5 Spiral galaxies

In many ways the Mily Way galaxy is a proto-typical spiral galaxy. Our view of the MW is extremly close up. This is advantageous when trying to resolve the distinct structural, kinematic and chemical components of the Galaxy.

However, our interior view quite literally prevents us from seeing the complete picture, i.e. a panoramic, exterior view of a galaxy. In this sense, observations of Andromeda (M31) are important as we can resolve brighter stellar populations and relate them to the structure of the entire galaxy.

Studies of both the MW and M31 are further limited by the fact that they are individual galaxies. Only by studying the bulk properties of large numbers of spiral galaxies can we understand the physics behind their population distributions, e.g. the LF, mass-metallicity relations and the Tully-Fisher relation.

5.1 The structure of a spiral galaxy

A spiral galaxy may be broadly classified as containing a disk, a bulge and a halo of stars.

- The disk, as in the MW, may have multiple components, in this case a thick and thin disk component.

- The bulge of the MW displays $V/\sigma \sim 1$ and is therefore rotationally flattened as opposed to triaxial. However, one might ask what fraction of the bulge is really a bar-like feature that we are viewing edge on?

- The halo—the stars that is and not the dark matter component—is composed of globular clusters and individual stars.

5.2 Vertical structure in the MW disk

We can parameterise the vertical struture of the MW disk by extending our description of the surface brightness (density) distribution as an exponential function to a third dimension, i.e.

$$\rho(R,z) = \rho_0 \exp(-R/a) \exp(-|z|/h), \tag{1}$$

where $h \sim a/10$ is a scale height whose value relative to the disk scale length indicates that disks are relatively thin, flattened structures. With reference to our previous surface brightness model, we can write that $I_0 = 2\,h\,\rho_0$.

The vertical structure of the MW disk may be investigated by considering the number-magnitude, $N(m)$, distribution of stars of fixed spectral type, i.e. colour. By considering a narrow range of spectral types, e.g. K type stars, we are effectively considering a narrow range of stellar absolute magnitudes (note that K-dwarfs are much more numerous than K-giants). The number of stars as a function of apparent magnitude is then a function of distance (which also corresponds to a volume) and the stellar density at that distance.

$$N(m) = \int \phi(m = M + 5\log[D/10\mathrm{pc}] + A[l, b, D])\, \rho(D)\Omega D^2 \mathrm{d}D, \tag{2}$$

where ϕ is local stellar LF of that spectral type, D is the distance, A is the absorption due to dust, ρ is the stellar density and $\Omega D^2 \mathrm{d}D$ is the volume element.

The results of these studies indicate that the MW disk is stratified by age, which is the same as saying that it is stratified by age:

- $h \sim 300 - 350$pc for K dwarfs.
- $h < 200$pc for A dwarfs.
- $h < 150$pc for [HI] gas.
- $h < 60 - 70$pc for giant molecular clouds (GMCs).

Older stars have larger scale heights for two reasons:

1. The disk is "lumpy" with the lumps being GMCs of individual masses up to $10^7 \mathrm{M}_\odot$. The cumulative effect of multiple star-GMC gravitational encounters is to kinematically excite the disk. Older stars experience more encounters and larger values of KE_z will result in a larger scale height.

2. A second effect, particularly for G and K dwarfs is that the MW disk does not appear to be a single structure.

   At about 600 pc, or two scale heights, above the MW disk, the distribution of GK stars shows an excess above the assumed single exponential distribution.

   From 600 pc to 2 kpc the distribution of GK stars follows an exponential distribution of scale height $h = 1$ kpc. This is the so-called "thick disk", with stars of scale height $h < 500$ pc defining the "thin disk."

   The thick disk has no continuing star formation and contains no stars of type OBA. Taken together, this indicates that the thick disk must be older than approximately 3 Gyr.

5.3 The Galactic halo

Beyond 2 kpc above the disk one observes a further excess of GK stars in excess of the thick disk contribution. This is the Galactic halo, a distribution of old, metal poor stars that continues to large Galactocentric radius.

Note that the absolute magnitude of a main sequence K-type star is $M_V(\mathrm{KV}) = +6$. Using the distance modulus, $m = M + 5\log(d/10\mathrm{pc}$, one notes that $m_V = 21(26)$ at $d = 10$ (100) kpc, excluding any absorption. Note also that the depth of a photographic plate (used in wide field astronomy up until 1990) is of the order of 19-20th magnitude in comprison.

At the mid-plane of the Galactic disk one calculates that of order 90% of stars belong to the thin disk, 10% belong to the thick disk and 0.1% belong to the halo. Within the extent of the thin disk approximately 30% of stars belong to the thick disk.

Note that, as thick disk stars are, on average, older than thin disk stars, they will have lower M/L ratios and thus their contribution by luminosity to Galactic structure will be lower than the numerical proportions outlined above.

Each structural component is also segregated by kinematics (σ_z) and metallicity (see Table 2.1 of Sparke and Gallagher). Quantitatively, one observes,

- thin disk: $Z_\odot/2 < Z < Z_\odot$
- thick disk: $Z_\odot/10 < Z < Z_\odot/2$
- halo: $Z_\odot/100,000 < Z < Z_\odot/10$

It is interesting to note that halo stars were first identified by Baade and others as rare, metal-poor stars in the solar neighbourhood that exhibited high proper motions and radial velocities (of order 100 km/s). These high velocity stars follow vertical orbits that take them several kpc above and below the disk (where they spend most of their time as KE + PE = constant).

5.4 Explaining the origin of the thin disk, thick disk and halo

The present day thick disk is currently interpreted as the remains of an earlier thin disk which was excited kinematically as the result of an encounter with a MW satellite galaxy. The energy of the impact was largely transferred into the increased random motion of the disk stars. It is assumed that the disk gas was able to dissipate the energy of the encounter and collapse under its own gravity to form the stars of the present day thin disk. The Galactic halo is thought to represent a distinct component associated with an earlier (the earliest) period in the history of the formation of the MW.

The evidence for the halo as an old structure includes

1. Low metallicity—the gas from which the stars formed had not been enriched significantly by previous generations of stars.
2. Extent—halo stars were formed as the MW was collapsing and retain the dissipationless kinematic signature of the of the collapsing cloud (note the following discussion of ELS versus Searle and Zinn).

5.5 ELS (1962) versus Searle and Zinn (1978)

The main assertion of Eggen, Lynden-Bell and Sandage is that the MW collapsed from a single, slowly rotating, spherical gas cloud to create a rotationally supported disk. The key observable was the existence of low angular momentum, metal-poor halo stars and globular clusters.

As the MW collapses in freefall the associated timescale is $t_{ff} = \left(\frac{3\pi}{32}\frac{1}{G\rho_0}\right)$ where ρ_0 is the original density of the cloud. Locally dense regions in the MW would have collapsed faster than the overall cloud to form stars and globular clusters which would exhibit low metallicities and preserve the original motion and scale of the collapsing could. As the cloud collapsed further, new peaks would become denser and Jeans unstable, populating the halo with local, collapsed structures in an outside-in manner.

In addition to an extended, low angular momentum halo, this model also predicts radial gradients in age and metallicity of collapsed objects in the halo. However, this model fails to explain the following aspects of the MW:

- The halo does not exhibit low angular momentum, it has *zero* angular momentum. There are equal numbers of prograde and retrograde globular cluster orbits.
- There is no radial gradient in metallicity for metal-poor globular clusters observed at $r > 8$ kpc (Searle and Zinn 1978).
- The G-dwarf problem (see later section on Chemical Evolution).

Searle and Zinn (1978) measured the metallicities of 200 RGB stars in 19 globular clusters and observed an effectively random distirbution of metallicities at $r > 8$ kpc. They explained this randomness in terms of a halo built from the stochastic accretion of dwarf galaxy-sized units and their associated globular clusters (see the main PAndAS figure). If the minor mergers that led to the creation of the MW halo were random in nature, this would naturally result in a $\bar{J} = 0$ halo.

5.6 Chemical evolution: Metal production in the MW

Star formation converts gas into stars. Metals are produced and returned to the gas phase from which new, metal enriched stars are formed. Therefore, if we can describe this process mathematically, we can describe the evolution of both the gas and the stars. Our model makes three principal assumptions:

- Closed: No gas flows into or out of the region we are considering.
- Instantaneous recycling: Stellar mass is returned to the interstellar medium (ISM) via high mass, short lifetime stars.
- Uniform composition: The chemical composition of the region we are considering is well mixed.

We begin by letting $g(t)$ be the gas mass fraction at a time t, therefore, $g(t = 0) = 1$. We further let $Z(t)$ be the metal mass fraction at a time t and $S(t)$ be the total mass that has existed in a star at any time between $t = 0$ and t.

We let α be the mass fraction locked in long lived stars and stellar remnants and $1 - \alpha$ therefore represents the fraction returned to the ISM. For Salpeter initial mass function (IMF) and normal stellar evolution theory, letting $\alpha = 0.8$ is a reasonable assumption. (You might want to consider what stellar mass is implied by this fraction?)

If we let p' be the mass fraction of metals produced and returned to the ISM, then $p = p'/\alpha$ therefore represents the mass fraction of metals locked up in stars. In this case p or p' refer to so-called primary elements that are produced directly from nuclear reactions involving hydrogen or helium, e.g. Carbon or Oxygen (via the triple alpha reaction). Under these conditions it is assumed that the metal production rate does not depend upon the current metallicity. The values of p and p' are considered to be constants and models suggest that $p \sim 0.01$. The above statements lead us to write that

$$g = 1 - \alpha S \tag{3}$$

and

$$\frac{\mathrm{d}g}{\mathrm{d}S} = -\alpha. \tag{4}$$

Two of the principal questions to be answered by a chemical evolution model concern the metallicity evolution of the ISM (the gas phase) and of the stellar popuplation itself. Considering first the gas phase metallicity (Z), we can write the following relation:

$$\begin{aligned}
\Delta\text{gas phase metallicity} &= \text{mass fraction of metals made and returned to the ISM} \\
&\quad - \text{mass fraction of metals locked in long-lived stars}
\end{aligned}$$

$$\begin{aligned}
\mathrm{d}(gZ) &= p'\mathrm{d}S - \alpha Z \mathrm{d}S \\
\frac{\mathrm{d}(gZ)}{\mathrm{d}S} &= p' - \alpha Z
\end{aligned} \tag{5}$$

We can then recast this equation as

$$\frac{\mathrm{d}(gZ)}{\mathrm{d}g} = \frac{\mathrm{d}(gZ)}{\mathrm{d}S}\frac{\mathrm{d}\dot{S}}{\mathrm{d}g} = (p' - \alpha Z)(-1/\alpha), \tag{6}$$

i.e.

$$g\frac{\mathrm{d}Z}{\mathrm{d}g} = -p. \tag{7}$$

Which has the solution

$$Z(t) = -p\ln[g(t)], \tag{8}$$

and indicates that the gas phase metallicity increases logarithmically as the gas fraction decreases.

But what of the metallicity of the stars formed from the gas? A star forming at a time t from gas of metallicity $Z(t)$ will also have a metallicity $Z(t)$. As Z increases monotonically as g decreases, then all stars forming at $t < t_0$ will have $Z_* < Z(t_0)$ and we can write

$$s(< Z_*) = 1 - g(t) = 1 - \exp[-Z_*/p]. \tag{9}$$

This relation predicts the existence of large numbers of low metallicity stars. The number of stars with any given metallicity can be written as

$$n_*(Z_*) \propto \frac{\mathrm{d}s(< Z_*)}{\mathrm{d}Z_*} \propto \exp[-Z_*/p]. \tag{10}$$

In addition, the mass-weighted mean stellar metallicity will be

$$\langle Z_*(t) \rangle = \frac{1}{s(t)} \int Z \mathrm{d}s = \frac{1}{1 - g(t)} \int \frac{Z}{p} \exp[-Z/p] \mathrm{d}Z, \tag{11}$$

where $\mathrm{d}s = -\mathrm{d}g$ and $g\frac{\mathrm{d}Z}{p} = \exp[-Z/p]\mathrm{d}Z/p$. Integrating this expression by parts and taking the limits $0 < Z_* < Z(t) = -p\ln[g(t)]$, one obtains

$$\langle Z_* \rangle = \frac{1 - g + g\ln g}{1 - g} \times p, \tag{12}$$

which tends to p as g tends to zero.

When we observe the metallicity distribution of G and K type giants towards the Galactic bulge (Baade's window), we note a good fit to $s(Z_*)$ for $p \sim Z_\odot$, i.e. 0.02. However, the simple, closed box model fails to reproduce a number of observations:

1. Globular clusters contain no gas yet all of the stars have the same metal-poor composition. The gas in GCs must have been partially enriched, formed a single generation of stars, and then expelled any remaining gas. A similar situation is encountered in dwarf galaxies. We assume that a burst of SNe type II imparted sufficient kinetic energy into the gas to expel it from the potential of the GC or dwarf galaxy. We can parameterize this by setting $p = 0$ as the effective yield in the GC or dwarf galaxy.

2. Stars of fixed age display varying metal content and therefore indicate the effects of inhomogenous mixing.

3. Not all metals are produced and returned to the ISM simultaneously. SNe type II return elements with $A < 30$ to the ISM while those with $A > 30$ are locked in the resulting remnant. SNe type Ia release abundant iron and iron-mass elements and their isotopes. The timescales for these two types of SNe are different, with $\tau(\mathrm{SNII}) < 100$ Myr and $\tau(\mathrm{SNeIa}) > 1$ Gyr. As a result of this the ratio of element abundances such as [O/Fe] varies with time.

4. The G-dwarf problem. Although the closed box model provides an acceptable description of the MW bulge, it fails to describe the disk.

   The gas fraction in the solar neighbourhood is $g(t_0) \sim 0.25$. Taking $Z(t_0) \sim Z_\odot$, we have

   $$Z_\odot \approx p \ln 4 \Rightarrow p \sim 0.7\ Z_\odot = 0.015. \tag{13}$$

   The mass fraction of low metallicity stars in the solar neighbourhood, e.g. $Z < Z_\odot/4$, should be

   $$s(Z < Z_\odot/4) = 1 - \exp[-Z_\odot/4p] = 0.4, \tag{14}$$

   whereas some 25% of local G-dwarfs show $Z < Z_\odot/4$. Pushing further, one expects $\sim$ 20% of stars to show $Z < Z_\odot/10$, whereas almost no solar neighbourhood stars display this metallicity (van den Bergh 1962).

5.7 Solutions to the G-dwarf problem

Solutions to the G-dwarf problem require modifications to our chemical evolution model.

1. Pre-enrichment. We consider the effect of a prior generation of zero metallicity, population III stars that raise the initial metallicity of our model from $Z(t = 0) = 0$ to $Z(t = 0) = Z_0$. We then note that

   $$Z(t) = Z_0 - p \ln g \tag{15}$$

   and

   $$s(Z < Z_*) = 1 - \exp[Z_0 - Z/p]. \tag{16}$$

   Setting $Z_0 = 0.15\ Z_\odot$ and $Z(t_0) = Z_\odot$ generates $p = 0.63\ Z_\odot$ and $s(Z < Z_\odot/4) \approx 0.2$. Problems with this idea include the non-observation of any population III stars and the challenge of achieving uniform mixing of these pre-enriched metals.

2. Inflows of low- or zero-metallicity gas. The inflow of low metallicity gas into our model has the effect of diluting the metallicity of the extant gas and prevents the metal abundance of new stars from rising too rapidly. However, overall the average metal abundance still increases with time.

The ELS collapse model was originally viewed as a closed box model. However, if we imagine $\rho(r)$ such that ρ increases towards the galactic centre, then this would appear to create a gradient in the free fall time as a function of radius with $t_{ff}(r_{inner}) < t_{ff}(r_{outer})$. The galaxy will collapse "inside-out" with lower metallicity outer gas falling onto the inner galaxy at later times. The explanation of Searle and Zinn also incorporates effective infall if one assumes that accreted dwarf galaxies are gas rich and metal poor. A "real" galaxy is probably a mixture of effects incorporating a closed bulge and a disk characterised by early inflow followed by late, closed evolution.

5.8 Does the G-dwarf problem really exist?

Sidney van den Bergh's (1962) original formulation of the G-dwarf problem was based upon stars in the solar neighbourhood. As we have seen, this population is dominated by the thin disk (70%) and thick disk (30%) stars. The G-dwarf problem indicates that the thin disk did not evolve as a closed box but must have incorporated low metallicity gas perhaps coming from an inside-out ELS-type collapse or from the accretion of gas-rich dwarfs à la Searle and Zinn.

What if we instead consider the stellar metallicity distribution of the Galaxy as a whole? Taking value from Reid and Majewski (1993) we can write

- Thin disk (exponential): $z_{thin} = 325$ pc, $\rho_{thin} = 1$.
- Thick disk (exponential): $z_{thick} = 1200$ pc, $\rho_{thick} = 0.02$.
- Halo (de Vaucouleurs): $b/a = 0.85$, $R_e = 2700$ pc, $\rho_{halo} = 0.0015 \times 2000$.

The full equation for the Galaxy is then

$$\rho_*(z) = \rho_{thin} \exp[-|z|/z_{thin}] + \rho_{thick} \exp[-|z|/z_{thick}] + \rho_{halo} \exp[-7.667(|z|/R_e)^{1/4} - 1] \tag{17}$$

and $N_*(D)\mathrm{d}D = \rho_*(D)\Omega D^2 \mathrm{d}D$. If we compare the total number of stars associated with each Galactic component and label, in an approximate manner, respectively halo, thick disk and thin disk stars as low, medium and high metallicity, then one notes that the inferred metallicity distribution is not unlike the exponential distribution predicted by the closed box model.

One might therefore conclude that the total metallicity evolution of the MW could be described by an open box model, i.e. one with three, linked compartments where gas is permitted to flow within the MW in order to match the thin disk metallicity distribution.