

# Chapter 14

## CHEMICAL EVOLUTION AND DISK GALAXY FORMATION

### 14.1 Introduction

In this set of lectures I will discuss some aspects of the formation of galaxies in view of the various properties and their distributions that I have discussed in my other contributions to this course. For that purpose I will need to review first some aspects of the distributions of metallicity in disks and bulges and some schematic background relating to the chemical evolution in galaxies. Those aspects that are particularly related to the solar neighborhood are covered by the other lecturers of this course.

### 14.2 Abundance gradients in galaxies and chemical evolution

Previously in this course I have discussed in detail the evidence for abundance gradients in bulges. This comes from color changes of the integrated light as a result of a hotter giant branch, more to the blue extended horizontal branch and less line blanketing in old, metal-poor populations. As a result the radial bluing in bulges is interpreted as a progressive decrease in mean stellar metallicity with increasing galactocentric radius. The amplitude of the effect is comparable to the total color variation in the integrated light of Galactic globular clusters, where the known abundance  $[Fe/H]$  has been measured independently and therefore the color variation can be calibrated. Consequently, the observed variation of abundance is of the rough order of one to two dex. Here it should be noted that the

notation  $[\text{Fe}/\text{H}]$  or any two other chemical elements stands for the logarithm of the density ratio, in this case normalized to the solar value.

In disks the abundances are most easily measured in the gas from emission lines from the HII regions. It was advocated first by Searle (1971), that the radial change in the line ratio of  $[\text{OIII}]$  at 4959 and 5007 Å and  $\text{H}\beta$  (often called the “excitation”) that he observed in a few nearby galaxies, was a result of variations of the oxygen to hydrogen ratio and therefore probably the heavy element abundance in the HII regions. This is not entirely obvious, because the observed line ratio’s in emission line spectra depend also on the continuum spectrum (effective temperature) of the ionizing star(s) and therefore the ionization level of the gas and on the electron temperature. Searle proposed three alternatives to interpret his observations, namely a gradient in the stellar temperatures, a changing dust content, which would affect the stellar radiation field and an abundance gradient. He then concluded that the main effect of the changing line ratio’s had to be attributed to variations in chemical composition of the gas. Since then many studies have been done that confirm both the observed change in line ratio and the interpretation that these are most likely due to abundance gradients in the disks.

The excitation  $[\text{OIII}]/\text{H}\beta$  increases with increasing stellar temperature and decreases with increasing oxygen abundance relative to hydrogen  $[\text{O}/\text{H}]$ . The latter is the case for the following reason. The cooling of the HII regions occurs through emission in optical lines, in particular the nebular lines from  $\text{O}^+$  and  $\text{N}^+$  in the neutral He-zone and  $\text{O}^{++}$  in the  $\text{He}^+$ -zone dominate this effect. Now, a lower abundance (of oxygen) reduces the cooling and increases the electron temperature. Then the  $[\text{OIII}]$ - (and also  $[\text{OII}]$ -) lines become stronger relative to  $\text{H}\beta$ .

As a example of observed gradients, I will take the recent observations of the Sab spiral M81 (Garnett and Shields, 1987), which is the earliest type in which such a study has been done. These authors use two methods to derive abundances from the observed line ratio’s. For this interpretation of the spectra one needs to know the electron temperature. The first method is the empirical one of Pagel et al. (1980), which is based on a range of HII regions in various galaxies that span the range of observed spectra. Pagel et al. have used bright regions to either measure the electron temperatures (which is possible using weak emission lines, such as  $[\text{OIII}]$  at 4363 Å so that the population of two upper levels of an atom can be inferred), so that the abundance can be inferred, while for others they used photoionization models. This calibrates the relation between excitation and metallicity. Furthermore, Garnett and Shields used their own photoionization models to infer abundances. The result is a radial gradient of  $[\text{O}/\text{H}]$  of  $-0.08 \text{ dex kpc}^{-1}$  between 3 and 15 kpc radius. This is a typical value, also for late type galaxies. There is no discernable gradient in  $[\text{O}/\text{N}]$ .

Before proceeding to discuss the observations, we need some general results from the theory of chemical evolution. For this define in a closed region of a galaxy the mass in gas

as  $M_g$ , in heavy elements  $M_Z$  and in stars  $M_*$ . The abundance then is  $Z(t) = M_Z(t)/M_g(t)$ . Next define the yield  $y$  (Searle and Sargent, 1972) as the mass in new metals ejected by heavy stars, when a unit mass is locked in long-lived stars. Then assuming that short-lived stars evolve instantaneously and that the products of nucleosynthesis are also instantaneously mixed in the interstellar medium, we have

$$\frac{dM_Z}{dt} = y \frac{dM_*}{dt} - Z(t) \frac{dM_*}{dt}, \quad (14.1)$$

$$\frac{dM_g}{dt} = -\frac{dM_*}{dt}. \quad (14.2)$$

Starting with an initial abundance  $Z_o = 0$ , this gives the “simple model”, which has as solution

$$Z(t) = y \ln \left( \frac{1}{x} \right) \quad \text{with} \quad x = \frac{M_g(t)}{M_{\text{tot}}}. \quad (14.3)$$

The fraction of stars with abundance  $\leq Z$  at time  $t$  (when the gas abundance is  $Z(t) = Z_1$  and  $x(t) = x_1$ ) is then

$$F(Z) = \frac{1 - x}{1 - x_1}. \quad (14.4)$$

This solution gives for the solar neighborhood the well-known G-dwarf problem, which states that there are too few stars of low metal abundance. It is also possible to calculate the mean stellar metallicity at any time. For this  $F(Z)$  is first converted by differentiation into a number distribution. The result then is after dropping the subscript 1

$$\langle Z \rangle = y \frac{1 - x(1 - \ln x)}{1 - x}. \quad (14.5)$$

We see, that as gas consumption is completed and  $x \rightarrow 0$ , that  $\langle Z \rangle \rightarrow y$ , the yield itself. This is the mean metallicity of the stars, which should be distinguished from the abundance  $Z(t)$  of the gas, which goes to infinity in this instantaneous recycling approximation.

There are various ways of solving the G-dwarf problem with minor extensions of the simple model. The first is to set  $Z_o \neq 0$  (prompt initial enrichment). This can occur, if chemical enrichment of the gas has proceeded in another position of the galaxy prior to settling in the disk, such as in particular in the “thick disk” as has been suggested by Gilmore and Wyse (1986). A second possible solution is that in the early phases formation of massive stars is more pronounced compared to that in light stars, so that few stars of low metallicity survive until the present time. This is a possibility in Larson’s (1986) hypothesis of bimodal star formation, in which formation of both mass-groups of stars is independent and should then be heavily weighed towards massive stars in the initial

phases of disk star formation. The equations above for the simple model survive then as long as one everywhere replaces  $Z$  with  $Z - Z_o$ . For complete gas consumption  $\langle Z \rangle \rightarrow y + Z_o$ . I will refer to this as the extended simple model.

A second possibility is to postulate an inflow of unprocessed material into the disk, such that

$$\frac{dM_g(t)}{dt} = -\frac{dM_*}{dt} + f(t). \quad (14.6)$$

The extreme model of this kind has  $M_g = \text{constant}$  and then the solution becomes

$$Z(t) = y\{1 - \exp(-\mu)\} \quad \text{with} \quad \mu = \frac{M_*(t)}{M_g}, \quad (14.7)$$

$$F(Z_1) = \frac{\mu}{\mu_1}, \quad (14.8)$$

$$\langle Z \rangle = y - \frac{y}{\mu} + \frac{y}{\mu} \exp(-\mu). \quad (14.9)$$

Eventually we will have  $M_* \gg M_g$  and  $\mu \gg 1$ ; then  $\langle Z \rangle \rightarrow y$ . The important feature of this model is that  $Z(t)$  very quickly increases in the early phases (because there is less gas, it also is more easily enriched) and  $Z(t)$  approaches  $y$  asymptotically.

For the discussion of the gradients in bulges we also have to consider the case that enriched material is lost from the area. In that case we have

$$\frac{dM_Z}{dt} = y \frac{dM_*}{dt} - Z(t) \frac{dM_*}{dt} - Z(t)g(t), \quad (14.10)$$

$$\frac{dM_g}{dt} = -\frac{dM_*}{dt} - g(t). \quad (14.11)$$

The illustrative model in this case has  $g(t)$  proportional to the star formation rate  $g(t) = \alpha dM_*/dt$ . This could in practice be the case, when star formation through supernova explosions heats up the interstellar medium, which subsequently is blown away. It is straightforward to see that the equations of the simple model recover in this case, but with  $dM_*/dt$  multiplied by a factor  $(1 + \alpha)$  and an “effective yield”  $y' = y/(1 + \alpha)$ . The solution then has the same form as the simple model with  $y$  replaced with  $y'$ .

Now let us look at observations. First the abundance gradients in the bulges. I will discuss below that any gradient is probably conserved through the collapse. The hypothesis then is that in the inner regions of the protogalaxy the interstellar medium, sitting in the bottom of the potential well, had more difficulty to escape than in the outer parts. So, although star formation would heat up the gas, the value of  $\alpha$  in the above illustrative model was much lower in the central area's than in the extremities. After depletion of the gas, the mean stellar metallicity will be equal to the effective yield, as we have just seen. This process thus sets up an abundance gradient as a result of gas loss.

In comparison to the Galaxy, we should note that the observed gradients are currently limited to the brighter parts of the bulges. Although such a gradient may be present for the inner bulge in the Galaxy as well, it has been suggested by Searle and Zinn (1978) that there is no such gradient in the outer parts of the globular cluster system (roughly beyond the solar radius). They suggested that the loosely bound outer clusters have a broader range of ages and resulted from continuing infall of fragments after completion of the inner bulge formation. Likewise, these fragments lost gas and the mean abundances of the stars in the resulting globular clusters remained low with no radial gradient in their present space distribution. The discussion of the color gradients above and their possible origin then only applies to the inner parts of halo population II.

Observations of elliptical galaxies also have revealed a correlation between the metallicity and the luminosity. If the latter corresponds to mass, this relation can also be explained qualitatively by the simple description. After all, in small systems the potential well is less deep and the gas more prone to evaporation from the system. Such systems will therefore have lower effective yields and therefore lower mean stellar abundances after completion of gas consumption by star formation. The correlation is indeed in the sense expected in this picture.

Now let us return to the disks. In their study of M81, Garnett and Shields have also estimated the gas fraction in the disk as a function of radius. The resulting variation of the O/H ratio from their HII-region spectra with this fraction is shown in fig. 14.1. Similar figures have been produced also for other galaxies by other workers. The full-drawn line is that for the simple model with a constant yield with radius. In spite of the fact that molecular gas was not included, the figure shows reasonable qualitative agreement. The other curves are more complicated models. Although this should not be taken as confirmation of the simple model, it follows that general considerations of chemical evolution (the extended simple model) can explain the basic trends observed.

There also is a correlation between the mean gaseous abundance (e.g. measured at the effective radius) and the galaxy's integrated magnitude. This ranges by two orders of magnitude between small systems like the LMC and the largest spirals with the latter having the largest abundances. To some extent this is also a trend in morphological type, since late-type galaxies tend to be less luminous. Also the gas fraction decreases toward earlier galaxies and this trend therefore also fits qualitatively with the predictions of the extended simple model.

There is some uncertainty concerning the abundances of nitrogen. In most galaxies there is little radial trend in the N/O ratio, but sometimes in the N/S ratio. The latter also seems to correlate with galaxy luminosity. Nitrogen may at least in part be a secondary element, which means that it can only be produced in nucleosynthesis from other ("seed") elements and therefore its production in a particular generation of stars depends on the abundance in primary elements in the gas out of which these stars form. The simple model

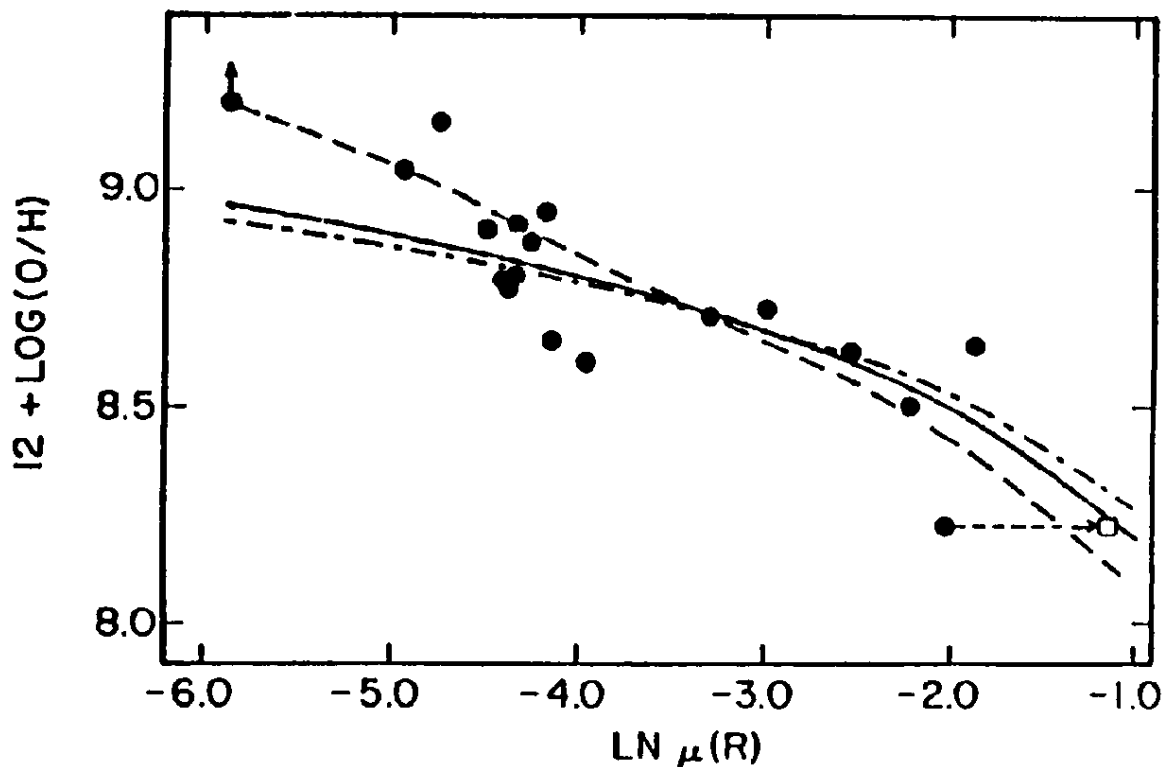


Figure 14.1: The correlation between O/H abundance ratio and the gas fraction in the disk of M81. The data points are from observed HII regions. The full line is a fit with the simple model of chemical evolution; the other two are more complicated models. From Garnett and Shields (1987).

then predicts that its abundance is roughly proportional to the square of the abundance of primary elements. The evidence on this point is not yet conclusive.

In his discussion of the K-giants used to determine velocity dispersions in the old disk of our Galaxy, Lewis (1986, see also Lewis and Freeman, 1989) found that these stars show essentially no radial gradient in their abundance  $[\text{Fe}/\text{H}]$ . This is in contrast to the observed abundance gradient in the gas. This feature, however, also follows from the models given above. In the extended simple model we see that the dependence of  $\langle Z \rangle$  on  $x$  is rather different from that of  $Z(t)$  in the gas (compare eqs. (14.3) and (14.5)). For  $Z_0 \neq 0$ , the mean stellar abundance is very close to this initial abundance, especially if we remember that we should take the relevant value for  $x$  at the end of formation of this population, which is then not much less than unity. The resulting small dependence upon

$x$  might then be the reason for the absence of such a gradient. For the gas we need to take  $x$  at the present time and  $Z(t)$  becomes  $Z_0$  plus a few times the yield and has a strong dependence on  $x$ .

### 14.3 Formation of spiral galaxies

I will now discuss a simple scenario for the formation of disk galaxies with a view towards an understanding of the various observations that I have discussed for external systems. Before doing this I will first describe some general notions, that I will then use to describe a possible scenario.

In the first place I recall the basic two-component structure of disk galaxies in their light distributions (van der Kruit and Searle, 1982). The possible “thick disks” in our and other galaxies, which have a small total luminosity compared to that of the disk, are a minor component and their existence would not contradict this basic feature of the stellar distributions. In NGC 7814, van der Kruit and Searle did observe a color gradient in the spheroid and noted that the isochromes and the isophotes have similar flattening. Now models have been presented, where bulge formation proceeds by a dissipational process, in which stars form out of a relatively slowly contracting gas sphere, which is continually enriched (e.g. Larson, 1976). It seems a general property of such models, especially when rotation is included, that the isochromes are considerably flatter than the isophotes, while the inner ones are flatter than those further out. This is so because on this hypothesis a galaxy is a superposition of a chronological sequence of stellar populations in which flattening and metal abundance both increase with time. The observed color gradients suggest that the binding energy of the stars is correlated with metallicity. These observations indicate that the bulge stars formed before the galaxy collapsed and that the abundance structure existed in the non-equilibrium protogalaxy. I have discussed above how such a gradient could have originated. Disks clearly form in a dissipational process, so that the basic two-component structure suggest two different epochs of star formation, one occurring before and the other after virialisation of the bulge and the collapse of the disk.

Van Albada (1982) has performed n-body simulations of the dissipationless collapse process of galaxy formation. His code was able to describe the process of violent relaxation (relaxation of stellar orbits resulting from strong variations of the gravitational field) sufficiently well for the results to be indicative of the evolution involved and the behavior of a protogalaxy under dissipationless collapse. He found two surprising things. In the first place, he found that the resulting distributions in collapsed systems with a variety of irregular initial conditions could be fit excellently to the  $R^{1/4}$ -law over a region between radii containing 10 and 99% of the total mass, provided that the collapse factor is large.

The range of surface brightness represented was 12 mag. A second important finding was that the binding energy of the particles before and after collapse was correlated. This means that any structure in the protogalaxy, such as an abundance gradient would survive statistically in the process of dissipationless collapse and violent relaxation. This study thus indicates that bulges with their observed luminosity distributions and abundance gradients can indeed be produced without invoking dissipational processes.

There is some dispute about the assumption that bulges and population II are the first components to collapse, mainly based on the fact that apparently there are many stars in the inner bulge of the Galaxy (within 0.5 kpc or so), that are considerably younger than the globular clusters of the outer halo (Harmon and Gilmore, 1988 and references therein). It should however be kept in mind that the inner bulge probably consists of more than just the oldest halo population. At small  $R$  (say smaller than 0.5 kpc) there is no longer a clear-cut distinction between the population II bulge stars and those of the old disk (including the “thick disk”), as such stars will have similar kinematics (velocity dispersions of order  $100 \text{ km s}^{-1}$  and little rotation) and all will be relatively metal-rich. But also stars that formed in the inner disk after completion of disk formation over a period of several Gyrs (until the gas supply was exhausted) will have large velocity dispersions due to the expected strong effects of secular evolution of their orbits. It would not be surprising therefore, if at distances less than 0.5 kpc or so from the plane in the inner bulge there is a mix of stars from population II, old disk population as well as somewhat younger disk generations, that cannot be distinguished by their kinematics or metallicities. The younger stars in the inner bulge should then be identified with those that formed over the first few Gyrs after disk formation and this fits with their observed vertical scaleheight of about 370 pc (Harmon and Gilmore, 1988).

Disk galaxies rotate. Peebles (1969) has studied the possibility that protogalaxies acquire angular momentum from tidal torques from their neighbors in the early universe. He found that the amount of angular momentum predicted by this mechanism could be described by a single dimensionless parameter

$$\lambda = J |E|^{1/2} G^{-1} M^{-5/2}, \quad (14.12)$$

where  $J$  is the total angular momentum,  $E$  the total energy and  $M$  the total mass. Peebles estimated that  $\lambda$  would be about 0.08, while numerical experiments have shown it to be about 0.07 with a standard deviation of about 0.03. A problem at that time was that the estimated amount of angular momentum that results from this mechanism was far too little to explain the observed rotation.

This problem was addressed by Fall and Efstathiou (1980). They included the effects of a dark halo and assumed that the luminous part of a galaxy settles in the potential well of this dark halo. They considered the case that the halo (of which the bulge may or may not be a part) and the disk have the same distribution of specific angular momentum



(angular momentum per unit mass) and that the dissipational collapse of the disk occurred with detailed conservation of angular momentum; in other words, the distribution of specific angular momentum before and after collapse is the same. This makes it possible to calculate from observations the angular momentum distribution of the protodisk and therefore of the protohalo. For this purpose they used observed density distributions (from surface photometry) and rotation curves to calculate the properties of the protogalaxy. Their specific analytical fit to flat rotation curves will be used below; it is given by

$$V_{\text{rot}}^2(R) = \frac{V_{\text{m}}^2 R^2}{R_{\text{m}}^2 + R^2} \left\{ 1 - \gamma \ln \left( \frac{R^2}{R_{\text{m}}^2 + R^2} \right) \right\}. \quad (14.13)$$

This applies up to the radius  $R_{\text{H}}$  of the halo, after which it will become Keplerian. For actual galaxies  $R_{\text{m}}$  is 0.1 to 0.5 scalelengths  $h$ . From these fits to galaxies, Fall and Efstathiou could also calculate the value of Peebles'  $\lambda$  if the radius  $R_{\text{H}}$  is assumed. The requirement that tidal torques provide the angular momentum observed then translated in the requirement that the mean collapse factors  $R_{\text{H}}/h$  are about 20 and mean halo-to-disk mass ratio's of order 10. It is important to stress that they did not presuppose any distribution parameters for the protogalaxy and started only from observations with the assumption stated.

Mestel (1963) had made an interesting inference from the then known rotation curve and mass distribution of the disk of the Galaxy. He calculated the distribution function of specific angular momentum and noted that it was rather similar to that of a sphere with a uniform density distribution and with uniform (angular) rotation speed. He then proposed that galaxies indeed collapsed with detailed conservation of angular momentum (as used later also by Fall and Efstathiou) and started from uniform, uniformly rotating spheres. This is sometimes referred to as Mestel's hypothesis. Gunn (1982) used this hypothesis to show that if such a sphere settles in the force field of a dark halo with a flat rotation curve a roughly exponential disk results. Actually, Freeman (1970) had noted that the self-gravitating exponential disk also has a specific angular momentum distribution similar to that of the Mestel sphere, but this was before the days that flat rotation curves were discovered.

I will now take the following approach in my discussion of a possible scenario of disk galaxy formation (see also van der Kruit, 1987). Assume that protogalaxies form in the early universe and acquire angular momentum from tidal torques from neighbors and that these protogalaxies can roughly be described as uniform, uniformly rotating spheres. Then let the dark matter quickly settle (dissipationlessly) in a distribution that resembles an isothermal sphere. This means that the dark halos have at large  $R$  a density distribution proportional to  $R^{-2}$  and provide a potential field that at larger radii corresponds to a flat rotation curve. Early on also some stars form, especially in the inner regions, which also settle quickly thereafter in the inner bulge. If indeed this star formation is concentrated

towards the central area's of the protogalaxy, it is mostly self-gravitating and will collapse as van Albada simulated in a density distribution that closely follows that of the  $R^{1/4}$ -law. These bulges will have a small amount of rotation, as is observed. The amount of mass that is turned into stars at this stage is controlled by an unknown process, but it is this amount that determines the Hubble type of the galaxy that is forming. It may have to do with details of the angular momentum distribution in the protogalaxy. An abundance gradient, that will naturally develop in the protogalaxy, will survive the collapse and result in the abundance structure observed in the bulges.

As Fall and Efstathiou, I will assume that the dark halo has the same angular momentum distribution as the protogalaxy. This means that at all positions in the protogalaxy there is an equal ratio of dark to luminous matter. The distribution of specific angular momentum in the Mestel sphere is given by  $M(h_s)/M$ , which is the fraction of matter with specific angular momentum  $\leq h_s$ :

$$\frac{M(h_s)}{M} = 1 - \left(1 - \frac{h_s}{h_{\max}}\right)^{3/2}. \quad (14.14)$$

The protogalactic sphere has a density  $\rho_o$  and therefore radius  $R_m = (3M/4\rho_o)^{1/3}$ . The gravitational potential energy is  $\Omega = -3GM^2/5R_m$  and the total angular momentum  $J = 2/5Mh_{\max}$ . Now assume that at the time of the start of the collapse, when the protogalaxy detaches itself from the expanding universe and reverses its expansion into contraction, its total energy is essentially gravitational. Then  $|E| = |\Omega|$ ; it cannot be much smaller: if a factor 2 smaller the protogalaxy is in virial equilibrium. Then we find from Peebles' parameter  $\lambda$  that

$$h_{\max} = \frac{5}{2} \left(\frac{5}{3}\right)^{1/2} G^{1/2} \lambda M^{1/2} R_m^{1/2}. \quad (14.15)$$

Now assume that a mass fraction  $1 - \Gamma$  is in the form of dark matter and that it settles in a roughly isothermal sphere with radius  $R_H$ . Then its gravitational energy after collapse is in good approximation

$$\Omega_H = -\frac{GM_H^2}{R_H} = -G(1 - \Gamma)^2 \frac{M^2}{R_H} \quad (14.16)$$

and according to the virial theorem its total energy

$$E_H = \frac{\Omega}{2} = -G(1 - \Gamma)^2 \frac{M^2}{2R_H}. \quad (14.17)$$

The total energy of the protohalo is under the assumptions made of equal distribution of specific angular momentum equal to  $(1 - \Gamma)$  times that of the protogalaxy or

$$E_H = -\frac{3}{5}(1 - \Gamma) G \frac{M^2}{R_m}. \quad (14.18)$$

Since the dark halo collapses without dissipation of energy these two energies need to be equal and we get

$$R_H = \frac{5}{6}(1 - \Gamma) R_m \quad (14.19)$$

and the asymptotic circular speed in the rotation curve is (this is of course not the rotation of the halo itself)

$$V_m^2 = \frac{GM}{R_H} = \frac{6}{5} \frac{G}{1 - \Gamma} \frac{M}{R_m}. \quad (14.20)$$

Now let us look at the remaining material, which is gas that will eventually settle in the disk. Its mass is  $\Gamma M$  and its specific angular momentum distribution that of the Mestel sphere. Let this settle in a flat disk under conservation of specific angular momentum in the potential field of the dark halo. For convenience I will use for this that corresponding to the Fall and Efstathiou rotation curve given above. Let us first see what to expect for an exponential disk. The lower full-drawn line in the left-hand part of fig. 14.2 is that of an exponential disk in such a rotation curve with  $\gamma = 0.1$  (this is not important) and  $R_m = 0.2h$ . The specific angular momentum is expressed in units of  $hV_m$ . Then we find the best representing curve for a Mestel sphere; this is the dashed line. The only free parameter to do this comparison is  $h_{\max}$  (which determines where the distribution approaches unity) and it can be seen that we have to choose it as about 4.5 in the units used. For smaller values it will rise too steeply and for larger values too slowly. The upper full-drawn line then shows the distribution for an exponential disk, but now with a cut-off at 4.5 scalelengths in order to reproduce the fact that the Mestel sphere has no  $h_s$  larger than  $h_{\max}$ .

Then turn it around and calculate what the resulting surface density distribution is for the angular momentum distribution of a Mestel sphere, where I have chosen  $h_{\max} = 4.5hV_m = 22.5V_mR_m$  following the arguments given above. This is shown in the right-hand part of fig. 14.2. The full-drawn line compares this to the expected exponential disk with  $h = 5R_m$  and a cut-off at  $4.5h$ . The comparison is very good; the vertical scale is in natural logarithms, which is close to magnitudes. Deviations of a few tenths of a magnitude are not uncommon in actual surface photometry of disks. So, the conclusion from this is, that if we let material in a Mestel sphere settle in the force field corresponding to a flat rotation curve under detailed conservation of angular momentum, we end up with a surface density distribution that is exponential and truncates at 4.5 scalelengths. This is what has actually been observed in the stellar distribution in disks of galaxies, as I discussed extensively above.

A few further remarks need to be made before proceeding to calculate further properties of this scenario. The first is that the prediction in fig. 14.2 shows a strong excess at small radii. This is material that was originally in the protogalaxy close to the rotation axis. Some of this material might actually have been used to form the bulge stars, so that

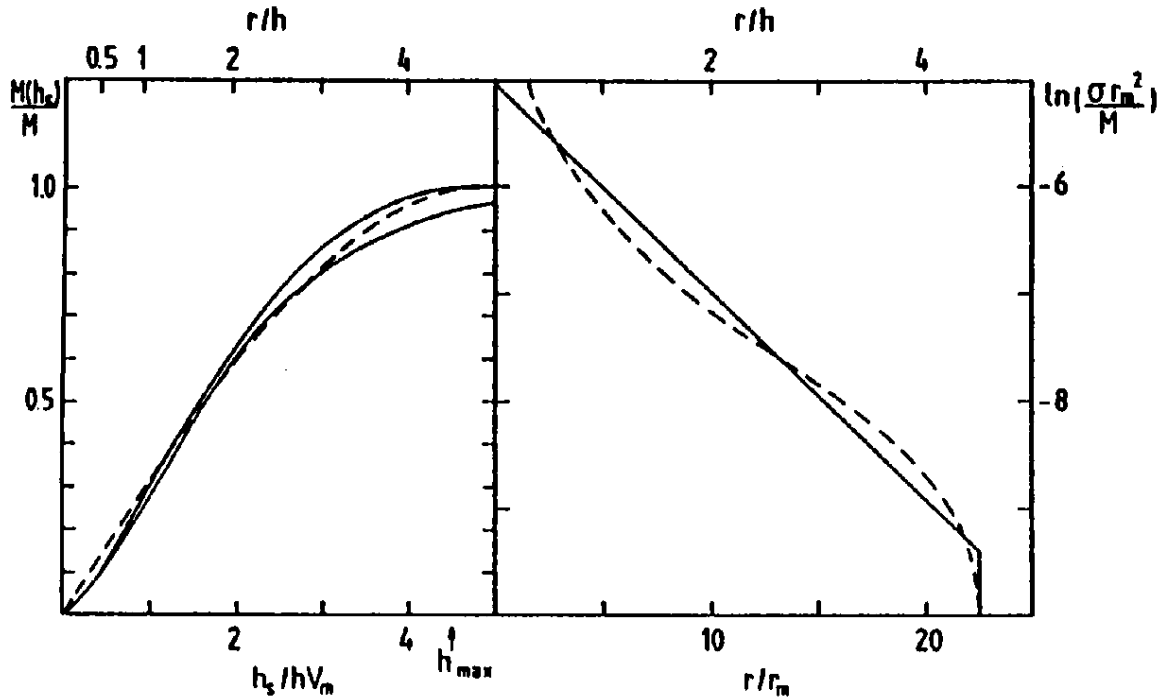


Figure 14.2: The left-hand panel shows the distribution of specific angular momentum in three cases. The lower full-drawn curve is that of an infinite exponential disk in a Fall and Efstathiou (1980) rotation curve (eq. (14.13)), using  $\gamma = 0.1$  and  $R_m = 0.2h$ . The upper full-drawn line is the same, but for an exponential disk with a sharp cut-off at  $4.5h$ . The dashed line is the distribution for a uniform, uniformly rotating sphere with maximum specific angular momentum  $h_{\max} = 4.5hV_m$ . The right-hand panel shows the surface density of a disk with the angular momentum distribution of a Mestel sphere in the same rotation curve and for the same  $h_{\max}$ , which is also equal to  $22.5V_mR_m$ . The full-drawn line is an exponential disk with scalelength  $h (= 5R_m)$  and a cut-off at  $4.5h$ . From van der Kruit (1987).

this is not a worrying difference. Also the precise form of the expected distribution depends much on the assumed rotation curve, while the disk potential itself will modify this as well. This is also of no great consequence, because—as Freeman (1970) has shown—the self-gravitating exponential disk also has the specific angular momentum distribution of the Mestel sphere. So, even if disk self-gravity seriously changes the potential field, we still expect the material to settle in an exponential disk.

The gas that settles in the disk will quickly start to form stars. The final process of disk

formation will certainly be accompanied by cloud collisions and enhanced star formation. In this initial phase there is much turbulent gas motion and many gas concentrations and we must expect that strong effects on the orbits of these stars formed at the time of disk settling, as in the Spitzer-Schwarzschild mechanism, will occur. This initial generation of disk stars will therefore be expected to have large random motions (and intermediate metallicities) and must then settle in a thicker distribution. I suggest that this is now evident in the “thick disk” population and maybe also in Zinn’s (1985) subsystem of disk globular clusters. The subsequent star formation will form the old disk population and will then naturally display the edges at 4 to 5 scalelengths.

This leaves the question of the HI gas that is observed beyond the optical edges. We can only assume that this is gas that fell in at a later stage and has been falling in from larger initial radii than the extent of the protogalaxy. There is then no requirement for it to settle in the same plane as the disk. This would explain the larger angular momentum than present in the stellar disk and also why the observed warps usually start at about the optical edges. Actually, there is no reason, why this process would not also provide gas at smaller radii, which would fit in with the observation that the HI surface density shows no feature at the optical edge.

We have thus seen that the gas with mass  $\Gamma M$  of the protogalaxy will settle dissipationally in a flat disk with a scalelength  $h = h_{\max}/\beta V_m$  and  $\beta$  about 4.5. In the following I will for simplicity ignore self-gravity, but this does not affect the conclusion concerning the formation of an exponential disk. The assumption roughly corresponds to the situation, where the dark halo everywhere dominates the rotation curve. Otherwise, we will have to assume that some unknown mechanism provides the “disk-halo conspiracy”. In any case, the resulting rotation curve will be approximately flat at the level of the asymptotic value  $V_m$  for the dark halo. With the equations above for  $h_{\max}$  and  $V_m$  we can then write

$$h = \frac{25}{6} \frac{1}{2^{1/2}} \frac{\lambda}{\beta} \frac{1}{(1-\Gamma)^{1/2}} R_m. \quad (14.21)$$

The central surface density of the disk is

$$\sigma_o = \frac{36}{625} \left(\frac{4}{3}\right)^{2/3} \frac{1}{\pi^{1/3}} \left(\frac{\beta}{\lambda}\right)^2 \frac{\Gamma}{1-\Gamma} \rho_o^{2/3} M^{1/3}. \quad (14.22)$$

The initial density  $\rho_o$  is the density of the protogalaxy when it detaches itself from the rest of the universe. If galaxies form at about the same time, this density may not vary by large factors. Indeed in the discussion of Fall (1979) on density perturbations for hierarchical clustering, the density spectrum in a critical universe  $\Delta\rho/\rho$  is proportional to a low negative power of  $M$  and therefore the dependence of  $\sigma_o$  on  $\rho_o$  and  $M$  almost disappears. So we see, that if  $\Gamma$  is constant for all galaxies, we recover Freeman’s law

of constant central surface brightness. This says that everywhere in the early universe we should have had equal ratio's of dark and luminous matter and that this still applies between present-day galaxies.

Substituting  $\beta = 4.5$  and  $\lambda = 0.07$ , we get

$$\Gamma(1 - \Gamma)^{1/2} = 1.5 \frac{\sigma_o h}{V_m^2}, \quad (14.23)$$

$$R_m = \frac{22}{(1 - \Gamma)^{1/2}} h, \quad (14.24)$$

$$R_H = 18(1 - \Gamma)h, \quad (14.25)$$

$$M = 4.2 \cdot 10^6 (1 - \Gamma)^{1/2} V_m^2 h, \quad (14.26)$$

$$\rho_o = 9.7 \cdot 10^{-8} (1 - \Gamma)^2 \frac{V_m^2}{h^2}. \quad (14.27)$$

Here length is in kpc, velocity in  $\text{km s}^{-1}$ , mass in  $M_\odot$  and density in  $M_\odot \text{ pc}^{-3}$ . Using for the Galaxy  $h = 5$  kpc,  $V_m = 220 \text{ km s}^{-1}$  and  $\sigma_o = 400 M_\odot \text{ pc}^{-2}$ , it follows that  $\Gamma = 0.06$ ,  $R_m = 115$  kpc,  $R_H = 90$  kpc,  $M = 1.0 \cdot 10^{12} M_\odot$  and  $\rho_o = 2 \cdot 10^{-4} M_\odot \text{ pc}^{-3}$ . Similar values obtain for other galaxies for which the relevant data are available, namely  $\Gamma$  in the range 0.04 to 0.11 and  $\rho_o$  a few times  $10^{-4} M_\odot \text{ pc}^{-3}$ . The collapse factor  $R_m/h$  is of order 20 and the radius of the halo is about  $18h$ . Of course,  $\lambda$  may be different for individual galaxies within the range  $0.07 \pm 0.03$ , indicated by numerical experiments and the estimates for the properties then changes accordingly. We are in a few cases approaching such radii for the observed extent of HI rotation curves. The large percentage of the mass in the form of dark material is essentially needed to explain the amount of observed rotation in disk galaxies by the tidal torque hypothesis. In the clustering model of Fall (1979) we expect  $\Delta\rho/\rho$  at about  $9\pi^2/16$ ; then it follows that  $\rho_o$  is about  $5.5\langle\rho\rangle$  and for the values above for  $\rho_o$  this occurs at a redshift  $z$  of about 3.5 in a universe with  $\Omega = 1$  and  $H = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

The luminosity of the disk is

$$L_{\text{disk}} \propto (L/M)\Gamma^2(1 - \Gamma) \frac{V_m^4}{\mu_o}. \quad (14.28)$$

So, for constant  $M/L$ ,  $\Gamma$  and central surface brightness  $\mu_o$  we get the Tully-Fisher relation.

It should be stressed that the scenario given here is very schematic and should not be construed as more than a working hypothesis. Yet it qualitatively explains the general features that we know about spiral galaxies.

## 14.4 References

- Fall, S.M. 1979, *Rev. Mod. Phys.* *51*, 21  
Fall, S.M., Efstathiou, G. 1980, *Monthly Notices Roy. Astron. Soc.* *193*, 189  
Freeman, K.C. 1970, *Astrophys. J.* *160*, 811  
Garnett, D.R., Shields, G.A. 1987, *Astrophys. J.* *317*, 82  
Gilmore, G., Wyse, R.F.G. 1986, *Nature* *322*, 806  
Gunn, J.E. 1982, *Astrophysical Cosmology*, ed. H.A. Brück, G.V. Coyne and M.S. Longair, Vatican, p. 233  
Harmon, R., Gilmore, G. 1988, *Monthly Notices Roy. Astron. Soc.* *235*, 1025  
Larson, R.B. 1976, *Monthly Notices Roy. Astron. Soc.* *176*, 31  
Larson, R.B. 1986, *Monthly Notices Roy. Astron. Soc.* *218*, 409  
Lewis, J.R. 1986, Ph.D. Thesis, Australian Natl. Univ.  
Lewis, J.R., Freeman, K.C. 1989, *Astron. J.* *97*, 139  
Mestel, L. 1963, *Monthly Notices Roy. Astron. Soc.* *126*, 553  
Pagel, B.E.J., Edmunds, M.G., Smith, G. 1980, *Monthly Notices Roy. Astron. Soc.* *193*, 219  
Peebles, P.J.E. 1969, *Astron. Astrophys.* *11*, 377  
Searle, L. 1971, *Astrophys. J.* *168*, 327  
Searle, L., Sargent, W.L.W. 1972, *Astrophys. J.* *173*, 25  
Searle, L., Zinn, R. 1978, *Astrophys. J.* *225*, 357  
van Albada, T.S. 1982, *Monthly Notices Roy. Astron. Soc.* *201*, 939  
van der Kruit, P.C. 1987, *Astron. Astrophys.* *173*, 59  
van der Kruit, P.C., Searle, L. 1982, *Astron. Astrophys.* *110*, 79  
Zinn, R. 1985, *Astrophys. J.* *293*, 424

