sigma T^4}{\kappa\rho} \tag{6}$$

where κ is the Rosseland mean opacity, R the radius, ρ the mean density, T the central temperature, and t_{ff} the free–fall time of a collapsing fragment of mass M. Condition (6) reduces to

$$M > 0.1(T/1000K)^{7/6}M_{\odot} \tag{7}$$

for typical ice–grain opacities. Accretion can obviously increase the final mass, but the coincidence between the prediction (7) and the observed peak in the stellar luminosity function (Figure 1) is tantalizingly suggestive that there is at least a grain of truth in the underlying physics of fragmentation and core formation.

In the case of a primordial cloud, one arrives at a similar minimum core mass, provided that H_2 is the dominant constituent of the fragment. Several arguments have been advanced which suggest that formation of much more massive cores may be favored. If the H_2 is destroyed in shocks, it will not reform when the density exceeds $\sim 10^4$ cm^{-3} and the post–shock temperature is $\gtrsim 10^4$ K. Moreover, even if the H_2 survives, the primordial clouds are thermally unstable at a density and temperature when the instantaneous Jeans mass is about 100 $M_{\odot}$. In the absence of H_2, two-photon and Lyman alpha cooling (Nakada and Yoneyama 1976; Hasegawa et al 1981) guarantee that a protostellar core forms of mass about $\sim$ 20 $M_{\odot}$, according to (6), if $T \approx 10^4$K.

Finally, accretion onto the protostellar core, which occurs at a rate of order V_s^3 G^{-1} for a spherically symmetric, isothermal collapse, where V_s is the sound velocity, would certainly have been enhanced in the absence of heavy elements by a factor $\sim$ 100 relative to its rate in conventional clouds (Stahler *et al.* 1985). Again this favors massive star formation, but does not preclude the formation of low mass stars.

In summary, the theoretical uncertainty is such that the minimum stellar mass in primordial clouds may be similar to that today ($\sim$ 0.1 $M_{\odot}$), or possibly could be $\sim$ 10 -100 $M_{\odot}$ (Kashlinsky and Rees 1983). In any event, stars in the mass range 0.1 to 100 $M_{\odot}$ could almost certainly have formed in primordial clouds, although the IMF may have been deficient in the lower mass stars. Once heavy elements are produced, however, at a level $\gtrsim 10^{-4}$ that of solar abundance, a minimum protostellar mass of $\sim 0.1 M_{\odot}$ and a maximum mass of $\sim$ 100 $M_{\odot}$ seem indicated by these simple theoretical arguments.

3. THE STAR FORMATION EFFICIENCY

3.1. The Efficiency In The Solar Vicinity

The efficiency of star formation in a given region requires specification of the initial mass function as well as the duration of the star forming phase, or equivalently, the rate of star formation. It will be useful to define star–forming efficiency of a cloud (SFE) as the mass–fraction in stars after one dynamical time–scale has elapsed. It is important to distinguish the local SFE defined for a cloud, from the global SFE for an entire star–forming region or a galaxy, where many clouds form and die, and many generations of stars form, within a global dynamical time. The IMF is only reasonably well known between 0.2 $M_\odot$ and $1 M_\odot$ in the solar vicinity: it is the low mass stars that are crucial to knowing the efficiency. Adoption of a Miller-Scalo (1979) IMF allows one to estimate the star formation efficiency in nearby molecular clouds. The local gas mass (HI and H_2) is 0.06 $M_\odot$ per $M_\odot$ of total mass in a cylinder perpendicular to the galactic plane at the solar radius, and the local star formation rate is 3.6×10^{-11} $M_\odot$ yr^{-1} $M_\odot^{-1}$ for a Miller-Scalo IMF. Now giant molecular cloud complexes are probably where most stars, and certainly the massive stars, are forming. Hence in these regions, the inferred efficiency is about 0.7 percent per 10^7 yr; all gas would be recycled through stars in 1.5×10^9 yr. About 40 percent of this mass is returned to the interstellar medium; the rest is locked up in white dwarfs, neutron stars, and low mass stars. The observed ratio of mass in stars to total gas mass in well studied examples such as the λ Sco complex (Duerr *et al.* 1982) is of order 1 percent, from which we conclude that for typical giant molecular clouds (GMC), the massive star forming lifetimes are of order 10^7 yr. This estimate presumes that low mass stars are also forming: if they are not, and gas is recycled without disrupting the cloud, longer lifetimes are possible.

Cold clouds appear to have higher star formation efficiencies. For example, in the core of the ρ Oph cloud, about 40 percent of the gas mass is in stars (Wilking and Lada 1983), and in Taurus-Auriga, about 5 percent of the molecular gas may have been converted into stars (Cohen and Kuhi 1979). This simply confirms what one expects from the Miller–Scalo IMF: that low mass stars dominate the mass density and accordingly determine the local efficiency. There does seem to be a significant variation in stellar content between dark clouds, but one may more plausibly attribute this to variation in the duration of the star forming phase rather than to variation in the IMF.

However, what is not so apparent is why, in a region where low mass stars are seen to be forming, the star forming efficiency should considerably exceed that in a large molecular cloud complex where stars of a wide range of masses are forming. The duration of the star forming phase could be very long in the regions of low mass star formation, while massive stars may form over a shorter timescale. Evidence for this was originally presented by Herbig (1962); more recent discussions cite the $\sim 10^8$ yr lag between main–sequence turn–on time of low mass stars and turn–off of the most massive stars in the Hyades Cluster (Stauffer 1984). Star formation in OB associations appears to be non–coeval: the less massive stars$(8-20 M_\odot)$ form first and are $(10-20)\times 10^6$ yr older than the most massive stars $(> 25\ M_\odot)$ (Doom *et al.* 1985). A similar effect has been claimed for young clusters (Iben and Talbot 1965; Adams *et al.* 1984), although when account is taken of the duration of time spent on the main sequence, only the lowest mass stars appear to show any early peak in their formation rate(Stahler 1985).

One notable difference between open clusters and associations is that the latter are expanding. This can be understood in terms of differing star formation efficiency. For a cluster to remain bound, it must have formed with high ($\gtrsim$ 30 percent) efficiency according to Mathieu (1983) and Lada *et al.* (1985), while formation in associations could have been much less efficient, with gas disruption by OB stars leading to the observed expansion. Such a variation in efficiency is most simply attributed to the duration of the low mass star formation phase: an extended period of low mass star formation enhances the efficiency of star formation (Elmegreen 1983), as seen in the ρ Oph core. Open clusters must have formed from clouds that were forming low mass stars over 10^7 -10^8 yr, whereas OB associations formed from clouds that underwent massive star formation over a period of $\sim 10^6$ -10^7 yr, but with little prior low mass star formation. The reason why some clouds can undergo an extended period of low mass star formation while others do not can be plausibly attributed to differences in cloud mass, as will be argued below.

3.2. Bimodal Star Formation at the Present Epoch

The preceding discussion shows that star formation occurs efficiently in some regions, where most of the low mass stars form, and much less efficiently in other regions, where the massive star formation rate is high and massive star formation may indeed predominate. This difference between low mass and massive star formation appears to be reflected in the local IMF (Section 2.1). The bimodal star formation model of Gusten and Mezger (1983) takes this one step further. These authors argue that the global star formation rate in our galaxy, in particular the radial gradient, can only be understood if star formation is very inefficient in the inner spiral arms, where the star formation rate is highest, thereby not exhausting the local gas reservoir. Relative to the solar vicinity, the massive star surface density increases by a factor of 15 at 4 kpc, while the total surface density inferred from the rotation curve rises by only a factor of 3. Hence if stars (and stellar remnants) are responsible for the disk mass, a radial variation in the IMF is required, and moreover, even a Miller-Scalo IMF would overproduce low mass stars in the inner disk.

Since the massive star formation rate is inferred via infrared and thermal radio continuum observations, the only remaining adjustable parameter involves low mass star formation. The Gusten-Mezger prescription for bimodal star formation is as follows: low mass stars ($\lesssim 3\ M_\odot$) form only in the interarm regions, while more massive stars form both in and outside the arms but at a rate that is larger in the arms by the enhanced formation of giant molecular clouds (GMC) due to the increased local gas streaming velocity. Their model yields approximately equal numbers of OB stars inside and outside the arms at the solar radius, whereas spiral arm OB stars dominate by a factor of 3 over interarm OB stars in the inner galaxy at 4-6 kpc. Thus the bimodal hypothesis means that the integrated stellar lock–up rate is reduced by a factor of 2 at the solar radius, and by a factor of 4 in the inner galaxy. The net gas consumption time–scale is accordingly increased, at the solar radius, to about 6×10^9 yr. The global gas consumption time is somewhat less: the star formation rate is presently 5 $M_\odot$ yr^{-1} in the bimodal model, with about 2 $M_\odot$ yr^{-1} being permanently locked up in stellar remnants and low mass stars. Since the total mass of interstellar HI and H_2 is about $4 \times 10^9\ M_\odot$, this means that if our galaxy is a closed system with no infall, the bulk of star formation will decrease with an e–folding time of about 2×10^9 yr.

This may be an uncomfortably short fraction of the age of the disk, which is probably at least 10^{10} yr, especially given the fact that spiral galaxies of similar morphological type to our own galaxy are quite common. Infall offers one way of augmenting the gas supply in the disk, but this seems a much less attractive possibility now than it did several years ago. There is little observational evidence for the necessary reservoir of infalling intergalactic gas clouds; much of the high velocity HI gas appears to have a more recent origin, either in the Magellanic stream or in a galactic fountain model.

Sandage (1985) resolves the time-scale discrepancy by adopting a more extreme bimodal star formation model. If no low mass stars at all below 2-3 $M_\odot$ presently form in the inner galaxy (and there is no evidence that they do!), the lock-up rate can be reduced by about an order of magnitude, relative to that in a model with a universal IMF. Larson (1985b) argues that the minimum stellar mass forming in the inner galaxy cannot be too low ($\lesssim 2\ M_\odot$), otherwise excessive mass in dark remnants would be produced. Hence our galaxy can continue its present star formation activity for another 2×10^{10} yr. Only then will the bright lights begin to go out throughout the universe, as spiral galaxies exhaust their gas supplies. One could certainly get by with a more modest version of bimodal star formation: an e–folding time of about 5 x 10^9yr would suffice, and this would allow some low mass star formation, at a level of 10 to 20 percent, to accompany the massive star formation. That there is likely to be at least a shred of truth in this conclusion comes from recognizing that in starburst regions, low mass star formation does indeed appear to be suppressed by roughly an order of magnitude in terms of the gas consumption rate (Section 2.2). The principal deficiency in the more extreme hypothesis is that it offers no explanation of the formation of low mass stars in the past: these stars contribute practically all the stellar mass in galaxies. A successful model of galactic evolution must incorporate both massive and low mass star formation.

3.3. Bimodal Star Formation In The Past

Important steps towards developing a self-consistent bimodal star formation model of galactic evolution have been taken in papers by Gusten and Mezger (1983) and by Larson (1985b). The former paper tackles the radial abundance gradient, which is explained by the enhancement of massive star formation in spiral arms. The model nucleosynthetic yield systematically increases within the solar circle, as the massive star fraction rises towards 2 kpc. The principal alternative to bimodal star formation is to postulate radial inflow: Lacey and Fall (1985) obtain reasonable abundance gradients if the inflow velocity is nearly 1 kms^{-1}.

Larson (1985b) addresses the origin of the dark matter in the disk of our galaxy, and develops a bimodal model in which a greatly enhanced massive star formation rate, taken to decay exponentially with an e-folding time of 3×10^9 yr, dominates the early galaxy. The massive star remnants, mostly white dwarfs, constitute the dark matter that amounts to about fifty percent of the local disk mass density (Bahcall 1984). There is a possible difficulty with this prediction, since too few very low luminosity white dwarfs are seen to allow any increase of the white dwarf birth rate at early times (Liebert *et al.* 1983) unless the cooling of these old white dwarfs is more efficient than expected in current models (Iben and Tutukov 1984).

Larson also applies the bimodal model to the chemical evolution of the solar neighborhood disk population. A rapid increase in metallicity is produced over the first 5×10^9 yr. Over the past 10^{10} yr or so, the enrichment is very modest, amounting to about 10 percent. This is found by Larson to be in reasonable agreement with the empirical age-metallicity relation (Twarog 1980; Carlberg *et al.* 1985), especially if the minimum mass of the massive star mode at early times is taken to be about 4 $M_\odot$. In fact, this is precisely the value required by Larson's model at present in the inner galaxy in order to avoid overproducing the mass in dark remnants.

Perhaps the most unsatisfactory aspect of Larson's model is that it overproduces heavy elements in the early galaxy. Larson argues that one can obtain a satisfactory yield only if stars between about 10 $M_\odot$ and 16 $M_\odot$ explode as supernovae and contribute to the observed oxygen enrichment. The difficulty with this is that theoretical studies suggest that these relatively low mass supernova progenitors do not give large enough yields (Hillebrandt 1985). While the absolute yields are very uncertain, the relative yields should be more reliable. In particular, Wilson *et al.* (1985) found that a 25 $M_\odot$ progenitor gave good agreement with solar system abundances, whereas a 15 $M_\odot$ star did not.

3.4. Towards A Physical Mechanism for Bimodal Star Formation

Fragmentation into preferentially more massive clumps of molecular cloud gas (Section 2.4.1.) provides one possible means of forming more massive stars. But this mechanism requires a considerable number of OB stars to be present in order to provide the heat input that leads to enhanced pressure support and larger fragment masses. Why would many massive stars form in a localized region to initiate this process?

The resolution may come from recognizing three effects. Firstly, if the mass of a star is randomly distributed according to some IMF, then the most abundant, least massive stars should form initially and most frequently. Massive stars are delayed in their appearance, and Elmegreen (1983) shows for example that a 10 $M_\odot$ star may take 40 times longer to form than a 1 $M_\odot$ star by stochastic fragmentation of the parent cloud. This means that the most massive stars are so rare that they will only be found in the massive star clusters, which in turn can only form from the most massive molecular clouds.

Secondly, the most massive molecular clouds are associated with, and some might say, even define, the spiral arms (Solomon 1987; Stark 1987). One can understand this if spiral density waves play an important role in enhancing molecular cloud growth (see below). If the mass distribution of molecular clouds is such that the average interarm cloud is considerably less massive ($\sim 10^4 - 10^5$ $M_\odot$) than the average arm cloud ($\sim 10^5 - 10^6 M_\odot$), one immediately obtains an enhancement of OB star formation in the spiral arms relative to interarm regions. However, were this all one could say, then low mass star formation would occur both in the arms and outside the arms: star formation could not be bimodal in the sense of suppressing low mass star formation and reducing the efficiency in the arms.

Thirdly, once several OB stars form in a sufficiently massive molecular cloud, dynamical and thermal feedback is likely to play an important role in subsequent star formation. Two-dimensional hydrodynamical simulations (Klein et al 1983) have demonstrated that gas clumps can be triggered into collapse when engulfed by an ionization front driven by nearby massive stars. The fragmentation argument of Section 2.4.1 suggests that in such a situation, gas heating, due for example to molecule–grain collisions (Falgarone and Puget 1985), will maintain large clump masses and favor predominantly massive star formation. Hence massive stars are likely to induce further massive star formation, until the local gas reservoir is depleted. The smallest clumps will tend to be ablated, providing further focussing of the induced IMF towards massive stars (Klein et al 1986). Thus induced OB star formation occurs primarily in massive clouds whose growth is driven by density waves: this is the proposed physical mechanism for bimodal star formation.

3.5. The Efficiency of Star Formation in Starburst Regions and in Protogalaxies

Comparison of available gas masses and star formation rates in well-studied starburst regions such as the core of M82 suggests that the star formation efficiency, even when no allowance is made for low mass star formation, is enhanced by about an order of magnitude relative to the inner disk of our galaxy. The gas exhaustion time-scale by lock-up into remnants is about 10^9 yr and implies that such energetic starbursts are a relatively rare phenomenon.

The star formation efficiency is also known for protogalaxies. For example, a luminous elliptical galaxy is inferred, from population synthesis studies that reproduce its spectral energy distribution, to have converted about half of its initial gas content into long-lived stars after about 10^9 yr (Tinsley 1977; Bruzual 1983). One cannot rule out a shorter, more intense initial star formation burst for forming ellipticals, lasting only over a free–fall time of $\sim 10^8$ yr, but this would have required a star–forming efficiency that seems inordinately high. Moreover the build–up of a metallicity gradient requires at least several dynamical times: violent relaxation in free–fall collapse models results neither in metallicity gradients nor in sufficiently dense cores. Disk galaxies evolve more slowly, and are found to have nearly constant low mass star formation rates, corresponding to a lock–up time–scale of $\sim 10^{10}$ yr (Gallagher *et al.* 1984; Sandage 1985). It appears that protoellipticals and starbursts, together on the one hand, and protodisks and the inner disk of our galaxy, on the other hand, share similar global star formation efficiencies, as measured by the stellar lock–up time–scale.

This at first sight seems rather surprising. After all, physical conditions in a protogalaxy differed considerably from the present interstellar medium, and protoellipticals, unlike starbursts, made many low mass stars. However starbursts are less than 10^8 yr old, and low mass star formation is likely to be occurring but at a slower rate than that of the massive stars. Giant molecular clouds may provide a common link, where the specific lock–up rate is known to be about 10^9 yr ^{-1}a and one can try to ascertain how efficiently GMC would have formed stars in a protogalaxy. Unfortunately the theoretical uncertainties are so great that one cannot even predict the sign of any possible differences in star forming efficiency. Consider the following arguments, for illustrative purposes, that respectively support a lower and a higher efficiency.

Star Formation Was Less Efficient In Primordial Clouds:

(a). There were no, or certainly very weak, magnetic fields in a protogalactic gas cloud. Protogalactic clouds are known to acquire their specific angular momentum via tidal torquing against nearest neighbors of comparable mass. The lack of any magnetic fields means that it would have been very difficult to overcome the angular momentum barrier via Alfven wave breaking, the mechanism believed to be operative in the present interstellar medium (Nakano 1984). One suspects that the fragmentation of such clouds would be very inefficient. The three–dimensional hydrodynamical simulations of collapsing, rotating clouds provide a crude measure of efficiency: a mass fraction $\Delta M/M$ of the original cloud could form fragments with specific angular momentum reduced by a similar factor relative to the original value (Bodenheimer 1978; Bodenheimer *et al.* 1980). Since one has to shed at least four orders of magnitude in specific angular momentum to form even rapidly rotating stars, the inferred efficiency would appear to be perhaps one percent of that today, when about one percent of a giant molecular cloud is observed to form stars per dynamical time.

(b). OB star disruption of ambient gas would be effective in a primordial cloud of mass and density comparable to that of a protoglobular cluster. HII region formation is the primary disruption mechanism (Tenorio-Tagle *et al.* 1985), and the lack of heavy element coolants means that gas is readily driven out of the shallow potential well of the parent cloud. The numerical simulations find that efficient mass loss occurs once a few O stars have formed coevally: the residual gas is dispersed although the cluster remains bound.

Star Formation Was More Efficient In Primordial Clouds:

(a). The much higher collision velocities between clouds in the protogalactic potential well prior to disk formation, together with the diminished cooling efficiency, suggests that cloud collisions would have been far more disruptive than in the present interstellar medium, where density wave-induced streaming plays a primary role in the build-up of all but the most massive GMCs (Tomisaka 1984). There would have been much less coagulation; hence primordial clouds should have been appreciably less massive than conventional giant molecular clouds. If our argument about enhanced OB star formation in massive clouds applies, one concludes that fewer massive stars per unit cloud mass would have formed in primordial clouds, leading to less disruption by HII regions and winds, and a greater efficiency of star formation. Low mass stars should certainly have formed.

(b). Compton cooling plays a role in ionized gas clouds at a redshift $z \gtrsim 20$, when the cooling time–scale is less than an initial collapse time for a cloud that is condensing out of the cosmic background matter. This process can cool the gas to a temperature equal to that of the background radiation, namely $3(1+z)$K, and thereby may enhance fragmentation of a layer of gas swept up by a shock front. Ostriker and Cowie (1981) and Ikeuchi (1981) found that explosion–driven shocks in the Compton cooling era would drive shells that became gravitationally unstable on scales of stellar or massive stellar object masses, thereby providing an amplification mechanism when these objects formed and eventually exploded, and also an efficient residue of massive stellar remnants.

Evidently, theory can provide plausible arguments both for and against higher efficiency of star formation in primordial clouds. I conclude that the only reasonable option is to assume that these theoretical arguments are all inconclusive, and that as inferred from simple population synthesis models, the primordial efficiency was similar to that today. If this turns out to be an unacceptable hypothesis, one can always reopen the Pandora's box of parameters that may (or may not) determine the primordial efficiency.

This still leaves us with an unresolved issue, namely how to determine the global efficiency of star formation. If we have a theory that applies today, we can also apply it to a protogalaxy and help confirm our speculation about the efficiency being unchanged between protoellipticals and starbursts, and between protodisks and spiral arms. We have argued that low mass star formation is suppressed relative to massive star formation in regions of high star formation rate. Thus to understand the efficiency, we now turn to the final clue, the star formation rate: why does this vary between the inner and outer galaxy, for starbursts, and for protodisks and protoellipticals?

4. THE STAR FORMATION RATE

4.1. The Observed Star Formation Rate

While the low mass star formation rate is only known indirectly in our galaxy away from the solar neighborhood through constraints on gas depletion rates and on the total mass density, the rate of formation of massive stars is well-known. More or less direct observations, notably of far infrared emission, measure the massive star formation rate in starbursts, and similar arguments involving the gas reservoir constrain the low mass star formation rate. The preceding discussion inferred that in starburst regions, and in our inner galaxy, the massive star formation rate is most probably enhanced at present relative to that of the low mass stars.

The past star formation rate can be inferred for different morphological types and components of galaxies by population synthesis modelling of the observed colors. As previously mentioned, one finds that typical old populations of elliptical galaxies and spheroidal bulges underwent greatly enhanced low mass and massive star formation for the first 10^9 yr or so of their protogalactic phase, when most of the observed stars were formed. After 10^9 yr, the initial gas fraction has been depleted by about 50 percent. By contrast, a protodisk develops much more slowly. The bulk of its star formation is spread out over $\sim 10^{10}$ yr. This information has been extracted, with relatively coarse time resolution, for disk galaxies (Gallagher *et al.* 1984), and with finer time resolution, via the age-metallicity relation (Twarog 1980) for stars in the solar vicinity. These results are shown schematically in Figure 2, which is adapted from a recent discussion by Sandage (1985).

4.2. Protogalaxies and Starbursts

This variation in star formation rate and integrated lockup rate (or global efficiency) between bulge and disk components is the key to understanding the morphological characteristics of galaxies. The slow formation mode allows considerable dissipation and gas recycling, leading to formation of a disk; the rapid formation mode results in a spheroidal stellar distribution formed with much less dissipation–regulated collapse. One can arrive at any desired bulge–to–disk ratio by adjusting the gas supply rate that determines whether star formation is rapid or slow.

For example, it seems likely that the gas reservoir is environmentally regulated. In a region of high galaxy density, tidal interactions will strip outlying gas clouds, whereas an isolated protogalaxy will have its initial gas reservoir remain intact. In this way, one is able to explain why early-type galaxies predominate in galaxy clusters, and late-type galaxies in the field (Dressler 1983). One has to appeal to some process such as mergers of proto–spirals to account for the enhanced abundance of ellipticals in the regions of highest galaxy density (Silk and Norman 1983). The prevalence of enriched intracluster gas produced by stellar ejecta from early phases of star formation, gas that otherwise would have settled to form stellar disks, lends some support to this scheme.

It is here that there is a second close parallel between starbursts and protoellipticals. For extreme starburst activity often appears to be triggered by close interaction with a companion galaxy (Joseph and Wright 1985, and references cited therein). Tidal interactions appear to provide the physical mechanism responsible for the violent star formation, which has previously been shown to occur at a similar rate in both environments. There is also evidence that a central bar can trigger a starburst, the triaxiality and associated resonant orbits enhancing the radial gas streaming rate and the ensuing build–up of giant molecular clouds (Combes and Gerin 1985). Density waves stimulate molecular cloud growth, which in turn drives star formation. Collision–induced density disturbances are strongly amplified by disk self–gravity according to the swing–amplifier effect (Toomre 1981), and the non–linear inelastic response of gas clouds should then lead to a dramatic outburst of star formation. A greatly scaled down version of this effect presumably operates in spiral density waves. Note that one can only adopt a similar mechanism for proto–ellipticals if these systems form by mergers of flattened systems, presumably protodisks. The disks are necessary in order for resonant build–up and collisional triggering of molecular clouds to occur.

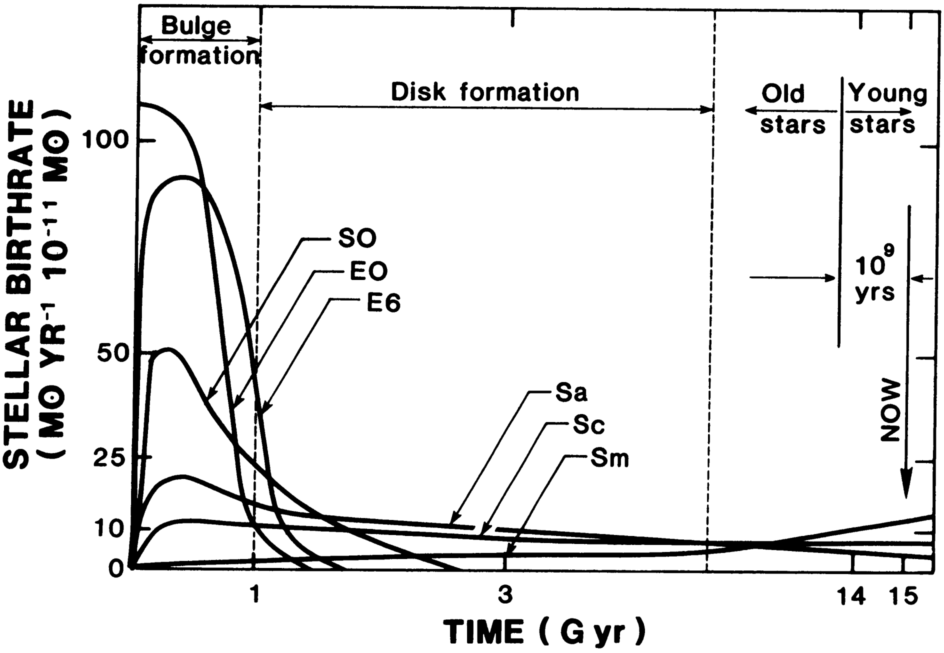

Figure 2

Past star formation rates for various morphological types of galaxy. Schematic star formation rates are shown in units of $M_\odot$ yr^{-1} $M_\odot^{-1}$ (adapted from Sandage 1985). The integral under each curve represents the total number of stars formed per unit galaxy mass.

4.3. A Simple Theory for the Star Formation Rate

There are two crucial ingredients that must be incorporated in order to obtain any insight into the rate of star formation. One is the density–wave induced rate of formation of giant molecular clouds (Tomisaka 1984). These clouds are the preferential sites of massive star formation. A second is feedback of energy from massive and possibly low mass stars into the ambient gas clouds. Molecular clouds do not collapse in a free–fall time. Nor did protoellipticals form in a free–fall time: typically, it took $\sim$ 10 free–fall times or $\sim 10^9$ yr to form the bulge population. During this period of vigorous star formation, there must have been sufficient feedback of energy to inhibit formation of a cold disk.

The modes of feedback include protostellar and post– main sequence winds, HII regions, and supernova explosions. In our own interstellar medium, these all play a role. Within molecular clouds, winds and radiation heating by OB stars appear to be most important, while the overall velocity dispersion of clouds is controlled by HII regions and supernova explosions. It is this energy input that determines the thickness of the molecular cloud disk and the scale height of the newly formed stars.

A simple feedback model for the massive star formation rate and global efficiency may be constructed as follows (Silk 1985). Define the feedback to be the energy input to the interstellar medium per unit rate of star formation. I use Type II supernova energy input as illustrative of energy input from massive stars. Normalizing to our own interstellar medium, the energy input per gram of interstellar gas is

$$f\lambda = \frac{fE_{SN}}{\dot{M}_*\tau_{SN}} = \frac{10^{51} f \text{ erg}}{(5M_\odot \text{yr}^{-1})(50 \text{ yr})} = 2 \times 10^{15} f \text{ erg g}^{-1} \tag{8}$$

where τ_{SN} is the mean time between supernovae, E_{SN} is the kinetic energy injected per supernovae, and f is the fraction of the initial energy that is imparted to the interstellar gas. According to Spitzer (1978), the radiative losses in the pressure–driven snowplow phase occur when the supernova remnant has decelerated below $v_{\rm cr} \approx 85$ kms^{-1}, and the efficiency at which energy is transferred to clouds moving at $v << v_{\rm cr}$ is approximately given by $f = v/4v_{\rm cr}$. For $v >> v_{\rm cr}$, the supernova remnants are energy conserving, and $f = 1$. This crude scaling only applies to a supernova remnant expanding into a uniform medium, and may be quite different if the interstellar gas is inhomogeneous or has multiple phases. The energy input suffices to maintain a velocity dispersion v in the molecular gas clouds provided that

$$\dot{M}_* = \frac{2vv_{cr}M_g}{\lambda t_{\rm diss}}, \tag{9}$$

where M_g is the total mass of molecular gas in the protogalaxy and $t_{\rm diss}$ is the mean dissipation time–scale of bulk kinetic energy. Applied to our own galaxy, equation (9) yields, with $M_g = 4 \times 10^9 M_\odot$, $v = 6\mathrm{kms}^{-1}$, and $t_{\rm diss} = 10^7$ yr, characteristic of the mean HI cloud collision time, $\dot{M}_* = 3 M_\odot$ yr^{-1}.

In order to apply (9) to other environments, it is necessary to rescale the dissipation time. On dimensional grounds, one expects $t_{\rm diss} \propto n^{-\frac{1}{2}}$, where n is the mean gas density. A more elaborate derivation goes as follows. The mean collision time between clouds of mean cross-section σ, number density $n_{\rm cl}$, and velocity dispersion v is a measure of the dissipation time, so that

$$t_{\rm diss} = (n_{\rm cl}\sigma v\sqrt{2})^{-1} \tag{10}$$

Assume that the clouds are in approximate gravitational equilibrium, with an internal pressure (not necessarily thermal, but due to stirring by winds, magnetic support, etc.)P_i that provides support against collapse. Writing $\mathrm{P}_i = \mathrm{P}_c(\Delta v)^2$, where P_c is the mean internal density and Δv the internal velocity dispersion, one obtains the following relations betweeen $\mathrm{P}_c, \Delta v$ and cloud radius R_c:

$$\frac{R_c}{(\Delta v)^2} P_i^{\frac{1}{2}} = \frac{\rho_c R_c}{P_i^{\frac{1}{2}}} = \left(\frac{3}{4\pi G}\right)^{\frac{1}{2}} \tag{11}$$

The fact that $(\Delta v)^2 \propto R_c$ over a wide range in cloud size, and that independently $\rho_c R_c \approx$ *constant* over this range, equation(4), lends considerable support to the inference that cloud pressure is approximately constant in nearby molecular clouds. Theoretical arguments are readily devised that can explain the coincidence between inferred gravitational support and constant pressure in terms of cloud stability subject to external pressure (Chieze 1985) or wind-driven shock stirring (Silk 1985b).

However regardless of any such model, one can rewrite the dissipation time-scale in terms of the mean internal cloud pressure P_i and the mean pressure due to cloud random motions P, defined as being equal to $n_{\rm cl} M_{\rm cl}\, v^2$ for clouds of mean mass $M_{\rm cl}$. This leads to a new form for (9), namely

$$\dot{M}_* = (6\pi G)^{\frac{1}{2}} M_g (v_{\rm cr}/\lambda) P / P_i^{\frac{1}{2}}. \tag{12}$$

Next, we note that if similar hydrodynamical stirring associated with star formation is responsible both for the mean cloud dispersion in velocity and for internal support against gravity, the internal pressure should be in approximate

Table 1
STAR FORMATION RATE AND EFFICIENCY

Environment	$\upsilon/\upsilon_\odot$	$\rho/\rho_\odot$	$\upsilon/\upsilon_\odot$	$p/p_\odot$	M_g/M_*	$\dot{M}_*/M_*$	$\dot{M}_*/M_g$	ϵ
Solar						$M_\odot/M_\odot$Gyr		
Neighborhood	1	1	1	1	0.05	0.04	0.8	0.01
Inner Galaxy	3	10	1	10	0.04	0.12	3	0.01
Protodisk	3	10	1	10	1	3	3	0.01
Starburst	10	100	1	100	1	10	10	0.01
Protoelliptical	1	1	10	100	1	10	10	0.1

equipartition with the kinetic energy in bulk motions, or P $\approx P_i$. We shall adopt this simplifying assumption in order to arrive at a scaling relation for the mean SFR, while realizing that it is unlikely to be strictly valid even in our own interstellar medium, where the pressure in molecular cloud cores considerably exceeds that in the diffuse gas. Finally, we arrive at

$$\dot{M}_*/M_g = (6\pi GP)^{\frac{1}{2}}\upsilon_{cr}\lambda^{-1}, \tag{13}$$

namely that the massive star formation rate per unit gas mass is proportional to the square root of the mean pressure in the interstellar medium. Expression (13), or more generally, equation (9), provides not only the massive star formation rate; it also yields the global efficiency. By global efficiency, I mean the total mass cycled through massive stars relative to the initial gas mass within a dynamical time–scale. I presume that low mass stars also form as long as the gas reservoir is present, but the feedback argument does not constrain their formation rate. The ratio of stellar mass to gas mass per unit dissipation time, or equivalently dynamical time for the gas, is

$$\epsilon = 2\upsilon\upsilon_{cr}\lambda^{-1}, \tag{14}$$

with

$$\dot{M}_*/M_g = \epsilon/t_{\rm diss} \tag{15}$$

and

$$t_{\rm diss}^{-1} = (3\pi G\bar{\rho}/2)^{\frac{1}{2}}, \tag{16}$$

where $\bar{\rho}$ is an appropriately defined mean density.

The analysis of bimodal star formation by Gusten and Mezger (1983) leads to an equation very similar to (15), except that $t_{\rm diss}^{-1}$ is replaced by the giant molecular cloud coagulation rate ν due to streaming of gas clouds entering spiral arms. This is equal to $[\Omega(R) - \Omega_p]/\Delta\theta$ where Ω_p is the spiral pattern speed, $\Omega(R)$ is the angular velocity, and the azimuthal angle $\Delta\theta = 2\pi/m$ for m spiral arms. In the solar neighborhood, we find that $t_{coag} \approx 3 \times 10^7$ yr, falling to $\sim 10^7$ yr in the inner galaxy, in agreement with the previously estimated value of t_{diss}. Expression (14) now also yields the global efficiency in the bimodal star formation model. Representative values of coagulation rate, density, cloud velocity dispersion, and pressure, all normalized to the solar neighborhood, are shown in Table 1, together with the massive star formation rate, gas fraction and efficiency for various environments. The parameters are very uncertain in protodisks and ellipticals but similar star formation rates are inferred, whether scaled via the pressure or via the density wave streaming velocity. The massive star formation rate per unit mass of gas is similar between our inner galaxy and the protodisk, and between starbursts and protoellipticals. Only in protoellipticals is the global efficiency enhanced by an order of magnitude, according to (14) because of the considerable velocity dispersion between colliding clouds. This largely suppresses radiative losses in the expanding supernova remnants and wind–driven shells.

5. CONCLUSIONS

I have argued that there are suggestive similarities between star formation in the solar neighborhood and in the inner galaxy and starburst regions, and between the past history of star formation in the disk and, more generally, in protogalaxies. Bimodal star formation offers the most attractive resolution of the longevity of the gas supply in our galactic disk and of the radial gradient in star formation rate. Feedback is necessary to account for the inefficiency at which molecular clouds form stars. This input led to a simple theory for the massive star formation rate. There is one immediate astronomical application, namely one can account for the variation in L_{IR}/L_B measured in spiral galaxies and in starburst galaxies. IRAS observations have shown that L_{IR}/L_B is boosted by up to ~ 100 or even larger, with L_{IR} measuring the rate of massive star formation, and L_B the total stellar mass. The key time–scale that regulates the star formation rate is the cloud coagulation or energy dissipation time. It has been suggested that this is largely controlled by the presence of density waves and non–circular streaming motions that are driven by tidal interactions or by a radial flow induced by a central bar. The gas–stirring time–scale can be increased by up to a factor of 10, as can the gas fraction by a similar factor, relative to our own galaxy and also the global star formation efficiency: this yields the observed enhancement.

In a protogalaxy, similar physical effects are dominant. For example, it is likely that protogalaxy mergers are responsible for formation of slowly rotating luminous ellipticals. Ellipticals are found predominantly in dense regions of the universe, such as galaxy clusters, and these provide a plausible environment for inducing frequent mergers during the cluster formation phase. This alone argues for a similarity between protoellipticals, which then would have formed by mergers of protospirals, and starbursts, which are tidally driven, an argument which has been strengthened by consideration of the past star formation rate in elliptical galaxies. The global star formation efficiency is also inferred from population synthesis modelling to be an order of magnitude larger than in galactic disks today, and the simple gas cloud coagulation model can account for this difference if the primordial gas cloud streaming motions were an order of magnitude larger than in the disk today such a situation would be induced in mergers. One can now qualitatively understand why ellipticals did not develop disks, star formation having been sufficiently efficient to keep the gas stirred up, yet underwent sufficient dissipation to develop dense cores and metallicity gradients. It takes $\sim 10^{10}$ yr to grow a disk, and $\sim 10^9$ yr to form a bulge or an elliptical. The duration of the star forming phase in a protoelliptical is seen simply to be of order (λ/v^2) dynamical time–scales. One fossil memory of the protogalaxy phase is that the present luminosity should be a measure of the burst luminosity, which in turn is expected to be of order $\dot{M}_* \sim \mathrm{M}_g\ (\mathrm{v}\mathrm{v}_{cr}/\lambda)\ \mathrm{t}_{\mathrm{diss}}^{-1} \sim \mathrm{v}^4\ \mathrm{v}_{cr}/\lambda\ \mathrm{G}$ or $\sim 100 v_{200}^4 M_\odot$ yr^{-1}, where $v_{200} \equiv (v/200 kms^{-1})$. If low mass stars indeed formed along with the massive stars, then the fact that the spheroidal components of galaxies satisfy a relation of the form $\mathrm{L} \propto \mathrm{v}^4$ (Faber and Jackson 1976) may be a fossil relic of the protogalactic formation phase. Perhaps a somewhat lower efficiency, intermediate between that of protodisks and luminous protoellipticals, allowed sufficient prolongation of the star formation phase for the recycled matter in low luminosity ellipticals and bulges to dissipate sufficiently to become rotationally supported.

Two other properties of galaxies can be accounted for once the IMF and past star formation rate are known. The chemical evolution of the solar neighborhood can be understood: the paucity of low metallicity stars and the flat age-metallicity relation over the past 10^{10} yr are consequences of bimodal star formation, if the nucleosynthetic yield were higher over the first few gigayears than today. This can be accomplished if the rate of massive star formation was much larger in the past. An alternative suggestion, that post–main–sequence mass loss may have been inhibited in low metallicity stars so that all stars above the Chandrasekhar mass would have become supernovae (Jura 1985), probably does not give satisfactory elemental abundance ratios in metal–poor stars. A modest amount of infall is necessary in the inner galaxy, unless strong radial inflow occurred, in order to produce the radial abundance gradient in our disk.

Finally, the stellar remnants contribute appreciably to the dark matter of the disk. As Larson (1985b) emphasizes, the galactic rotation curve provides the best constraint on the past history of massive star formation in the inner galaxy, and a significant density of white dwarfs is produced. It is possible that this can also account for the dark matter in the solar neighborhood. About fifty percent of the matter by surface density is in remnants in the solar neighborhood, and the bimodal model implies that about two-thirds should be in stellar remnants in the inner galaxy, this being the fraction of O stars that form in the spiral arms. One expects a similar mass fraction of remnants to have formed in elliptical galaxies, with obvious implications for the mass-to-light ratios observed directly and also inferred from population synthesis of old stellar populations.

In summary, it seems likely that the history of the star formation rate and the bimodality of the IMF may provide the keys to unravelling the origin of galactic morphology, the chemical evolution, the dark remnant content of luminous populations, and the future evolution of star–forming regions.

I acknowledge helpful discussions with several individuals in the course of preparing this review, including R. Pudritz, L. Cowie, R. Gusten, L. Hernquist, S. Lizano, F. Palla and other participants in the 1985 ITP, Santa Barbara Workshop on Star Formation. I am especially indebted to Rosemary Wyse for collaborating with me on many aspects of the research described here, and for a critical reading of this manuscript. My research has been supported in part by NSF under grant AST82-13345 and by Calspace.

REFERENCES

Adams, M. T., Strom, K. M. and Strom, S. E. 1983, Ap.J. Suppl. 53, 893.

Augarde, R. and Lequeux, J. 1985, Astr. Ap. (in press).

Bahcall, J. 1984, Ap. J., 276, 169.

Beers, T., Preston, G. and Shectman, S. 1985, A.J., 90, 2089.

Bessel, M. S. and Norris, J. 1984, Ap.J., 285, 622.

Blitz, L. and Stark, A. 1986, Ap.J. Letters (in press).

Bodenheimer, P. 1978, Ap.J. 224, 488.

Bodenheimer, P., Tohline, J. E. and Black, D. C. 1980, Ap.J. 242, 209.

Bond, J. R., Carr, B. and Hogan, C. 1985, preprint.

Bruzual, G. 1983, Ap.J. 273, 105.

Carlberg, R. G., Dawson, P. C., Hsu, T. and Vander Berg, D. A. 1985, Ap.J. (in press).

Chieze, J. P. 1985, preprint.

Cohen, M. and Kuhi, L. V. 1979, Ap.J. Suppl. 41, 743.

Combes, F. and Gerin. 1985, Astr. Ap., 150, 327.

Doom, C., DeGreve, J. P., and de Loore, C. 1985, Ap.J. 290, 185.

Dressler, A. 1980, Ap.J., 236, 351.

Duerr, R., Imhoff, C. L., and Lada, C. J. 1982, Ap.J. 261, 135.

Elmegreen, B. G. 1983, M.N.R.A.S. 203, 1011.

Faber, S. M. and Jackson, R. E. 1976, Ap.J., 204, 668.

Falgarone, E. and Puget, J. L. 1985, preprint.

Gallagher, J. S., Hunter, D. A. and Tutukov, A. V. 1984, Ap.J. 284, 544.

Gaustad, J. E. 1963, Ap.J. 138, 1050.

Goldsmith, P. F., Langer, W. D. and Wilson, R. W. 1985, preprint.

Gusten, R. and Mezger, P. G. 1983, Vistas in Astronomy 26, 159.

Hartwick, F. D. A. 1976, Ap.J. 209, 418.

Hasegawa, T., Yoshii, Y. and Sabano, T. 1981, Astr. Ap., 98, 186.

Hasegawa, T., and Yoneyama, T. 1976, P.A.S.J. 28, 186.

Hayashi, C. and Nakano, T. 1965, Prog. Theor. Phys. 34, 754.

Herbig, G. 1962, Ap.J. 135, 736.

Hillebrandt. W. 1985, in Cosmological Processes (in press).

Hirasawa, T. 1969, Prog. Theor. Phys. 42, 523.

Hoyle, F. 1953, Ap.J., 118, 513.

Iben, I. and Talbot, R. J. 1966, Ap.J. 144, 968.

Iben, I. and Tutukov, A. V. 1984, Ap.J. 282, 615.

Ikeuchi, S. 1981, P.A.S.J., 33, 211.

Joseph, R. D. and Wright, G. S. 1985, M.N.R.A.S. 214, 87.

Jura, M. 1985, preprint.

Kahn, F. D. 1974, Astron. Ap. 37, 149.

Kashlinsky, A. and Rees, M. J., 1983, M.N.R.A.S., 205, 995.

Klein, R. I., Sandford, M. T. and Whitaker, R. W. 1983, Ap.J. Letters 271, L69.

Klein, R. I., Sandford, M. T. and Whitaker, R. W. 1986, Ap.J. (in press).
Lada, C. J., Margulis, M. and Dearborn, D. 1985, Ap.J. (in press).
Lacey, C. G. and Fall, S. M. 1985, preprint.
Larson, R. B. 1981, M.N.R.A.S. 194, 809.
Larson, R. B. 1982, M.N.R.A.S. 200, 159.
Larson, R. B. 1985a, M.N.R.A.S. 214, 379.
Larson, R. B. 1985b, M.N.R.A.S. (in press).
Larson, R. B. and Starrfield, S. 1971, Astron. Ap. 13, 190.
Liebert, J., Jahn, C. C., Gresham, M. and Strittmatter, P. A. 1979, Ap.J. 233, 226.
Low, C. and Lynden-Bell, D. 1976, M.N.R.A.S. 176, 367.
Mathieu, R. D. 1983, Ap.J. Letters 267, 97.
Matsuda, T., Sato, H. and Takeda, H. 1969, Prog. Theor. Phys. 42, 219.
Matteucci, L. and Greggio, L. 1985, Astr. Ap. (in press).
Mezger, P. G. 1987, Proc. I.A.U. Symp. 115 (in press).
Miller, G. E. and Scalo, J. M. 1979, Ap.J. Suppl. 41, 513.
Myers, P. C. 1983, Ap.J. 270, 105.
Myers, P. C. 1985, in Protostars and Planets II(Tucson: University of Arizona Press) (in press).
Nakada, Y. and Yoneyama, T. 1976, P.A.S.J. 28, 61.
Nakano, T. 1984, Fund. Cosm. Phys., 9, 139.
Negroponte, J. 1986, M.N.R.A.S. (in press).
Olaffsson, K., Bergrall, N. and Ekman, A. 1984, Astr. Ap. 137, 327.
Ostriker, J. P. and Cowie, L. L. 1981. Ap.J. Letters, 243, L127.
Palla, F., Stahler, S. and Salpeter, E. 1983, Ap.J., 271, 632.
Rieke, G. H., Catri, R. M., Black, J. H., Kailey, W. F., McAlary, C. W., Lebofsky, M. J. and Elston, R. 1985, Ap.J. 290, 116.
Rieke, G. H., Lebofsky, M. J., Thompson, R. I., Low, F. J. and Tokunaga, A. T. 1980, Ap.J. 238, 24.
Sandage, A. 1985, preprint.
Scalo, J. 1985, Fund. Cosm. Phys. (in press).
Silk, J. 1977a, Ap.J. 214, 784.
Silk, J. 1977b, Ap.J. 211, 638.
Silk, J. 1980, in Star Formation, I. Appenzeller, J. Lequeux and J. Silk, 10th Saas-Fee Course (Geneva:Geneva Observatory, p. 133).
Silk, J. 1985a, Ap.J., 247, 1.
Silk, J. 1985b, Ap.J. Letters, 292, L71.
Silk, J. and Norman. C. 1981, Ap.J., 247, 59.
Solomon, P. 1987, in I.A.U. Symposium 115 (in press).
Spitzer, L. 1978, Physical Processes in the Interstellar Medium (New York:J. Wiley), p.260.
Stahler, S. 1985, Ap.J., 293, 207.

Stahler, S. W., Palla, F. and Salpeter, E. E. 1985, Ap.J. (in press).

Stark, A. 1987, in I.A.U. Symposium 115 (in press).

Stauffer, J. 1984, Ap.J. 280, 189.

Telesco, F. *et al.* 1985, Ap.J. (in press).

Tenorio–Tagle, G., Bodenheimer, P., Lin, D. N. C. and Noriega–Crespo, A. 1985, preprint.

Tinsley, B. M. 1977 Ap. J. 211, 621.

Tinsley, B. 1979, Ap. J. 229, 1046.

Tohline, J. 1980, Ap.J., 239, 417.

Tomisaka, K 1984, P.A.S.J., 36, 457.

Toomre, A. 1981, in The Structure and Evolution of Normal Galaxies, eds. S. M. Fall and D. Lynden–Bell (Cambridge:CUP), p. 111.

Truran, J. W. and Cameron, A. G. W. 1971, Ap. Sp. Sci. 11, 430.

Twarog, B. 1980, Ap.J. 242, 242.

Wilking, B. A. and Lada, C. J. 1983, Ap.J. 274, 698.

Wolfire, M. G. and Cassinelli, J. P. 1985, preprint.

Wilson, J. R., Mayle, R., Wassley, S. E., Weaver, T. A. 1985, Proc. XII Texas Symposium in Relativistic Astrophysics (in press).

Yoneyama, T. 1972, P.A.S.J., 24, 87.

Yorke, H. H. and Krügel, E. 1977, Astron. Ap. 54, 183.

KUTNER: Do abundances and gradients provide us with information on the initial and current star formation rates?

SILK: Abundance ratios are useful constraints on the initial mass function. Disk abundance gradients provide an integrated constraint on several parameters; including the initial rate of star formation, the diffusion of heavy elements in successive cycles of star formation, as well as the initial mass function. Presumably, most of this was determined in the first five billion years ago.

LO: There is an important difference between the 2 enhanced star formation epochs you mentioned: in the recent starburst nuclei, there is clear evidence for a preferential massive star formation, but in the galaxy formation era, large number of low mass stars were formed. Do you have an explanation for this difference?

SILK: A large number of massive stars also formed in a protogalaxy, since rapid chemical evolution occurred. It is likely that comparable mass fractions in low mass and massive stars formed, and this would also be consistent with star formation rates in starburst regions. If one considers bimodal star formation to denote the enhancement to this level of massive star formation relative to that inferred from the Miller-Scalo or Salpeter IMEs, then photoellipticals and starbursts are undergoing similar levels and efficiencies of star formation activity.