Mon. Not. R. astr. Soc. (1974) **169**, 229–245.

EFFECTS OF SUPERNOVAE ON THE EARLY EVOLUTION OF GALAXIES

Richard B. Larson

(Received 1974 July 5)

SUMMARY

During the early evolution of an elliptical galaxy, some of the residual interstellar gas is heated to high temperatures by supernova explosions and is driven out of the galaxy in a galactic wind. The energy supplied per supernova is typically reduced about an order of magnitude by radiative cooling of supernova remnants, but the remaining energy is still sufficient to cause significant gas loss, particularly for small galaxies. In galaxies of smaller mass, gas loss begins earlier and carries away a larger fraction of the initial mass, owing to the lower escape velocity. Model collapse calculations show that the effect of early gas loss is to cause galaxies of smaller mass to have less condensed nuclei, smaller average metal abundances, and smaller metal abundance gradients, in qualitative agreement with the observations.

1. INTRODUCTION

In order to explain the absence of observable interstellar matter in most elliptical galaxies, Mathews & Baker (1971) suggested that the gas lost from the stars in an elliptical galaxy is strongly heated by supernova explosions and driven out of the galaxy in a hot 'galactic wind'. An alternative explanation of the general shortage of gas in elliptical galaxies, suggested by Gallagher (1972), is that the gas produced by evolving stars may be efficiently consumed by continuing star formation. However, detailed model calculations for the formation and evolution of elliptical galaxies (Larson 1974, Paper I; Larson & Tinsley 1974, Paper II) show that, while continuing star formation may indeed reduce the gas content below the observed limit, the gas and the young stars formed from it are concentrated toward the centre of the galaxy and give the nucleus a bluer colour than the outer regions, in contradiction to the observations for most elliptical galaxies. Therefore it appears more likely that gas is indeed lost from elliptical galaxies, and it is of interest to consider whether supernova heating in the models of Papers I and II would be sufficient to generate a galactic wind, as proposed by Mathews & Baker (1971). Furthermore, it is of interest to consider when in the evolution of a galaxy a galactic wind might first become important, since there is some evidence that gas loss may occur at a relatively early stage in the evolution of an elliptical galaxy (Paper II).

Another reason for considering the possibility of early gas loss in elliptical galaxies is the fact that the galaxy models of Paper I, which allow no gas loss, do not readily account for the observed correlation between the metal abundances and the masses of elliptical galaxies (Baum 1959; McClure & van den Bergh 1968; Sandage 1972; Faber 1973). This correlation might be explainable if elliptical galaxies lose some of their recycled and metal-enriched gas at an early stage of evolution, and if the amount of gas lost is greater for galaxies of smaller mass. As was pointed out by van den Bergh (1972), loss of metal-enriched gas might be expected to be particularly important for galaxies of small mass because of their relatively low escape velocities. Observations suggest that galaxies less massive than M32, i.e. galaxies of mass $M \lesssim 4 \times 10^9 M_\odot$, are quite metal poor relative to more massive galaxies; thus early gas loss may have been especially important for galaxies in this mass range. We note that the least massive and most metal poor extragalactic systems known, i.e. the remote globular clusters and dwarf elliptical galaxies such as the Draco system, have such small escape velocities ($\lesssim 20$ km s^{-1}) that it seems unlikely that they could ever have retained a significant fraction of their recycled gas.

In the galactic wind model of Mathews & Baker (MB), it was assumed that in an elliptical galaxy of mass $10^{11} M_\odot$ there is one supernova explosion every 50 yr, and that gas is lost from evolving stars at a rate of about 3 $M_\odot$/yr; with these parameters, the supernova energy input is found to be sufficient to generate a galactic wind. However, Tammann (1974) has recently revised the supernova rate in elliptical galaxies downward by almost a factor of 30 from that assumed by MB to a rate of about one per 1400 yr per $10^{11} M_\odot$; this considerably reduces the supernova energy production rate. At the same time, the detailed galaxy evolution models of Paper I yield a gas production rate of only about 0·1 $M_\odot$/yr per $10^{11} M_\odot$, which is also about a factor of 30 smaller than that assumed by MB; thus the energy requirement for a galactic wind is also reduced. As will be shown in Section 2, the net result appears to be that the production of a galactic wind is, if anything, even more strongly favoured than in the case studied by MB. Using the evolutionary models of Paper I, it is possible to estimate when in the evolution of an elliptical galaxy a galactic wind first becomes important, and we find in Section 3 that this probably occurs within the first $\sim 2 \times 10^9$ yr of a galaxy's lifetime. After this time we expect further star formation to be terminated or at best to occur only sporadically, so that most of the time no effects of star formation would be visible. This prediction is in agreement with the conclusion of Paper II that recent star formation has probably not occurred in most elliptical galaxies.

2. CONDITIONS REQUIRED FOR A GALACTIC WIND

As discussed by MB, if radiative cooling is not important, it is expected that the gas released by evolving stars in an elliptical galaxy will be heated to very high temperatures of the order of 10^7 K, primarily by supernova explosions but also to a lesser extent by the kinetic energy imparted to the gas by stellar motions. As a result, after $\approx 5 \times 10^7$ yr a steady galactic wind is established which expels the gas from the galaxy as fast as it is produced. The condition required for a steady galactic wind to be possible is that the radiative cooling time for the gas in the central part of the galaxy must exceed the time required for the gas to flow out of this region. Whether this condition is satisfied depends on the gas production rate and the supernova energy input rate, and the range of values of these parameters for which a galactic wind is possible is illustrated graphically in Fig. 4 of MB.

In the models of Paper I, the gas production rate after an assumed galactic age of 12×10^9 yr is about 0·1 $M_\odot$/yr per $10^{11} M_\odot$. In units of s^{-1}, this corresponds in the notation of MB to a specific gas production rate of $\alpha_s \sim 3 \times 10^{-20}$ s^{-1}. To estimate the supernova energy production rate in the models of Paper I we first assume, as suggested by Tamman (1974) and discussed further in Paper II, that all supernovae come from relatively massive stars with masses $\gtrsim 4\, M_\odot$; the predicted supernova rate in units of supernova events per 100 yr is then approximately equal to the star formation rate in $M_\odot$/yr (Paper II). Since the star formation rate in the models is essentially equal to the gas production rate, i.e. $\sim 0{\cdot}1\, M_\odot/\mathrm{yr}/10^{11}\, M_\odot$, the predicted supernova rate is $\sim 0{\cdot}1$ SN/100 yr/$10^{11}\, M_\odot$. This is in close agreement with the observed supernova rate in elliptical galaxies, i.e. $\sim 0{\cdot}07$ SN/100 yr/$10^{11}\, M_\odot$ (Tammann 1974); therefore similar results would be obtained even if theoretical estimates were disregarded and only the observed supernova rate were used. Assuming that the kinetic energy of a supernova explosion is $\approx 10^{51}$ erg (e.g. Minkowski 1967), we obtain a total supernova energy production rate of $\approx 3 \times 10^{40}$ erg s^{-1} per $10^{11}\, M_\odot$; this corresponds to $\alpha_{\mathrm{SN}} T_{\mathrm{SN}} \approx 6 \times 10^{-13}$ K s^{-1} in the notation of MB. Referring to Fig. 4 of MB we then find that our predicted values of α_s and $\alpha_{\mathrm{SN}} T_{\mathrm{SN}}$ fall well within the region of the diagram corresponding to a ' hot wind ', and we conclude that an elliptical galaxy should produce a galactic wind if the values of the parameters are anywhere near those obtained here.

We note that if all supernovae come from massive stars, as suggested by Tammann (1974), then the cessation of star formation in an elliptical galaxy after a galactic wind has been established would lead also to the cessation of supernova production, and the primary energy source for the galactic wind would be turned off. If the galactic wind then stops, gas can again begin to accumulate in the galaxy and renewed star formation can occur, until a galactic wind is once again re-established and the gas is blown out. In this way it might be possible for star formation to recur in cycles in an elliptical galaxy, but it seems likely that any such recurrent star formation would take place only in brief bursts and that most of the time no effects of star formation would be observable, in agreement with the observations (Paper II).

Since the condition for a galactic wind appears to be satisfied by an ample margin at age 12×10^9 yr in the models of Paper I, it might be expected that this condition would also be satisfied during most of the earlier evolution of these models. Examination of the time dependence of the supernova rate and the gas production rate in the models shows that this is indeed the case; the condition for a galactic wind is satisfied almost from the beginning, assuming that the criterion of MB is always valid. However, during the earliest stages of galactic evolution the gas density is much higher than at age 12×10^9 yr, and radiative energy losses from expanding supernova remnants may be quite important. Since it is then no longer valid to assume, as was done by MB, that the expansion of supernova remnants is adiabatic and that all of the supernova energy is available for heating the interstellar gas, it is necessary to re-examine the energy input from supernova remnants in situations where radiative cooling of supernova remnants is important.

3. EVOLUTION AND COOLING OF SUPERNOVA REMNANTS

3.1 *Adiabatic phase and the onset of cooling*

The evolution and radiative cooling of supernova remnants (SNR's) has been studied by Cox (1972) using various analytic approximations, and numerical computations which generally confirm the conclusions of Cox have been made by Rosenberg & Scheuer (1973), Chevalier (1974), and Straka (1974). Here we summarize the results of Cox (1972) in so far as they are relevant for the present purposes.

During the early stages of expansion of a SNR, its temperature is so high ($\gg 10^6$ K) and its expansion time scale so short ($< 10^5$ yr) that radiative cooling is unimportant within the expansion time and the expansion is adiabatic. The evolution of the SNR during this period is described by the similarity solution for an adiabatic blast wave, which predicts that the SNR radius varies with time according to

$$R = 0{\cdot}34\,(\epsilon_0/n)^{1/5} t^{2/5}\ \mathrm{pc}, \tag{1}$$

where R is the radius of the shock front at the outer edge of the SNR, ϵ_0 is the initial blast energy in units of 10^{51} erg, n is the (uniform) density of the surrounding interstellar medium in cm^{-3}, and t is the time in years since the supernova explosion. Eventually, after deceleration of the blast wave has reduced the temperature of the outer shell to a value of the order of 10^6 K, the time scale for radiative cooling of the shell becomes comparable with the expansion time, and radiative energy losses become important. The shell then undergoes a transition in structure and becomes much thinner, denser, and cooler, and it begins to efficiently radiate away all of the kinetic energy of newly swept-up interstellar gas. The time at which complete cooling of the shell occurs is given by Cox (1972) as

$$t_c \simeq 5{\cdot}7 \times 10^4\, \epsilon_0^{4/17} n^{-9/17}\ \mathrm{yr}, \tag{2}$$

and the corresponding shell radius is given by equation (1) as

$$R_c \simeq 27\, \epsilon_0^{5/17} n^{-7/17}\ \mathrm{pc}. \tag{3}$$

Because of its much lower density and higher temperature, the interior region of the SNR has a much longer radiative cooling time than the shell and continues to expand adiabatically without significant radiative energy losses. Thus after time t_c the SNR consists of two quite distinct regions: an outer thin, cool shell which contains most of the mass and kinetic energy of the SNR, and an interior hot dilute ' bubble ' of gas which has not experienced significant radiative cooling and contains nearly all of the thermal energy of the SNR.

3.2 *Importance of SNR cooling in model galaxies*

Radiative cooling of SNR's will significantly reduce the energy available to power a galactic wind if cooling becomes important before the SNR's in a galaxy collide and merge with each other, thus becoming part of the general interstellar medium. The lifetime of a SNR before it collides with other SNR's depends on the supernova rate and hence, if supernovae are produced by massive stars, on the star formation rate.* We assume that one supernova is produced for each 100 $M_\odot$ of stars formed (Paper II); then if S denotes the star formation rate in units of $M_\odot/\mathrm{pc}^3/10^9$ yr as used in Paper I, the supernova rate in units of $\mathrm{SN/pc^3/yr}$ is $10^{-11}\,S$. We define a characteristic SNR collision time t_{SNR} as the time required for SNR's forming at the rate $10^{-11}\,S$ to fill up half of the volume of space; a typical SNR of age t_{SNR} will then have half of its volume intersected and reheated by younger SNR's. If the time dependence of SNR radii is given by equation (1), we then obtain

$$t_{\mathrm{SNR}} \simeq 2{\cdot}3 \times 10^5\, (n/\epsilon_0)^{3/11} S^{-5/11}\ \mathrm{yr}. \tag{4}$$

* Even if some supernovae come from low mass stars, the observed fact that the supernova rate is much higher in galaxies with appreciable star formation activity (Tammann 1974) indicates that most of the supernovae in a galaxy with active star formation come from relatively short-lived massive stars.

Radiative energy losses from SNR's will be important if the cooling time t_c is shorter than the collision time t_{SNR}. Using the expressions for t_c and t_{SNR} given in equations (2) and (4), the condition $t_c < t_{SNR}$ can be expressed as a condition on the star formation rate S:

$$S < S_{crit} \simeq 20\, \epsilon_0^{-1 \cdot 12} n^{1 \cdot 76}. \quad (5)$$

(The numerical coefficient in this equation is probably uncertain by at least a factor of 2.) For further applications we shall adopt $\epsilon_0 = 1$ (i.e. a supernova energy of 10^{51} erg); we then have

$$S_{crit} \simeq 20\, n^{1 \cdot 76}. \quad (6)$$

We expect SNR cooling to be important if $S < S_{crit}$, and unimportant if $S > S_{crit}$.

For any region of a galaxy where we know the gas density n (cm^{-3}) and the star formation rate S ($M_\odot/pc^3/10^9$ yr), or equivalently the supernova rate $10^{-11} S$ ($SN/pc^3/yr$), we can evaluate the possible importance of SNR cooling by comparing S with the quantity S_{crit} given by equation (6). For example, in the disc of our galaxy we have, in order of magnitude, $S \approx 10^{-2}$ and $S_{crit} \approx 1$; thus $S \ll S_{crit}$ and SNR cooling should be quite important. At the centre of the galactic wind solution of MB, the assumed supernova rate is equivalent to $S \sim 9$, whereas $S_{crit} \sim 4$; thus $S > S_{crit}$ and SNR cooling should not be important, although the difference between S and S_{crit} is not large. (The margin of safety against SNR cooling would be increased by the revision of parameters discussed in Section 2.)

For the galaxy models of Paper I, both n and S are given directly by the model

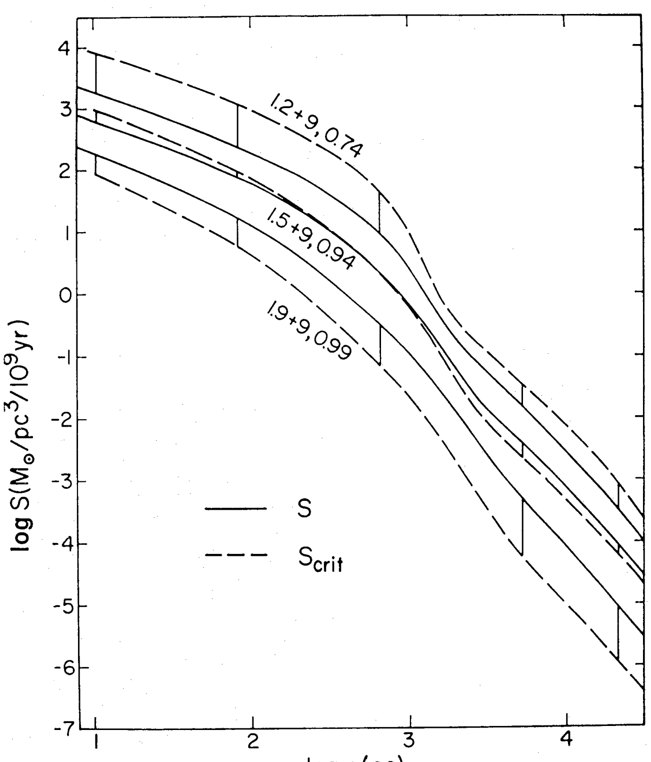

FIG. 1. *The star formation rate S (solid curve) and the quantity S_{crit} defined by equation (6) (dashed curve), both in units of $M_\odot/pc^3/10^9$ yr, are shown as functions of radius for three times in the early evolution of model D. Pairs of curves referring to the same time are joined by vertical lines and are labelled by the time t in years and the fraction of the total mass that has been converted into stars.*

calculations, and S and $S_{\rm crit}$ can be compared in detail as functions of radius and time. As an example, Fig. 1 shows S and $S_{\rm crit}$ plotted as functions of radius at three times in the early evolution of model D. We see that S and $S_{\rm crit}$ show a very similar variation with radius, S being initially less than $S_{\rm crit}$ at all radii, so that SNR cooling is initially important throughout the system. As the gas becomes depleted, $S_{\rm crit}$ decreases more rapidly than S, so that eventually $S > S_{\rm crit}$ at all radii and SNR cooling becomes unimportant everywhere. The transition from $S < S_{\rm crit}$ to $S > S_{\rm crit}$ occurs at nearly the same time ($t \sim 1{\cdot}5 \times 10^9$ yr) at all radii; at this time approximately 94 per cent of the initial protogalactic mass has been transformed into stars. The total stellar gas loss rate ($\sim 8\, M_\odot$/yr) and the supernova rate ($\sim 0{\cdot}4$ SN/yr) in the model at this time satisfy the condition for a steady galactic wind, according to Fig. 4 of MB, so that a galactic wind should be established by $t \sim 1{\cdot}5 \times 10^9$ yr, if it is not already present. The fact that SNR cooling becomes unimportant at about the same time at all radii suggests that gas loss may also begin at about the same time at all radii, as was assumed in the ' sudden gas loss ' models discussed in Paper II.

The above results hold for models A–D of Paper I, in which the star formation rate was assumed to be related to the dynamical time scale. It was also found possible in Paper I to produce plausible models with the assumption that the star formation rate varies as a power of the local gas density, and an example of such a model is given by model F. The star formation rate assumed in model F is $S = 1{\cdot}3\, n^{1{\cdot}85}$, which is very similar to the expression for $S_{\rm crit}$ given in equation (6) except that the numerical coefficient is considerably smaller. As a result, over the range of gas densities encountered in model F, S is always between 10 and 30 times smaller than $S_{\rm crit}$. Thus SNR cooling is always important, and the energy available per supernova for powering a galactic wind is always considerably less than the initial blast energy ϵ_0. In order to determine whether a galactic wind is still possible in such circumstances and to estimate the total amount of gas lost from a galaxy in a galactic wind, it is necessary to consider the later evolution of SNR's and to estimate how much energy is still available for driving a galactic wind in cases where SNR cooling is important.

3.3 *The later evolution of SNR's*

After radiative cooling of the outer shell of a SNR has occurred, the remaining energy of the SNR consists almost entirely of the kinetic energy of the shell and the thermal energy of the hot interior. Since the cooling time for the dense, cool outer shell of an old SNR is very short, it seems probable that when old SNR shells eventually collide with each other or with dense interstellar clouds, their kinetic energy is rapidly dissipated and radiated away. Thus the kinetic energy of old SNR shells does not seem likely to play an important role in heating the interstellar medium or driving a galactic wind. The only remaining energy source for a galactic wind is then the thermal energy of the hot dilute gas in the interiors of old SNR's; because of its very long radiative cooling time, radiative energy losses from this hot gas remain insignificant even after SNR collisions have occurred.

The thermal energy content of an old SNR after complete cooling of the outer shell has occurred has been estimated by Cox (1972) to be

$$E_{\rm th} \sim 0{\cdot}22\, E_0(R/R_{\rm c})^{-2} \tag{7}$$

where E_0 is the initial blast energy and $R_{\rm c}$, given by equation (3), is the shell radius at the time t_c of complete shell cooling. To obtain an approximate expression for the time dependence of E_{th} for times $t > t_c$, we adopt the approximate relation $R \propto t^{0 \cdot 31}$ found from numerical computations by Chevalier (1974); we then have

$$E_{th} \sim 0{\cdot}22\, E_0(t/t_c)^{-0 \cdot 62}. \tag{8}$$

We wish to estimate the residual thermal energy E_r of a SNR at the time $t = t'_{SNR}$ when it collides and merges with other SNR's, thus halting further adiabatic expansion and cooling of its hot interior gas. Here t'_{SNR} differs slightly from t_{SNR} as given by equation (4), because the SNR radius R is no longer given by equation (1) at times $t > t_c$; if we assume instead that $(R/R_c) = (t/t_c)^{0 \cdot 31}$, we find

$$t'_{SNR} \sim 2{\cdot}7 \times 10^5\, \epsilon_0^{-0 \cdot 34} n^{0 \cdot 38} S^{-0 \cdot 52}\ \mathrm{yr}. \tag{9}$$

Putting $t = t'_{SNR}$ in equation (8) and making use of the definition of S_{crit} given in equation (5), we then obtain

$$E_r \sim 0{\cdot}22\, E_0(S/S_{crit})^{0 \cdot 32}. \tag{10}$$

From equation (10) we see that the available energy E_r per supernova depends only weakly on the star formation rate S, so that only an order-of-magnitude value of S/S_{crit} is needed to derive an estimate of E_r. Table I lists some values of the

TABLE I

S/S_{crit}	E_r/E_0
1	$\sim$0·22
0·3	0·15
0·1	0·11
0·03	0·07
0·01	$\sim$0·05

ratio E_r/E_0 for values of S/S_{crit} covering the range of greatest interest. For example, in models A–D of Paper I we have $S/S_{crit} \approx 0{\cdot}1$–$0{\cdot}3$ during the early stages of evolution (Fig. 1); in model F, $S/S_{crit} \approx 0{\cdot}03$–$0{\cdot}1$ at all times; and in the solar neighbourhood, $S/S_{crit} \approx 0{\cdot}01$. We see that under most circumstances likely to be encountered in practice, the residual thermal energy E_r of an old SNR after shell cooling has occurred is given within about a factor of 2 by

$$E_r \approx 0{\cdot}1\, E_0. \tag{11}$$

Since E_0 is of the order of 10^{51} erg, the thermal energy available for driving a galactic wind in cases where SNR cooling is important is of the order of 10^{50} erg per supernova.

In situations where SNR cooling is important, the detailed dynamics of gas loss from a galaxy (if any) will be rather different from the simple galactic wind model studied by MB. When the dense, cool outer shells of old SNR's collide with each other, they will be broken up and may partially coallesce with other SNR shells to form a fragmented, clumpy distribution of dense, cool ' clouds ' of interstellar matter. At the same time, the hot dilute bubbles of gas contained in the SNR cavities will break out and merge with other hot bubbles to form a hot dilute ' intercloud ' medium which fills up most of the space between the dense cool ' cloud ' regions. A very similar concept has been proposed by Cox & Smith (1974), who suggested that the collision of old SNR's in a galaxy would result in the formation of an interconnected network of ' hot tunnels ' containing very hot

($\gtrsim 10^6$ K) low density gas which is maintained at high temperature by continuing supernova explosions. We note that supernova explosions occurring in a region of low gas density are especially effective in reheating the gas because the SNR cooling time is relatively long in a region of low density (equation (2)).

As long as the dense cool ' cloud ' regions are able to survive in approximate pressure equilibrium with a surrounding hot dilute ' intercloud ' medium, it seems likely that the cloud material will stay in the galaxy and perhaps even collapse into new stars, rather than be driven out of the galaxy in a galactic wind. However, the hot dilute intercloud gas may be heated to sufficiently high temperatures by repeated supernova explosions that it acquires enough energy to escape from the galaxy; this is especially likely in the peripheral regions of the galaxy (*cf.* also Chevalier & Gardner 1974). Thus a galactic wind may be generated which involves only the intercloud component of the interstellar gas. If a significant amount of gas is lost in this way and the pressure of the interstellar medium is thereby reduced, the continued existence of dense cool clouds will be made more precarious, and the efficiency of supernova explosions in heating the remaining lower density gas will be increased. A considerable fraction of the gas might thus eventually be heated to high temperatures and driven out of the galaxy in a galactic wind, even though initially $S < S_{\mathrm{crit}}$ and SNR cooling is important.

The true situation is probably even more complicated than sketched above, since in reality magnetic fields and cosmic rays may contribute an amount of energy comparable to the thermal energy of old SNR's, and may play a comparable role in causing or influencing a galactic wind. This possibility is suggested by the fact that in our galaxy the magnetic and cosmic ray energies appear to be comparable to the thermal energy of the interstellar medium. Apparently supernovae or pulsars release a significant fraction of their energy in the form of cosmic rays, and dynamical instabilities resulting from the interaction between cosmic rays and magnetic fields may be instrumental in causing gas to escape from the galactic plane (Wentzel 1972). If such processes are important, the dynamics of gas loss from a galaxy is clearly too complex for any simple quantitative treatment to be possible, and only order-of-magnitude estimates of the energetics can be made.

In order to estimate the amount of gas lost from a galaxy during its evolution, we shall adopt the simplifying assumption that the occurrence of gas loss depends only on the thermal energy of the gas (plus the energy of magnetic fields and cosmic rays, if significant), and that gas is lost from a galaxy whenever its energy content exceeds the energy required for escape. We assume also that the energy released by supernovae is simply stored in the gas without significant losses (other than the radiative cooling of SNR shells) until the gas escapes from the galaxy. In the case of the thermal energy, which is mostly in the hot dilute intercloud component, the assumption of negligible losses is probably justified by the very long cooling time of this hot dilute gas ($\gtrsim 10^8$ yr). If the same assumption can be made for other possibly relevant forms of energy, the amount of gas lost from a galaxy can then be estimated simply on the basis of the amount of energy deposited in the gas.

4. GAS LOSS IN MODEL GALAXIES

4.1 *Estimated amount of gas lost*

In order to estimate when gas loss may become important for the evolution of a galaxy we shall assume, in accordance with equation (11), that approximately

90 per cent of the initial blast energy E_0 of each supernova is radiated away and that the remaining 10 per cent is available for heating the interstellar medium and powering a galactic wind. We also assume that the energy deposited in the interstellar gas by each supernova is stored without further losses until the total energy content of the gas becomes large enough to cause the gas to escape. In the model calculations to be described below, it will be assumed for simplicity that all of the gas in a galaxy is suddenly lost as soon as its total energy content exceeds the escape energy. In reality, the gas is probably not lost all at once, but we would obtain approximately the same prediction for the total amount of gas lost by assuming only that all of the available energy of $\sim 0{\cdot}1\, E_0$ per supernova is expended in removing gas from the galaxy.

If one supernova is produced for every $100\, M_\odot$ of stars formed and if the energy available per supernova for driving a galactic wind is $\approx 10^{50}$ erg, then the total energy available is $\approx 10^{48}$ erg per $M_\odot$ of stars formed. The energy required to cause escape from an elliptical galaxy of mass $10^{11}\, M_\odot$ can be estimated by noting that the escape velocity at the centre of the models of Paper I is about $800\ \mathrm{km\ s^{-1}}$, which corresponds to an escape energy of about $3 \times 10^{15}\ \mathrm{erg\ g^{-1}}$ or $6 \times 10^{48}\ \mathrm{erg}\ M_\odot^{-1}$. Hence if M_s is the mass ($M_\odot$) of stars formed and M_g is the mass remaining in gaseous form, we predict that loss of the remaining gas can occur when the available energy of $\approx 10^{48}\, M_s$ erg is equal to the required energy of $\approx 6 \times 10^{48}\, M_g$ erg, i.e. when $M_g/M_s \approx 1/6$. Thus, in the models of Paper I, gas loss can occur after roughly 6/7 or 86 per cent of the initial mass has been transformed into stars; the remaining $\approx 1/7$ or 14 per cent of the initial mass is lost in a galactic wind. This result is consistent with the conclusion previously obtained for model D that a galactic wind would have been established by the time that ~ 94 per cent of the initial mass has been converted into stars (Section 3.2). Thus it seems reasonable to conclude that for a galaxy of mass $10^{11}\, M_\odot$ gas loss will become important by the time that roughly 90 per cent of the initial mass has been transformed into stars, and that the remaining ≈ 10 per cent of the mass will be lost in a galactic wind.

Because of the lower escape velocity of a galaxy of mass less than $10^{11}\, M_\odot$, gas loss will begin earlier in such a galaxy and a larger fraction of the initial mass will be lost. To obtain an estimate of how the escape velocity scales with the total mass M, we assume that protoglaxies of different mass begin with approximately the same initial density, so that $M \propto R^3$ where R is the radius of the protogalaxy. (The results are qualitatively similar for any physically reasonable scaling relationship between M and R.) The escape energy per unit mass then varies as $M/R \propto M^{2/3}$, and for a protogalaxy or galaxy of mass M it is given approximately by

$$E_{\mathrm{esc}}/M_g \approx 6 \times 10^{48}\, (M/10^{11})^{2/3}\ \mathrm{erg}\ M_\odot^{-1}, \tag{12}$$

where we have made use of the previous estimate of $E_{\mathrm{esc}}/M_g \approx 6 \times 10^{48}\ \mathrm{erg}\ M_\odot^{-1}$ for a galaxy of mass $10^{11}\, M_\odot$. Equating the available energy $\approx 10^{48}\, M_s$ to the energy E_{esc} required to cause escape of the remaining gas, we predict that a protogalaxy of mass M $(= M_s + M_g)$ will lose its remaining gas when

$$M_g/M_s \approx \tfrac{1}{6}(M/10^{11})^{-2/3} \approx (M/6 \times 10^9)^{-2/3}. \tag{13}$$

For example, if the initial mass M is $6 \times 10^9\, M_\odot$, we find from equation (13) that gas loss occurs when $M_g \approx M_s \approx 3 \times 10^9\, M_\odot$, so that $3 \times 10^9\, M_\odot$ of gas is lost and a galaxy of mass $3 \times 10^9 M_\odot$ remains. Thus we expect the effects of early gas loss to be of major importance for galaxies of mass $\lesssim 3 \times 10^9 M_\odot$.

4.2 *Model calculations and results*

To illustrate the possible effect of early gas loss on the evolution of galaxies of different mass, we have computed several models which are identical to model D of Paper I except that the remaining gas is artificially removed from the galaxy at various stages when the fractional mass in stars M_S/M has attained values ranging from 0·20 to 0·90. It has been assumed that after gas loss occurs the system remains permanently devoid of gas and no further star formation occurs. For convenience in comparing the results the initial mass has been left constant at $10^{11} M_\odot$ in all cases, but each model is to be considered as representing a galaxy of different initial mass and radius, the total mass M being related to the assumed value of M_S/M through a relation similar to equation (13) (see below), and the radius R scaling with mass according to $R \propto M^{1/3}$. Models were computed in which gas loss occurs when M_S/M attains values of 0·20, 0·30, 0·40, 0·50, 0·60, 0·75 and 0·90; the corresponding galaxy masses M_S implied by equation (13) are approximately $1{\cdot}5 \times 10^8$, 5×10^8, $1{\cdot}3 \times 10^9$, 3×10^9, 7×10^9, 2×10^{10} and $1{\cdot}5 \times 10^{11} M_\odot$, respectively.

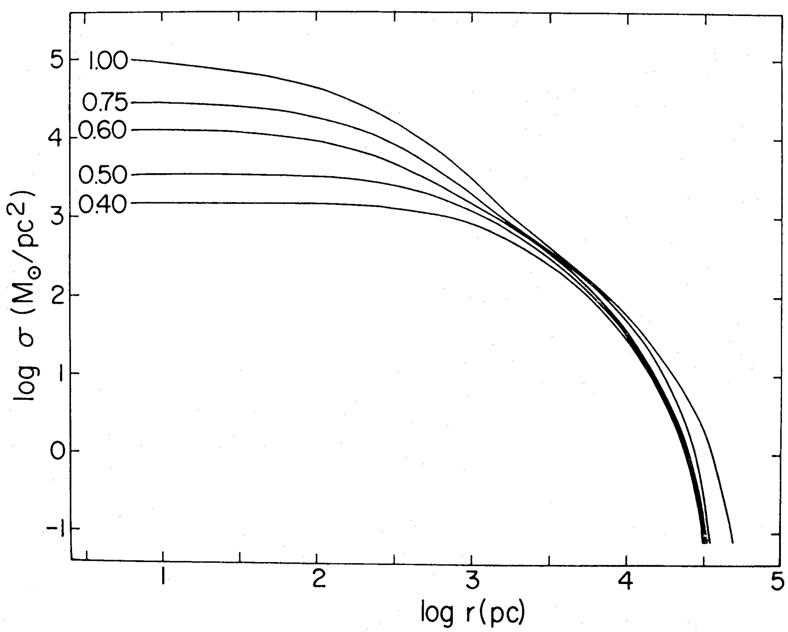

FIG. 2. *The projected stellar density distributions σ(r) obtained for models in which the remaining gas is removed when the fractional mass in stars M_S/M has attained values of* 1·00, 0·75, 0·60, 0·50 *and* 0·40.

There are two principal consequences of early gas loss for the evolution of these model galaxies. First, since the nuclear region of a galaxy tends to form last through the infall of gas left over after the formation of the halo (Paper I), one effect of early gas loss is to inhibit or prevent the formation of a condensed nucleus; thus systems which lose their residual gas at relatively early stages of evolution develop less pronounced nuclei. This is illustrated in Fig. 2, which shows the projected stellar density distributions obtained for models in which gas loss occurs when M_S/M is equal to 1·00, 0·75, 0·60, 0·50 and 0·40 (the first case is idential to model D of Paper I). Since these models represent galaxies of decreasing mass, we see that galaxies of smaller mass develop less prominent nuclei, and that galaxies of sufficiently small mass may form no distinct nucleus at all. This prediction is in good qualitative agreement with the observed fact that giant elliptical galaxies always possess distinct nuclei, while the dwarf ellipticals do not (van den Bergh 1972). According to van den Bergh, the dividing point between giant elliptical galaxies with condensed nuclei and dwarf ellipticals without nuclei occurs in the vicinity of absolute magnitude -15, which corresponds to a mass of the order of $2 \times 10^9 M_\odot$. For comparison, the present calculations predict that no distinct nuclear condensation is formed if more than ≈ 50 per cent of the initial protogalactic mass is lost; this corresponds to a final galactic mass less than $\approx 3 \times 10^9 M_\odot$, if equation (13) is adopted.

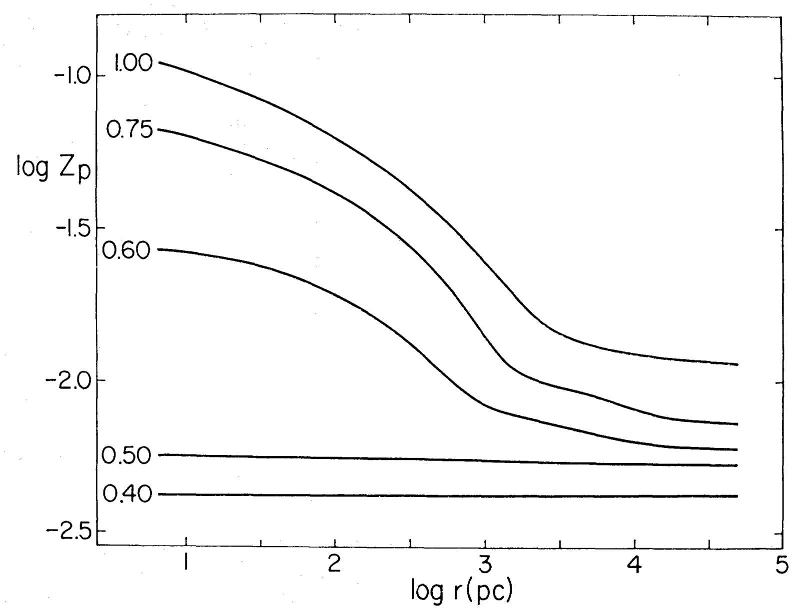

FIG. 3. *The projected stellar metal abundance distributions $Z(r)$ for models in which the remaining gas is removed when the fractional mass in stars M_S/M is equal to* 1·00, 0·75, 0·60, 0·50 *and* 0·40.

The second important consequence of early loss of the metal-enriched residual gas in a forming galaxy is that the average metal abundance of the resulting galaxy is reduced. Furthermore, the metal abundance gradient predicted in Paper I is reduced or eliminated, since the metal-enriched gas which would normally have condensed at the centre to form a ‘ super metal rich ’ nucleus is now lost from the system. These effects are illustrated in Fig. 3, which shows the projected metal abundance Z_p (i.e. the metal abundance averaged along lines of sight) as a function of radius in models in which gas loss occurs when the fractional stellar mass M_S/M attains values of 1·00, 0·75, 0·60, 0·50 and 0·40. It is evident that the effect of gas loss is to reduce the final metal abundance at all radii and to reduce the metal abundance gradient in the nuclear region. The metal abundance gradient disappears completely if gas loss occurs before ≈ 50 per cent of the initial mass has been converted into stars, which we expect to happen for galaxies of mass less than $\approx 3 \times 10^9 M_\odot$.

The detailed results are, of course, quantitatively somewhat different for models computed with different assumptions. For example, in gas loss calculations performed with model F of Paper I the resulting metal abundance gradient is steeper at large radii, and it is reduced but never completely eliminated by early gas loss. However, the results are still qualitatively similar to those described above, and we conclude that the effect of early gas loss in very small galaxies (mass $\leqslant 3 \times 10^9 M_\odot$) is to substantially reduce both the average metal abundance and the metal abundance gradient.

In order to predict how the average metal abundance $\langle Z \rangle$ of an elliptical galaxy should depend on its mass, we have calculated $\langle Z \rangle$ for each of the present models and plotted the result *vs* M_s/M, the fractional mass converted into stars, in Fig. 4.

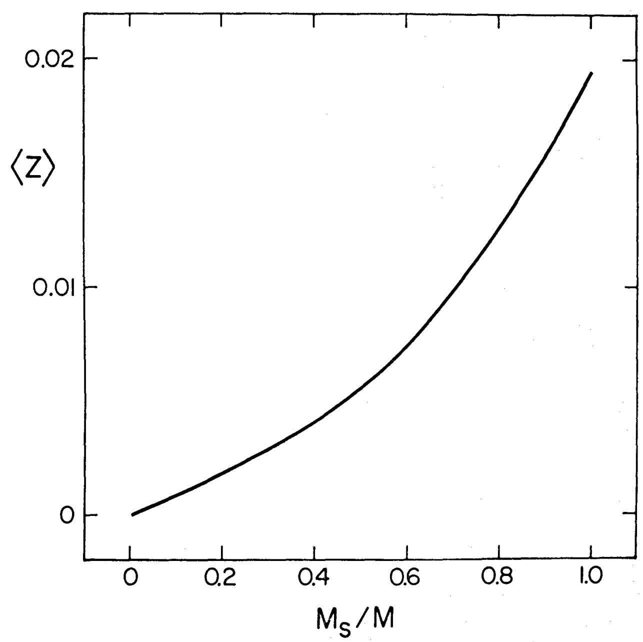

FIG. 4. *The average stellar metal abundance* $\langle Z \rangle$ *for models with gas loss is plotted vs the fractional mass in stars* M_s/M *at the time when the residual gas is lost.*

The quantity M_s/M can be calculated as a function of the initial mass M or of the final galactic mass M_s by using equation (13) or a similar relation. Since all of the numerical quantities used in deriving equation (13) are quite uncertain, we shall instead use the following two simple relations which are similar to equation (13) but may give a more realistic indication of the range of uncertainty:

$$M_g/M_s = (M/10^9)^{-2/3} \tag{14a}$$

$$M_g/M_s = (M/10^{10})^{-2/3}. \tag{14b}$$

Using these equations to calculate M_s/M as a function of the final galactic mass M_s, and obtaining the corresponding value of $\langle Z \rangle$ from Fig. 4, we find the results shown in Fig. 5 for the dependence of $\langle Z \rangle$ on M_s; here the two curves correspond to equation (14a) (upper curve) and equation (14b) (lower curve). The horizontal dashed line indicates the value of $\langle Z \rangle$ corresponding to $M_s/M = 0{\cdot}50$, which is the approximate dividing point between giant elliptical galaxies with condensed 'super metal rich' nuclei and dwarf elliptical galaxies without condensed nuclei or strong metal abundance gradients.

The results shown in Fig. 5 are in good qualitative agreement with various observational data indicating a systematic dependence of metal abundance on mass for elliptical galaxies. For example, the data of Baum (1959) for the variation of $B-V$ with luminosity, if interpreted as indicating a dependence of metal abundance Z on luminosity, indicate an increase in Z with increasing luminosity up to an absolute magnitude of about -17 ($M \approx 10^{10} M_\odot$), followed by a levelling off at higher luminosities. A somewhat similar result was obtained by Faber (1973) who

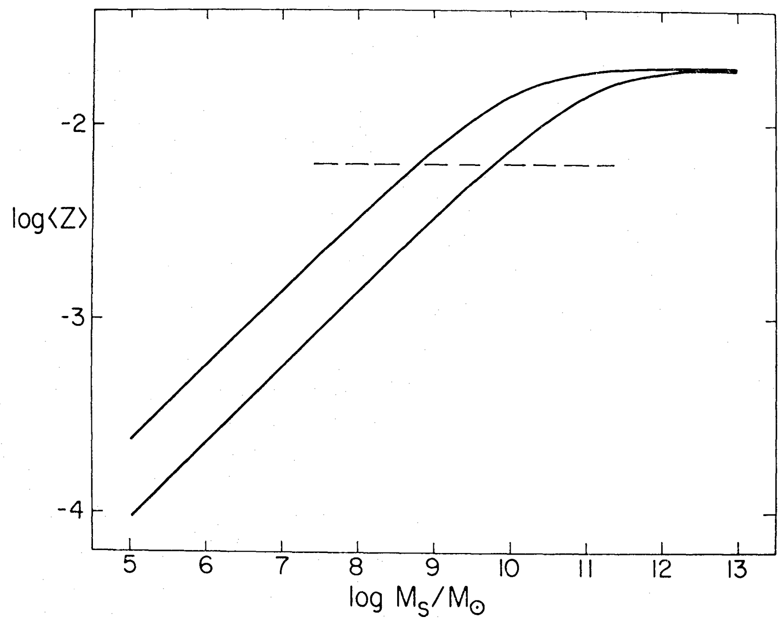

FIG. 5. *The predicted dependence of the average stellar metal abundance $\langle Z \rangle$ on the final galactic mass M_s (solar units). Thr upper curve has been calculated using equation (14a) and the lower curve using equation (14b). The horizontal dashed line indicates the value of $\langle Z \rangle$ corresponding to 50 per cent mass loss, i.e. M_s/M = 0·50.*

found a strong increase of Z with mass for galaxies less massive than M32 ($M \sim 4 \times 10^9 M_\odot$), followed by a more gradual increase in Z at larger masses, amounting to about a factor of 2 between M32 and the most massive galaxies studied ($M \approx 10^{12} M_\odot$). These observations are roughly in agreement with the predicted variation of Z with mass shown in Fig. 5, but a quantitative comparison is difficult because of the difficulty in determining accurately the relationship between colour and metal abundance for elliptical galaxies (Paper II). For what it is worth, we note from Fig. 5 that the predicted increase in Z from small systems (e.g. globular clusters) of mass $10^5 M_\odot$ to giant elliptical galaxies of mass $\gtrsim 10^{11} M_\odot$ is about two orders of magnitude, in rough agreement with the observed difference in Z between metal-poor globular clusters and ' normal ' metal-rich systems. Also, the predicted range in Z is consistent with the observed range of about 0·3 in $B-V$ (Baum 1959) if we adopt the rough relation $\Delta \log Z/\Delta(B-V) \sim 7$ obtained in Paper II. There is possibly a discrepancy in detail between the predicted and observed $Z(M)$ relations in that the data of Baum (1959) and Faber (1973) seem to indicate a more rapid variation of Z with mass than predicted in the mass range $10^8 \lesssim M \lesssim 10^{10} M_\odot$. Also, the data of McClure & van den Bergh (1968) and

Sandage (1972) do not show the predicted levelling-off of the $Z(M)$ relation at large masses.

There are many uncertainties affecting the comparison between predicted and observed metal abundances in galaxies of different masses. On the observational side, a better knowledge of the relation between photometric indices and metal abundances will be required before a detailed quantitative comparison is possible. We note also that the observations may not provide an unbiased measure of the average metal abundance of a galaxy if they refer mainly to a bright but possibly atypically metal rich central region; if this is the case, the observed correlation between metal abundance and mass could be influenced or caused in part by the predicted increase in concentration of heavy elements toward the centre with increasing galactic mass (Fig. 3). On the theoretical side, the crudeness of the present models and the uncertainties in the various parameters are such that only qualitative predictions can be considered reliable. Some complicating effects which have been neglected in the present models but are probably important in reality are discussed below in Section 5. In view of the many uncertainties, the agreement between the model predictions and the observed properties of elliptical galaxies seems satisfactory, and supports the conclusion that the mass dependence of the structure and composition of elliptical galaxies is largely a result of early gas loss, which has the greatest effect for the least massive galaxies.

5. POSSIBLE MODEL REFINEMENTS

In several respects, the models discussed above give an oversimplified picture of the early evolution of galaxies. For example, when applied to systems of very small mass these models imply that a large fraction of the initial protogalactic mass is lost at a very early stage of evolution, before there has been time for significant collapse to occur. If a protoglaxy were to lose most of its mass at this stage, the remaining system might become gravitationally unbound and not collapse to form a galaxy at all. However, such a situation would not arise in reality because star formation, supernova production, and gas loss can occur only after collapse to high densities has already taken place, at least in localized regions. If the temperature in a protogalaxy is of the order of 10^4 K, the Jeans mass is of the order of 10^7–$10^9\ M_\odot$ (Larson 1969), so we might expect gravitational collapse to occur first in regions of this size. Calculations of the collapse of interstellar clouds always show that a collapsing cloud develops one or more small condensations of much higher than average density, within which stars first begin to form; hence, by analogy, we expect that star formation in a protogalaxy will begin in dense condensations or protoclusters which are surrounded at first by extended envelopes of uncondensed gas. Since the time scale for formation and evolution of the massive stars that become supernovae ($\approx 10^6$–10^7 yr) is shorter than the time required for infall of the gaseous envelope of a protocluster ($\approx 10^7$–10^8 yr), it is probably in this type of situation that supernova explosions first become important.

If we consider for example a system whose initial mass is only $10^7\ M_\odot$ and therefore no larger than the Jeans mass, we expect gravitational collapse to produce a single central condensation or protocluster surrounded initially by an extended gaseous envelope containing most of the mass of the system. According to the assumptions of Section 4.2 (equation (14)), the energy produced by supernovae becomes sufficient to blow away the remaining uncondensed gas after about

$1\text{–}4 \times 10^5\, M_\odot$ of material has condensed into stars. If these stars form in a dense gravitationally bound cluster, this cluster need not be disrupted by supernova explosions, and only the surrounding envelope of uncondensed gas will be blown away. The resulting star cluster, with a mass of $\approx 1\text{–}4 \times 10^5\, M_\odot$ and a metal abundance of $\langle Z \rangle \approx 1\text{–}4 \times 10^{-4}$ (Fig. 5), would have properties very similar to those of globular clusters; thus it seems plausible that star formation in a protogalaxy actually begins with the formation of globular clusters. Also it seems possible that in the formation of a small stellar system like a globular cluster or a dwarf elliptical galaxy, a major fraction of the initial gas mass can be lost without the stellar system itself being disrupted.

It is obvious that our assumption (Paper I) that heavy elements are produced continuously and are always completely mixed with the surrounding gas cannot hold in detail if star formation and heavy element production occur in localized dense regions or protoclusters, as suggested above. The heavy elements produced in a forming cluster will not be uniformly mixed throughout the associated contracting cloud but will presumably remain concentrated in the central region where stars are actively forming. Therefore any subsequently formed stars will tend to have higher metal abundances than if the heavy elements were uniformly distributed throughout the contracting protocloud. On the other hand, it requires a finite time of the order of 10^7 yr or longer for heavy elements to be produced and incorporated in a second generation of stars, and since this is of the same order as the time required for supernova explosions to blow out the remaining gas from a forming cluster, it is possible that further star formation could be cut off before there has been time for significant metal enrichment to take place. In this case the metal abundance of a globular cluster or dwarf elliptical galaxy would be lower than predicted by Fig. 5. This might explain the extremely low metal abundances of some globular clusters and dwarf elliptical galaxies, such as the Draco system ($M \approx 10^5\, M_\odot$) which has a metal abundance Z of only $\approx 3 \times 10^{-5}$ (Hartwick & McClure 1974).

Another possible effect which may tend to reduce the metal abundances of small systems is that the heavy elements produced by supernovae may not become well mixed with the dense cool regions of the interstellar medium in which new stars are formed, but may instead remain mostly confined to the hot dilute component of the gas which cools very slowly and may be lost from the galaxy in a galactic wind. If this is an important effect, the gas which is lost from a galaxy in a galactic wind may contain a higher abundance of heavy elements than the gas which remains in the system to form new stars.

It is of interest to consider what might happen to the metal-enriched gas that is predicted to be lost from many galaxies at an early stage of evolution. Since galaxies generally occur in groups or clusters, and since the escape velocity of a cluster of galaxies usually exceeds the escape velocity of an individual galaxy, much of the gas lost from the galaxies in a cluster may remain bound within the cluster. Some of this metal-enriched intracluster gas may subsequently be accreted again by galaxies, particularly by the most massive galaxies in the cluster. If a massive galaxy acquires in this way a significant amount of the metal-enriched gas lost from other galaxies, it may attain an average metal abundance that is even higher than would be possible for a completely isolated galaxy. One might expect this effect to be particularly important for the central giant elliptical galaxies that dominate many clusters of galaxies. Possibly such an effect could account for the apparent continuing gradual increase of Z with mass which is indicated by the obsrevations (e.g. Faber 1973) even at very large masses where Fig. 5 predicts that Z should become asymptotically constant.

6. CONCLUSIONS

If conventional estimates of the effects of supernovae in heating the interstellar gas in a galaxy are correct, then supernova-driven gas loss must almost certainly occur at some stage in the evolution of most elliptical galaxies. Our discussion in Sections 3 and 4 indicates that gas loss probably begins during the first $1\text{–}2 \times 10^9$ yr of a galaxy's lifetime. If the type I supernovae observed in elliptical galaxies are produced at a sufficient rate by low mass stars, gas loss may occur continuously in a steady galactic wind; on the other hand, if type I supernovae are associated with relatively massive stars (Tammann 1974; Larson & Tinsley 1974) and thus occur only after recent star formation has taken place, gas loss would occur in more sporadic fashion, possibly alternating with periods of infall.

The time at which gas loss begins and the total amount of gas lost from a galaxy depend on the mass of the galaxy: a galaxy of smaller mass loses a larger fraction of its initial mass because of its lower escape velocity. Since radiative cooling of supernova remnants is probably important and may cause up to ≈ 90 per cent of the supernova energy to be radiated away, and since magnetic fields and cosmic rays may also play a role in causing gas to be lost, the dynamics of gas loss is probably very complex, and detailed calculations have not been attempted. However, simple model calculations (Section 4) based on the assumption that gas is lost from a galaxy as soon as it acquires sufficient energy to escape predict that (1) galaxies of smaller mass should show less prominent nuclei, and (2) galaxies of smaller mass should have both a lower average metal abundance and a smaller metal abundance gradient. These predictions are in good qualitative agreement with the observations; thus if gas loss is allowed for, collapse models like those of Paper I are able to account not only for the typical structure of elliptical galaxies but also for the variation of their properties with mass. The prediction that a galaxy develops a condensed nucleus only if most of the initial gas is retained in the galaxy and is allowed to condense at the centre illustrates clearly the fact that in these models the nucleus forms last through the central condensation of residual gas. As was pointed out in Paper I, this central condensation of residual gas may play a role in the production of quasar activity in the nucleus of a galaxy; if so, we would predict from the present results that quasar activity can occur only in relatively massive galaxies which develop highly condensed nuclei.

ACKNOWLEDGMENTS

I have benefited from conversations and correspondence with Drs D. P. Cox, S. M. Faber, J. P. Ostriker, and especially Beatrice Tinsley, who first called to my attention the importance of understanding the mass dependence of the properties of elliptical galaxies.

Yale University Observatory, New Haven, Connecticut 06520

Received in original form 1974 *May* 31

REFERENCES

Baum, W. A., 1959. *Publ. astr. Soc. Pacific*, **71**, 106.
Bergh, S. van den, 1972. In *External galaxies and quasi-stellar objects, IAU Symposium No.* 44, ed. D. S. Evans, p. 1, D. Reidel, Dordrecht, Holland.
Chevalier, R. A., 1974. *Astrophys. J.*, **188**, 501.
Chevalier, R. A. & Gardner, J., 1974. *Astrophys. J.*, **192**, 457.
Cox, D. P., 1972. *Astrophys. J.*, **178**, 159.
Cox, D. P. & Smith, B. W., 1974. *Astrophys. J. Lett.*, **189**, L10.
Faber, S. M., 1973. *Astrophys. J.*, **179**, 731.
Gallagher, J. S., 1972. *Astr. J.*, **77**, 568.
Hartwick, F. D. A. & McClure, R. D., 1974. *Astrophys. J.*, **193**, in press.
Larson, R. B., 1969. *Mon. Not. R. astr. Soc.*, **145**, 405.
Larson, R. B., 1974. *Mon. Not. R. astr. Soc.*, **166**, 585 (Paper I).
Larson, R. B. & Tinsley, B. M., 1974. *Astrophys. J.*, **192**, 293 (Paper II).
Mathews, W. G. & Baker, J. C., 1971. *Astrophys. J.*, **170**, 241.
McClure, R. D. & Bergh, S. van den, 1968. *Astr. J.*, **73**, 1008.
Minkowski, R., 1967. In *Nebulae and interstellar matter, Stars and stellar systems Vol.* 7, p. 623, eds B. M. Middlehurst & L. H. Aller, University of Chicago Press.
Rosenberg, I. & Scheuer, P. A. G., 1973. *Mon. Not. R. astr. Soc.*, **161**, 27.
Sandage, A. R., 1972. *Astrophys. J.*, **176**, 21.
Straka, W. C., 1974. *Astrophys. J.*, **190**, 59.
Tammann, G. A., 1974. In *Supernovae and supernova remnants*, ed. C. B. Cosmovici, D. Reidel, Dordrecht, Holland (in press).
Wentzel, D. G., 1972. *Publ. astr. Soc. Pacific*, **84**, 225.