Galaxies and their properties

Asaf Pe'er[1]

April 3, 2017

This part of the course is based on Refs. [1] - [3].

1. Observational facts

We shell begin the discussion about galaxies by looking at several key observational facts. Almost all the information we have about astronomical objects originate from the radiation we see from them. A radiation from a source is characterized by its spectral energy distribution (SED), $f_\lambda d\lambda$, which is the total energy of emitted photons in the wavelength range $\lambda..\lambda+d\lambda$. While in principle detectors exist at all wavelengths - from radio to γ-rays, in practice the earth's atmosphere is opaque to most of the spectrum. The earth is transparent only at the optical, near-IR and radio bands. This is why observations in the far-IR, UV, X- and γ rays are done from space.

The image of an astronomical object reflects its surface brightness distribution. The surface brightness is the energy (in photons) received by a unit area, per unit time from a unit solid angle. The units of the surface brightness I is $[I] = \text{erg s}^{-1}\ \text{cm}^{-2}\ \text{sr}^{-1}$. Integrating the surface brightness over the entire image reveals the the flux of the object, f, which have units $[f] = \text{erg cm}^{-2}\ \text{s}^{-1}$.

Integrating the flux over a sphere centered at the object with radius r which is the distance to it, gives the **bolometric luminosity**,

$$L = 4\pi r^2 f \text{ erg s}^{-1}. \tag{1}$$

As an example, the bolometric luminosity of the sun is $L\odot = 3.85 \times 10^{33}$ erg s^{-1}.

Magnitude and Color. For historical reasons, the flux of an astronomical object in the optical band is usually quoted in terms of apparent magnitude,

$$m_X = -2.5 \log(f_X/f_{X,0}), \tag{2}$$

where the zero-point flux, $f_{X,0}$ has traditionally been taken as the flux in the X band of the bright star Vega. In recent years, it has become more common to use 'AB-magnitude',

$$f_{X,0} = 3.63 \times 10^{-2} \text{ erg s}^{-1}\ \text{cm}^{-2}\ \text{Hz}^{-1} \int F_X(c/\nu)d\nu, \tag{3}$$

[1]Physics Dep., University College Cork

where $F_X(\lambda)$ is the transmission of the filter that defines the waveband.

Thus, if two stars differ in magnitude by 1 ($m_2 - m_1 = 1$), then star 1 appears 2.5 brighter than star 2.

Metallicities. The primary visible constituent of most galaxies is the combined light from their stellar population. Since nearly 100% of the elements heavier then He are produced inside the stars, it is common to denote the fraction in mass of elements heavier than Helium as the metallicity, which is denoted by Z. For our sun, $Z \approx 0.02$. When specifying the fraction of a particular element (e.g., oxygen) in a star, it is common to give its abundance relative to that in the sun:

$$[A/B] = \log\left[\frac{(n_A/n_B)_\star}{(n_A/n_B)_\odot}\right], \tag{4}$$

where $(n_A/n_B)_\star$ is the number density ratio between elements A and B in the star, and $(n_A/n_B)_\odot$ is the corresponding ratio for the sun.

1.1. Galaxies

Galaxies are the building blocks of the universe. Interestingly, they were recognized as objects outside of the Milky Way only in the 1920's, when Hubble identified Cepheid variable stars in Andromeda.

1.1.1. Classification of galaxies

Based on basic observational features, Hubble ordered galaxies in morphological sequence, known as **Hubble sequence**. Hubble's scheme classifies galaxies into 4 broad classes (see Figure 1):

1. **Elliptical galaxies**. Those have smooth, almost elliptical isophotes. They are divided into subtypes, $E0, E1, ...E7$, where the integer is the one closest to $10(1 - b/a)$, with a and b the lengths of the semi-major and semi-minor axes.

2. **Spiral galaxies** (see figure 2). These have disks with spiral arm structure. They are divided into two branches: **barred spirals** and **normal spirals**, according to whether or not a bar -like center is observed at the central part. On each branch there are three classes, a, b and c according to:

- The fraction of light in the central bulge.
- The tightness of the spiral arms.
- The degree into which the spiral arms are resolved into stars, HII regions and ordered dust lanes

These criteria are correlated. Pronounced bulge component usually correlates with with tightly wound spiral arms and relatively faint HII regions. These are classified Sa, etc.

3. **Lenticular** or $S0$ galaxies. This is intermediate class between ellipticals and spirals. They have smooth light distribution with no spiral arms (like ellipticals) and also thin disk and bulge (like spirals).

4. **Irregular galaxies.** These appear as patchy objects, dominated by few HII regions (= regions dominated by interstellar atomic, fully ionized hydrogen, i.e., H^+).

Elliptical and lenticulars together are often referred to as **early type galaxies**, while spirals and irregulars are **late type galaxies**. Note though that the origin of this classification is historical, more then physical.

Similarly, **dwarf galaxies** are really *faint galaxies* with $M_B \gtrsim -18$.

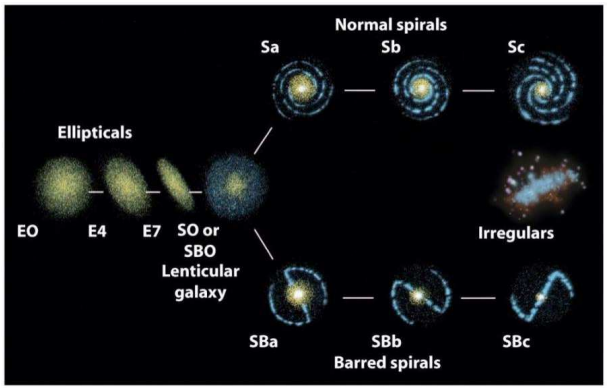

Fig. 1.— Hubble sequence of galaxy morphologies.

1.2. Elliptical galaxies

These are characterized by smooth, elliptical surface brightness distribution, contain little cold gas or dust, and have *red* photometric colors, which is a characteristic of old

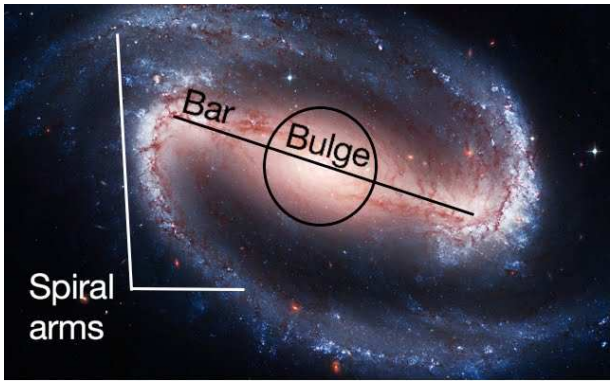

Fig. 2.— Basic morphological ingredients of a spiral galaxy.

stellar population.

Surface brightness profile. The 1-d surface brightness profile, $I(R)$ of an elliptical galaxy is usually defined as the surface brightness as a function of the isophotal semi-major axis length R.

The surface brightness profile of a spheroidal galaxy is generally well fit be the **Sersic profile**,

$$I(R) = I_0 \exp\left[-\beta_n \left(\frac{R}{R_e}\right)^{1/n}\right] \tag{5}$$

where I_0 is the central surface brightness, $n \approx 4$ is the Sersic index, $R_e \sim 1$ kpc is the effective radius that encloses half of the total light and $\beta_n \approx 2n - 0.324$.

Since $n = 4$ (also known as de Vaucouleurs profile) provides good fits to many elliptical galaxies, it is also called $R^{1/4}$ profile.

Colors. Elliptical galaxies in general have red colors, indicating that their stellar contents are dominated by old, metal rich stars. In general, the outskirts are bluer than central regimes.

kinematic properties. Giant ellipticals generally have low rotation velocities. A measure of the relative importance of the ordered and random motions within a galaxy can be set by comparing the maximum line of sight steaming motion, v_m to $\bar{\sigma}$, the average value of the line of sight velocity dispersion (interior to $R_e/2$). For bright ellipticals, this ratio is lower then the predicted for isotropic, oblate galaxy that was flattened by the centrifugal force generated by the rotation. This indicates that the flattening is due to velocity anisotropy, rather then rotation.

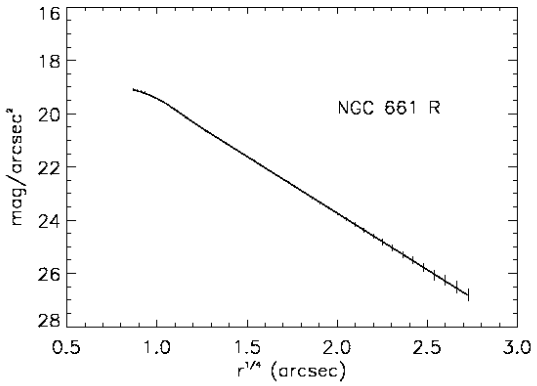

Figure 3.2: NGC 661 is a good example of de Vaucouleurs' law profile. The profile is a very good approximation of a straight line in the $r^{1/4} - \mu$ plane. The curvature at small r is mostly due to the effect of the PSF.

Fig. 3.— The elliptical galaxy NGC 661 is a good example of de Vaucouleurs $r^{1/4}$ law profile.

Scaling relations. The kinematics and photometric properties of elliptical galaxies are correlated. Elliptical galaxies with a larger central velocity dispersion are brighter; this is known as **Faber-Jackson relation**, and can be written as $L \propto \sigma^{\gamma}$ with $\gamma \simeq 4$.

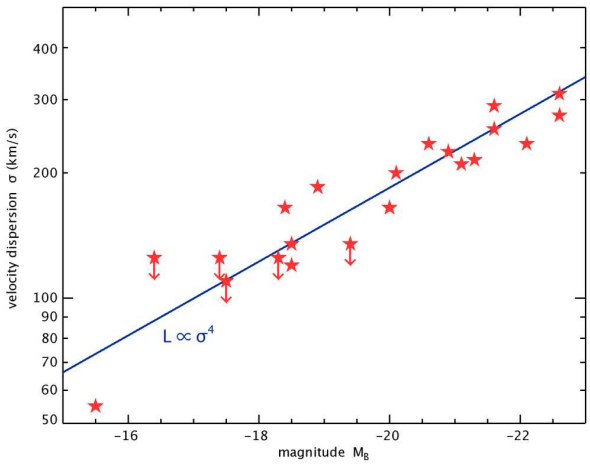

Fig. 4.— The luminosity - velocity dispersion (Faber-Jackson) relation in elliptical galaxies.

Gas content. In luminous ellipticals, the ISM is usually dominated by hot ($\sim 10^7$ K) x-ray emitting gas, which can contribute up to 10^{10} $M_o dot$ to the total mass of the galaxy.

1.3. Disk galaxies

Disk galaxies have far more complex morphologies than elliptical ones. They typically consist of thin, rotationally supported disk with spiral arms and often a bar, plus a central bulge component.

Surface brightness profile. Many disk galaxies are fitted with a Sersic $n = 1$ profile, namely

$$I(R) = I_0 \exp(-R/R_d), \tag{6}$$

where $I_0 = L/(2\pi R_d^2)$ is the central luminosity surface density, R_d is scale length and L is the total luminosity.

colors. In general, disk galaxies are bluer than elliptical galaxies of the same luminosity. This is mainly due to the fact that disk galaxies are still actively forming stars, and young stellar populations are blue. Similar to elliptical galaxies, more luminous disks are generally redder.

stellar halos. The Milky Way contains a halo of old, metal-poor stars with a density distribution that falls off as a power law, $\rho \propto r^{-\alpha}$ and $\alpha \sim 3$. Detection of stellar halos in other galaxies is extremely difficult, but was done in recent years.

These halos form as a result of a stellar stream associated with material that has been tidally stripped from satellite galaxies and globular clusters.

Bars and spiral arms. More than 50% of the spiral galaxies show bar-like structure in their inner regions, independent on their type.

Gas content. Unlike elliptical galaxies, the gas in spiral galaxies is mainly neutral hydrogen (HI), and molecular hydrogen (H_2).

kinematics. The stars and cold gas in galaxy disks move in the disk plane on roughly circular orbits. Therefore, the kinematics of a disk are largely specified by its rotation curve, $V_{\rm rot}(R)$, which expresses the rotation velocity as a function of galactocentric distance.

For massive galaxies, the rotation curves typically rise rapidly at small radii, and then are almost constant over most of the disk (see Figure 5).

The rotation curve is a direct measure of the gravitational force within a disk. Assuming, for simplicity, spherical symmetry, the total enclosed mass within radius r can be estimated from

$$M(r) = \frac{rV_{r,rot}^2(r)}{G} \tag{7}$$

In the outer region, where $V_{\rm rot}(r)$ is roughly constant, this implies $M(r) \propto r$. The enclosed

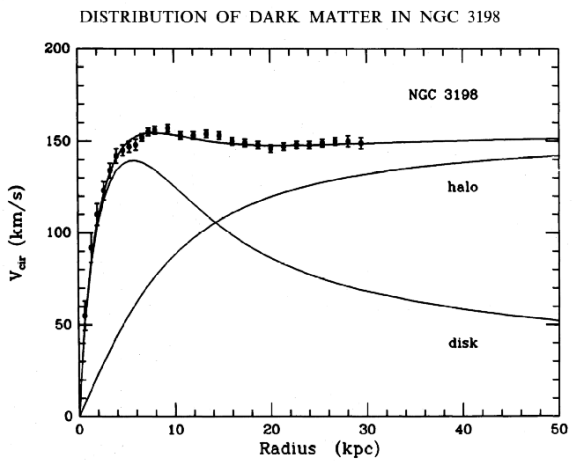

Fig. 5.— The rotation curve of the Sc galaxy NGC 3198.

mass in a galaxy (**unlike the luminosity!**) does not appear to converge. For NGC 3198 shown in Figure 5, the last data point corresponds to an enclosed mass of $\sim 1.5 \times 10^{11} M_{\odot}$, which is about a factor four larger than the inferred stellar mass. This is a strong, though indirect evidence for the existence of a dark matter halo. (Though note an alternative - modified Newtonian dynamics, or MOND theory, suggests that Newton's laws of motion need to be modified in these scales).

Tully-Fisher relation. Although spiral galaxies show great diversity in luminosity, size, rotation velocity and rotation-curve shape, they obey a well defined scaling relation between the luminosity, L and the rotation velocity (which is usually taken as $V_{\rm max}$, the maximum of the rotation curve well away from the center). This is known as the Tully-Fisher relation, and is shown in Figure 6.

The Tully-Fisher relation can be written as $L \propto V_{\rm max}^{\alpha}$, with α in between 2.5 and 4. A very crude derivation can be done by using Equation 7 and the assumption of constant light to mass ratio, resulting in $L \propto V_{\rm max}^2 r$, and assuming that all spiral galaxies have the same surface brightness, at their centers, namely $L/r^2 = const.$ This gives $L \propto V_{\rm max}^4$.

2. Formation and properties of disk galaxies

After completing a brief description of the observational properties, let us now consider more in depth the physical properties of the galaxies, beginning with the more complicated disk galaxies.

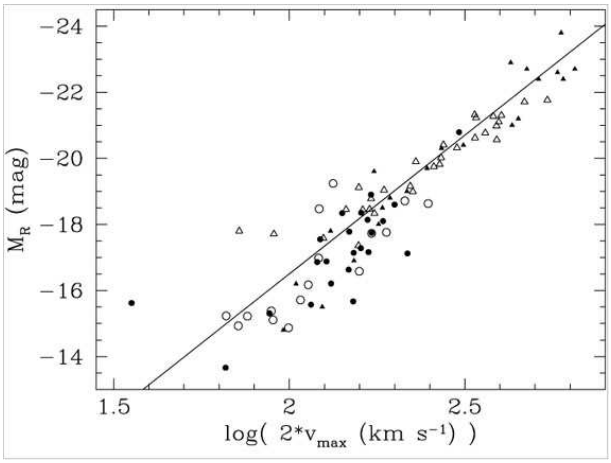

Fig. 6.— The luminosity - maximum rotation curve velocity (Tully-Fisher) relation in spiral (disk) galaxies.

2.1. The disk angular momentum

One of the most important properties of spiral galaxies is that their disks are supported by rotation. In order to quantify the importance of rotation, we introduce the (dimensionless) spin parameter as follows.

A galaxy of mass M, radius R and angular momentum L will rotate with angular speed $\omega \sim L/(MR^2)$. The angular speed ω_c of a circular orbit at radius R is given by $\omega_c^2 R \sim GM/R^2$. The energy of the system is $E \sim -GM^2/R$. We therefore look at the ratio

$$\lambda = \frac{\omega}{\omega_c} = \frac{L}{MR^2} \times \frac{R^{3/2}}{\sqrt{GM}} = \frac{L|E|^{1/2}}{GM^{5/2}}. \tag{8}$$

The parameter λ is known as the **galaxy spin parameter**. The value of λ tells how far the galaxy is supported against collapse by rotation, rather than pressure or random motion of its stars. Its importance lies on the fact that all 3 parameters: L, E and M are conserved for an isolated system during dissipationless gravitational collapse, hence λ should be conserved.

Observations show that for disk galaxies like our Milky Way, $\lambda \approx 0.5$. On the other hand, gravitational N-body simulations of galaxy formation show that the expected value of λ is $0.01 \lesssim \lambda \lesssim 0.1$. This is similar to what is observed in elliptical galaxies, but not spiral ones.

This already tells us that the Milky way must have a dark halo, otherwise it its disk

would not have time to form. Consider a self gravitating gas cloud containing no dark matter. As the cloud radiates and shrinks, its mass M and angular momentum L are conserved, but its binding energy, $-E$, increases in inverse proportion to its size, R, as $-E \propto R^{-1}$. Its spin parameter should therefore scale as

$$\lambda = \lambda_i \left(\frac{R}{R_i}\right)^{-1/2}. \tag{9}$$

Increasing λ from ~ 0.05 to $\lambda \sim 0.5$ thus requires a contraction by a factor of $R_i/R \sim 100$.

Consider a galaxy like the Milky way, with mass $M \sim 5 \times 10^{10}\ M_\odot$, and $R \sim 10$ kpc. Since for the Milky Way $\lambda \approx 0.425$, the radius of the initial cloud must have been 70 times larger than its current radius, or $R_i \sim 700$ kpc. The free fall time is

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32G\rho}} \simeq 4.3 \times 10^{10}\ {\rm yr}. \tag{10}$$

This is much longer then the age of the universe, thereby ruling out the possibility that disk galaxies could have formed from pure gas clouds with spin parameters expected from basic formation theory.

This argument is not valid in the presence of massive dark matter halo. Consider a halo with circular velocity independent of radius; This gives $V^2/r \propto M/r \propto r^0 \rightarrow M \propto r \rightarrow \rho \propto r^{-2}$. Assume that the gas cools and flows inward conserving its specific angular momentum (= angular momentum per unit mass, $\vec{L}/m = \vec{r} \times \vec{v}$). Thus, the gas that is currently at radius R and rotates at velocity V_c in the disk originated from radius

$$R_i = R(V_c/V_{\rm rot,i}), \tag{11}$$

where $V_{\rm rot,i}$ is the typical initial rotation velocity of the gas at radius R_i. Since the circular velocity is roughly constant, the specific angular momentum is linear with r. For a singular, isothermal halo, the average rotation velocity of the dark matter is $V_{\rm rot}(r) = \eta V_c$, where eta is a constant that depends on λ. For $\lambda = 0.05$, $\eta \simeq 0.14$. Thus, if the gas in the DM halo have the same specific angular momentum distribution as the DM particles (a reasonable assumption), then this gas needed to collapse only by a factor of $1/\eta \approx 7$ in order to bring its rotation speed up to V_c, reaching centrifugal equilibrium in the potential well of the DM halo.

The presence of DM halo thus reduces the collapse factor of the gas by an order of magnitude; Since the free fall time in isothermal potential is $t_{\rm ff} \propto R_i^{3/2}$, the gas can settle centrifugal equilibrium in $\approx 1.4 \times 10^9$ yr. This argument can serve as independent argument for the existence of DM halos.

3. Orbits in disk galaxies

Orbits in axisymmetric potential conserve both energy and angular momentum in the direction of symmetry axis (hereby the z direction). Writing the position of a star as

$$\vec{r} = R\hat{e}_R + z\hat{e}_z \tag{12}$$

and using the conversion from Cartesian to cylindrical coordinates,

$$x = R\cos\phi \quad y = R\sin\phi \quad z = z \tag{13}$$

and

$$\hat{e}_R = \cos\phi\hat{x} + \sin\phi\hat{y} \quad \hat{e}_\phi = -\sin\phi\hat{x} + \cos\phi\hat{y} \quad \hat{e}_z = \hat{e}_z \tag{14}$$

one obtains

$$\frac{d^2\vec{r}}{dt^2} = (\ddot{R} - R\dot{\phi}^2)\hat{e}_R + \frac{1}{R}\frac{d(R^2\dot{\phi})}{dt}\hat{e}_\phi + \ddot{z}\hat{e}_z. \tag{15}$$

One can now solve Newton's equation of motion, $d^2\vec{r}/dt^2 = -\vec{\nabla}\Phi$ in cylindrical coordinates, R, ϕ, z recalling that $\partial\Phi/\partial\phi = 0$ as the potential Φ is independent on ϕ. This gives

$$\ddot{R} - R\dot{\phi}^2 = -\frac{\partial\Phi}{\partial R}; \tag{16}$$

$$\frac{d}{dt}\left(R^2\dot{\phi}\right) = 0; \tag{17}$$

and

$$\ddot{z} = -\frac{\partial\Phi}{\partial z} \tag{18}$$

Equation 17 represents conservation of the angular momentum about the z axis, as $L_z = Rv_\phi = R^2\dot{\phi}$. Equations 16 and 18 describe coupled oscillations in the R and z directions. This can be seen as we can write Equation 16 in the form

$$\ddot{R} = -\frac{\partial\Phi}{\partial R} + \frac{L_z^2}{R^3} = -\frac{\partial\Phi_{\rm eff}}{\partial R} \tag{19}$$

where

$$\Phi_{\rm eff}(R,z) = \Phi(R,z) + \frac{L_z^2}{2R^2}. \tag{20}$$

We further have

$$\ddot{z} = -\frac{\partial\Phi}{\partial z} = -\frac{\partial\Phi_{\rm eff}}{\partial z}. \tag{21}$$

The effective potential has a minimum at $(R, z) = (R_g, 0)$ called the **guiding center**, where

$$\frac{\partial \Phi_{\rm eff}}{\partial R} = \frac{\partial \Phi}{\partial R} - \frac{L_z^2}{R^3} = 0 \tag{22}$$

or

$$\left.\frac{\partial \Phi}{\partial R}\right|_{R_g} = \frac{L_z^2(R_g, 0)}{R_g^3}. \tag{23}$$

Since $L_z = R v_\phi$, where $v_\phi = V_c(R)$ the right hand side of Equation 23 is simply

$$\frac{L_z^2(R_g, 0)}{R_g^3} = \left.\frac{v_\phi^2}{R}\right|_{R_g}, \tag{24}$$

which is the centripetal acceleration for perfect circular motion. Since the left hand side of Equation 23 is the radial force (per unit mass) exerted on the star by the gravitational potential, Equation 23 obtains a minimum for a perfect circular motion for which

$$\vec{F}_{\hat{R}}(R_g) = - \left.\frac{M v_\phi^2}{R}\right|_{R_g} \tag{25}$$

For this motion, the circular frequency will be

$$\Omega(R) = \frac{V_\phi(R)}{R} = \left[\frac{1}{R}\left(\frac{\partial \Phi}{\partial R}\right)_{(R_g,0)}\right]^{1/2} = \frac{L_z}{R^2} \tag{26}$$

In order to proceed, we define $x = R - R_g$ and Taylor expand $\Phi_{\rm eff}$ around its minimum point, $(x, z) = (0, 0)$. Since the first order terms are zero (used to find the minimum point), and the mixed term ($\partial^2 \Phi_{\rm eff}/\partial R \partial z = 0$) because of the symmetry of $\Phi_{\rm eff}$ around the $z = 0$ plane, we are left with two decoupled equations of motion,

$$\ddot{x} = -\kappa^2 x; \quad \ddot{z} = -\nu^2 z \tag{27}$$

where

$$\kappa^2 \equiv \left(\frac{\partial^2 \Phi_{\rm eff}}{\partial R^2}\right)_{(R_g,0)}; \quad \nu^2 \equiv \left(\frac{\partial^2 \Phi_{\rm eff}}{\partial z^2}\right)_{(R_g,0)}. \tag{28}$$

In this approximation, the motion in the x and z directions reduce to that of simple harmonic oscillations. This is known as the **epicyclic approximation**, as it allows the motion of stars in disk like potentials to be characterized by three frequencies: Ω, κ and ν.

Using the definition of the effective potential, it is straightforward to show that the epicyclic and circular frequencies are related through

$$\kappa = \left[R\frac{d\Omega^2}{dR} + 4\Omega^2\right]^{1/2} \tag{29}$$

For realistic galactic potentials, $\Omega < \kappa < 2\Omega$, where the limits correspond to a homogeneous mass distribution ($\kappa = 2\Omega$) and the Kepler potential ($\kappa = \Omega$). Therefore, in general, the period of the radial oscillation is incommensurable with the orbital period, so that the orbit does not form a closed figure in an inertial frame. Rather, the motion in (R, ϕ) plane can be described as a retrograde motion on an ellipse, whose guiding center is a in prograde motion around the center of the system.

Oort constants. It is useful to define two functions,

$$A(R) = -\frac{1}{2} R \frac{d\Omega}{dR} \tag{30}$$

and

$$B(R) = -\left(\Omega + \frac{1}{2} R \frac{d\Omega}{dR}\right). \tag{31}$$

These functions are related to the circular and epicycle frequencies by

$$\Omega = A - B \;\; ; \;\; \kappa^2 = -4B(A - B) = -4B\Omega. \tag{32}$$

The values of A and B at the solar radius can be measured directly from the kinematics of stars in the solar neighborhood, and are called **Oort constants**. Measurements give $A = (14.8 \pm 0.8)$ km s^{-1} kpc^{-1} and $B = -(12.4 \pm 0.6)$ km s^{-1} kpc^{-1}, from which we find that $\kappa_0 = (37 \pm 3)$ km s^{-1} kpc^{-1} and $\kappa_0/\Omega_0 = 1.35 \pm 0.05$. Thus, the sun oscillates about 1.35 times in the radial direction, by the time it completes an orbit around the galactic center.

4. The formation of disk galaxies

Disk galaxies are highly flattened systems supported by rotation. Thus, they are formed through the dissipational collapse of a gas cloud with some initial angular momentum. This occurs for hot clouds, $T > 10^4$ K, as they efficiently radiate their binding energy and contract, until its energy reaches a minimum.

In the absence of interactions with other masses (e.g., the dark matter halo), the cloud's angular momentum is conserved - simply because radiation is isotropic. This leads to a rotating disk. Although the lowest energetic state that conserves angular momentum is that all but a small fraction of Δm of the mass sinks into the central black hole (BH), this does not happen, since it requires an efficiency transfer of angular momentum between the clouds's ingredients. (Maybe?) we will discuss more on that later.

4.1. Non Self-Gravitating disks in Isothermal Spheres

As a simple illustration, let us consider an idealized case in which (i) we ignore the self-gravity of the disk; and (ii) we assume that dark matter halo has the density distribution of a singular, isothermal sphere, namely $\rho(r) = 3v_{\rm vir}^2/(4\pi G r^2)$. (This follows from comparing the centrifugal force, $mv_{\rm vir}^2/r$ with the gravitational force, GMm/r^2, and using $M = (4\pi/3)\rho r^3$). Here, $v_{\rm vir}$ is the circular velocity at the virial radius, $r_{\rm vir}$. We define the virial radius as the radius in which the average density within it is $\rho = \Delta_{\rm vir}\rho_c$, where ρ_c is the critical density, $\rho_c = 3H^2(z)/(8\pi G)$ (see "Galaxy Formation").

For flat Λ-CDM cosmology, $\Delta_{\rm vir}(z=0) \simeq 100$. We can therefore write the virial mass as the virial velocity as

$$r_{\rm vir} = \sqrt{\frac{3}{4\pi G \Delta_{\rm vir}\rho_c}} v_{\rm vir} = \sqrt{\frac{2}{\Delta_{\rm vir}}} \frac{v_{\rm vir}}{H(z)} \tag{33}$$

and

$$M_{\rm vir} = \sqrt{\frac{2}{\Delta}} \frac{v_{\rm vir}^3}{GH(z)} \tag{34}$$

Note that these relations hold independently of the density profile of the DM halo; they follow directly from the definition of the virial radius.

If we assume that the mass which settles into the disk is a fixed fraction, m_d of the halo mass, then the disk mass is

$$M_d \approx 1.3 \times 10^{11}\ h^{-1}\ M_\odot\ \left(\frac{m_d}{0.05}\right)\left(\frac{v_{\rm vir}}{200\ {\rm km\ s^{-1}}}\right)^3, \tag{35}$$

for $\Delta_{\rm vir} = 100$ and $H = H_0$.

We can calculate the angular momentum of the disk as follows. We first assume that the disk is infinitesimally thin and that gravitational effect of the disk can be neglected. We define the surface density of the disk $\Sigma(r)$ is $\Sigma(r) = \int \rho(r) dh$. The total disk mass is therefore $M_d = 4\pi \int \Sigma(r) r dr = 2\pi\sigma_0 R_d^2$, if $\Sigma(r) = \Sigma_0$ is constant and R_d is the disk radius.

For a disk that has a flat rotation curve (i.e., rotates at constant circular speed $V_c(r) = v_{\rm vir}$), its angular momentum is

$$L_d = 4\pi \int_0^{R_d} V_c(r)\Sigma(r) r^2 dr = \frac{2}{3} v_{\rm vir} R_d M_d \tag{36}$$

We assume that the angular momentum of the disk is some fraction l_d of the total angular momentum of the halo, namely $L_d = l_d L_{\rm vir}$. Relating $L_{\rm vir}$ to the spin parameter λ of the

halo (Equation 8), Equations 34 and 36 give the disk radius,

$$R_d = \frac{\lambda G M_{\rm vir}^{3/2}}{2 v_{\rm vir} |E|^{1/2}} \left(\frac{l_d}{m_d} \right). \tag{37}$$

The total energy is obtained from the virial theorem by assuming all particles are in circular orbits,

$$E = \frac{M_{\rm vir} v_{\rm vir}^2}{2} \tag{38}$$

and thus,

$$\begin{aligned} R_d &= \frac{1}{\sqrt{2}} \frac{\lambda G M_{\rm vir}}{v_{\rm vir}^2} \left(\frac{l_d}{m_d} \right) = \frac{1}{\sqrt{2}} \lambda r_{\rm vir} \left(\frac{l_d}{m_d} \right) \\ &\approx 10\ h^{-1}\ {\rm kpc}\ \left(\frac{l_d}{m_d} \right) \left(\frac{\lambda}{0.05} \right) \left(\frac{v_{\rm vir}}{200 {\rm km\ s^{-1}}} \right) \end{aligned} \tag{39}$$

and

$$\Sigma_0 \approx 207\ h\ M_\odot\ pc^{-2} \left(\frac{m_d}{0.05} \right) \left(\frac{l_d}{m_d} \right)^{-2} \left(\frac{\lambda}{0.05} \right)^{-2} \left(\frac{v_{\rm vir}}{200 {\rm km\ s^{-1}}} \right) \tag{40}$$

Equations 35, 39 and 40 thus relate the disk properties to the properties of the DM halo ($v_{\rm vir}$ and λ), as well as the uncertain efficiencies, m_d and l_d.

4.2. Comparison with observations.

For the Milky Way, we have $v_{\rm rot} \simeq 220$ km s^{-1}, $M_d \simeq 5 \times 10^{10}\ M_\odot$ and $R_d \simeq 3.5$ kpc. For $h = 0.7$ and assuming $v_{\rm rot} = v_{\rm vir}$, these values are obtained for $m_d \sim 0.01$ and $\lambda \sim 0.011$. This means that only a few % of the baryons initially n the halo, ended up in the disk. What happened to the other baryons? Good question (but recall that this is a very simplified model).

Furthermore, if we can assume a constant light-to-mass ratio, $L_d \propto M_d$, then since $V_{\rm obs} \simeq v_{\rm max}$, Equation 35 retrieves the **Tully-Fisher relation**,

$$L_d \propto V_{\rm obs}^3$$

5. Disk Instabilities

One crucial question that needs to be addressed is whether the disk galaxies are stable or not. There are two types of instabilities. **Global** instabilities, can destroy the disk; if a system is globally unstable, it will evolve towards a different configuration, erasing information about its initial conditions. On the other hand **local** instabilities determine

how perturbations on scales much smaller then the disk evolve - if the disk is unstable, local perturbations will grow, while if it is stable they will decay. Since for motion of stars requires the fragmentation and collapse of gas clouds, local instabilities are likely necessary for star formation in disk galaxies.

Here, I will focus only on local instabilities, which are believed to play an important role in the creation of density waves, which are responsible for the formation of the spiral arms (see below).

5.1. Basic Equations

We follow the standard procedure, namely write the dynamical equations in a steady state and add a small perturbations to them. We then linearize the equations, and check what are the conditions for the perturbation to grow.

Let us consider a thin, self-gravitating disk (i.e., we ignore for the moment DM halo). We use fluid dynamical equations, namely we assume that the disk is gaseous. We consider the disk to be axisymmetric, thereby we work in cylindrical coordinates, R, ϕ, z where $z = 0$ corresponds to the disk plane. Neglecting the thickness of the disk, the continuity equation $\partial\rho/\partial t + \vec{\nabla}(\vec{v} \cdot \rho) = 0$ becomes

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{R}\frac{\partial}{\partial R}(\Sigma R v_R) + \frac{1}{R}\frac{\partial}{\partial \phi}(\Sigma v_\phi) = 0 \tag{41}$$

where v_R and v_ϕ are the velocity components in the R and ϕ directions, and Σ is the surface density.

The momentum equation (first Euler equation), $\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla})\vec{v} = -\vec{\nabla}\left(\frac{p}{\rho}\right)$ in cylindrical coordinates are (don't forget to take the derivatives of the unit vectors)

$$\frac{\partial v_R}{\partial t} + v_R\frac{\partial v_R}{\partial R} + \frac{v_\phi}{R}\frac{\partial v_R}{\partial \phi} - \frac{v_\phi^2}{R} = -\frac{\partial}{\partial R}(\Phi + h) \tag{42}$$

and

$$\frac{\partial v_\phi}{\partial t} + v_R\frac{\partial v_\phi}{\partial R} + \frac{v_\phi}{R}\frac{\partial v_\phi}{\partial \phi} + \frac{v_\phi v_R}{R} = -\frac{1}{R}\frac{\partial}{\partial \phi}(\Phi + h) \tag{43}$$

where Φ is the gravitational potential at $z = 0$, which is related to Σ via Poisson equation,

$$\nabla^2\Phi = 4\pi G\Sigma\delta(z) \tag{44}$$

(recall that $\Sigma = \int \rho dz$).

The quantity h introduced into Euler equations represents the pressure force; we could have written the right hand of Equations 42 and 43 respectively as $-\frac{1}{\rho}\frac{\partial p}{\partial R}$ and $-\frac{1}{\rho R}\frac{\partial p}{\partial \phi}$, respectively, where ρ is the volume density and p is the pressure.

Since we don't care about the vertical structure of the disk, we can take the disk to be uniform in the z direction. Using the definition of the sound speed, $c_s^2 = (\partial p/\partial \rho)$, we have

$$\frac{1}{\rho}\frac{\partial p}{\partial R} = c_s^2 \frac{\partial \ln \rho}{\partial R} = c_s^2 \frac{\partial \ln \Sigma}{\partial R} \tag{45}$$

and similarly,

$$\frac{1}{\rho}\frac{\partial p}{\partial \phi} = c_s^2 \frac{\partial \ln \Sigma}{\partial \phi}. \tag{46}$$

We can therefore relate h to Σ by

$$dh = c_s^2 d \ln \Sigma. \tag{47}$$

Now we add a small perturbation, by writing $\Sigma = \Sigma_0 + \Sigma_1$, $v_R = v_{R,0} + v_{R,1}$, $v_\phi = v_{\phi,0} + v_{\phi,1}$ and $\Phi = \Phi_0 + \Phi_1$ and $h = h_0 + h_1$, where the subscripts '0' and '1' refer to the unperturbed and perturbed quantities, respectively. Since for the *unperturbed* disk $v_{R,0} = 0$ (the disk only rotates, but does not expand or shrinks), the first Euler Equation (Equation 42) becomes

$$\frac{v_{\phi,0}^2}{R} = \frac{\partial}{\partial R}(\Phi_0 + h_0) = \frac{\partial \Phi_0}{\partial R} + c_s^2 \frac{\partial}{\partial R} \ln \Sigma_0 \tag{48}$$

In the case of interest to us, the sound speed c_s is much smaller than the rotation speed, $v_{\phi,0}$; e.g., in Galactic interstellar gas, $c_s \simeq 10$ km s^{-1} $\ll v_{\phi,0} \simeq 200$ km s^{-1}. We can thus neglect the last term in Equation 48, to write

$$v_{\phi,0} \simeq \sqrt{R \frac{\partial \Phi_0}{\partial R}} = R\Omega(R) \tag{49}$$

where we used Equation 26.

Keeping only terms in the first order in perturbations, and using $v_{R,0} = 0$, $v_{\phi,0} = R\Omega(R)$, Σ_0 that is independent on ϕ and $v_{\phi,0}$ that is independent on ϕ, the continuity, Euler and the potential equations (Equations 41 – 44) become

$$\frac{\partial \Sigma_1}{\partial t} + \frac{1}{R}\frac{\partial}{\partial R}(\Sigma_0 R v_{R,1}) + \Omega(R)\frac{\partial}{\partial \phi}(\Sigma_1) + \frac{\Sigma_0}{R}\frac{\partial}{\partial \phi}(v_{\phi,1}) = 0, \tag{50}$$

$$\frac{\partial v_{R,1}}{\partial t} + +\Omega(R)\frac{\partial v_{R,1}}{\partial \phi} - 2\Omega(R) v_{\phi,1} = -\frac{\partial}{\partial R}(\Phi_1 + h_1), \tag{51}$$

$$\frac{\partial v_{\phi,1}}{\partial t} + v_{R,1}\left(\frac{\partial(\Omega(R)R}{\partial R} + \Omega\right) + \Omega\frac{\partial v_{\phi,1}}{\partial \phi} = -\frac{1}{R}\frac{\partial}{\partial \phi}(\Phi_1 + h_1) \tag{52}$$

Using the definition of κ in Equation 29, we can write Equation 52 as

$$\frac{\partial v_{\phi,1}}{\partial t} + \frac{\kappa^2}{2\Omega} v_{R,1} + \Omega \frac{\partial v_{\phi,1}}{\partial \phi} = -\frac{1}{R}\frac{\partial}{\partial \phi}(\Phi_1 + h_1). \tag{53}$$

Finally, the linearized Poisson equation becomes

$$\nabla^2 \Phi_1 = 4\pi G \Sigma_1 \delta(z). \tag{54}$$

We note that $h_1 \simeq c_s^2 \Sigma_1 / \Sigma_0$, which follows directly from Equation 47.

We next need to check the conditions for the perturbation to grow. The standard way is to expand any perturbed quantity Q_1 ($Q = \Sigma, v_R, v_\phi, h, \Phi$) as sum of modes a, each characterized by an angular frequency ω and azimuthal wavenumber $m > 0$, i.e.,

$$Q_1 = \mathrm{Re}\left(\sum_a Q_a(R) e^{-i(\omega t - m\phi)} \right) \tag{55}$$

Thus, e.g.,

$$\frac{\partial v_{\phi,1}}{\partial \phi} = i m v_{\phi,1} \ ; \quad \frac{\partial \Sigma_1}{\partial t} = -i\omega \Sigma_1 \ ; \quad \frac{\partial \Sigma_1}{\partial \phi} = im\Sigma_1 \ ; \ \text{ etc.} \tag{56}$$

Substituting in Equations 50, 51 and 53 and solving, we find

$$i(m\Omega - \omega)\Sigma_a + \frac{1}{R}\frac{\partial}{\partial R}(\Sigma_0 R v_{R,1}) + \frac{im\Sigma_0}{R} v_{\phi,a} = 0, \tag{57}$$

$$v_{R,a} = -\frac{i}{\Delta}\left[(m\Omega - \omega)\frac{d}{dR}(\Phi_a + h_a) + \frac{2m\Omega}{R}(\Phi_a + h_a) \right], \tag{58}$$

$$v_{\phi,a} = \frac{1}{\Delta}\left[\frac{\kappa^2}{2\Omega}\frac{d}{dR}(\Phi_a + h_a) + \frac{m(m\Omega - \omega)}{R}(\Phi_a + h_a) \right], \tag{59}$$

where

$$\Delta \equiv \kappa^2 - (m\Omega - \omega)^2, \tag{60}$$

and $h_a = c_s^2 \Sigma_a / \Sigma_0$. Note that in Equation 55 we sum over all modes Q_a. Since we assume that the modes are independent on each other, we can write equations 57 – 59 or each mode separately.

For the Poisson Equation, we write Equation 54 as

$$\nabla_{\vec{x}}^2 \Phi_1 + \frac{\partial^2 \Phi_1}{\partial z^2} = 4\pi G \Sigma_1 \delta(z). \tag{61}$$

We consider a plane wave perturbation in the disk, $\Sigma_1 = \tilde{\Sigma}_1 e^{i(\vec{k}\cdot\vec{x} - \omega t)}$. We thus seek a solution of the form

$$\Phi_1(\vec{x}, z, t) = \tilde{\Phi} e^{i(\vec{k}\cdot\vec{x}) - \omega t)} e^{-|kz|}. \tag{62}$$

This will imply

$$\tilde{\Phi}_1 = -\frac{2\pi G}{|k|}\tilde{\Sigma}_1 \tag{63}$$

5.2. Local Instability

Since we ar einterested in the formation of arms, we proceed by assuming that the characteristic size of the perturbation is much smaller than the size of the disk. We write the perturbation mode as

$$\Sigma_1(R,\phi,t) = A(R,t)e^{i[m\phi+f(R,t)]} \tag{64}$$

Here, $A(R,t)$ is a slowly varying function of R that sets the amplitude of the density wave, and $f(R,t)$ is the shape function.

At fixed time, $m\phi + f(R,t) =$ const defines a curve in the disk where the phase of the perturbation mode is the same. The curves defined by $m\phi + f(R,t) = 2\pi n$ where $n = 0, \pm1, \pm2, ...$ therefore correspond to the peaks in the density waves, and delineate the ridges of the spiral density arms. The radial separation between adjacent arms at a given azimuth is ΔR such that $|f(R+\Delta R,t) - f(R,t)| = 2\pi$.

Under the assumption of tight winding, $\Delta R \ll R$, we can Taylor expand $f(R+\Delta R,t) \approx f(R,t) + (\partial f/\partial R)\Delta R$, and so $\partial f/\partial R = 2\pi/\Delta R$. In this case, we can write the perturbation in the neighborhood of a point (R_0,ϕ_0) (expanding Equation 64) as

$$\Sigma_1(R,\phi,t) \approx A(R_0,t)e^{i[m\phi_0+f(R_0,t)]+(\partial f/\partial R)(R-R_0)} = \Sigma_a e^{ik(R_0,t)(R-R_0)}, \tag{65}$$

where

$$\Sigma_a = A(R_0,t)e^{i[m\phi_0+f(R_0,t)]} \tag{66}$$

and

$$k(R_0,t) \equiv \left[\frac{\partial f}{\partial R}\right]_{R_0} = \frac{2\pi}{\Delta R}. \tag{67}$$

Note that in the above expression, we neglected the variations with respect to the angle ϕ, because they are much slower than the radial variations for tightly wound waves. Under this assumption, a spiral density wave closely resembles a plane wave in the R direction., with wave-vector $k\hat{e}_R$ and wavelength ΔR.

Under this assumption, the potential perturbation Φ_a is related to Σ_a by

$$\Phi_a = -\frac{2\pi G\Sigma_a}{|k|} = -\frac{2\pi G}{|k|}A(R_0,t)e^{i[m\phi_0+f(R_0,t)]} \tag{68}$$

(see Equation 63). This means the following. In equation 58 and 59, we have, on the RHS, two terms: one $\sim \frac{d\Phi_a}{dR}$ and the second $\propto \frac{\Phi_a}{R}$. If we differentiate Φ_a with respect to R using Equation 68, we get

$$\frac{d\Phi_a}{dR} = if'(R)\Phi_a = i\frac{2\pi}{\Delta R}\Phi_a = ik\Phi_a \tag{69}$$

Thus, the ratio of the first to the second terms is $R/\Delta R \gg 1$. This means that we can neglect the second term (proportional to Φ_a/R). A similar conclusion holds for the h_a term, since $h_a \propto \Sigma_a$ (see below Equation 54).

Using Equation 69, we can now write Equations 58 and 59 as

$$v_{R,a} = \frac{(m\Omega - \omega)\, k(\Phi_a + h_a)}{\Delta}, \tag{70}$$

and

$$v_{\phi,a} = \frac{ik\kappa^2(\Phi_a + h_a)}{2\Omega\Delta}. \tag{71}$$

Similar arguments can be used to simplify Equation 57: Since Equations 70, 71 show that $v_{R,a}$ and $v_{\phi,a}$ are of the same order, the second term (the derivative term) is greater than the 3rd term by $\mathcal{O}(|kR|)$, and therefore the 3rd term can be neglected. Using Equation 70, we replace the derivative $d(R\Sigma_0 v_{R,a})/dR$ by $ikR\Sigma_0 v_{R,a}$. This allows us to write Equation 57 as

$$(m\Omega - \omega)\Sigma_a + k\Sigma_0 v_{R,a} = 0. \tag{72}$$

Finally, inserting the expression of $v_{R,a}$ into Equation 72, and using $h_a = c_s^2\Sigma_a/\Sigma_0$, we finally get the dispersion relation for a gaseous disk in the tight-winding approximation:

$$(m\Omega - \omega)^2 = \kappa^2 - 2\pi G\Sigma_0|k| + k^2c_s^2. \tag{73}$$

For **axisymmetric perturbations**, $m = 0$, and the dispersion relation reads

$$\omega^2 = \kappa^2 - 2\pi G\Sigma_0|k| + k^2c_s^2. \tag{74}$$

Note that without the κ^2 term, this is simply the Jeans criterion for gravitational instability in a thin disk. The extra κ term is owing to the Coriolis force associated with the disk's rotation, and which provides a centrifugal force that combines with the pressure to resist the growth of perturbations.

Equation 74 is used to investigate whether the disk is stable against perturbations with $m = 0$. Modes with $\omega^2 < 0$ grow exponentially in time (see Equation 55), and so are unstable. On the other hand, modes with $\omega^2 > 0$ are stable.

It is costume to define the two dimensionless parameters:

$$Q \equiv \frac{c_s\kappa}{\pi G\Sigma_0} \tag{75}$$

(the **Toomre's Q parameter**), and

$$\lambda_{\rm crit} \equiv \frac{4\pi G\Sigma_0}{\kappa^2}. \tag{76}$$

Using these parameters, we can write the dispersion relation (Equation 74) as

$$\omega^2 = \frac{4\pi^2 G\Sigma_0}{\lambda_{\rm crit}} \left[1 - \frac{\lambda_{\rm crit}}{\lambda} + \frac{Q^2}{4} \left(\frac{\lambda_{\rm crit}}{\lambda} \right)^2 \right] \tag{77}$$

where $\lambda \equiv 2\pi/|k|$.

It is evident from Equation 77 that for zero pressure ($c_s = 0$) disk, $\lambda_{\rm crit}$ is the largest unstable wavelength. The line separating stable and unstable modes is given by

$$Q(\lambda) = 2\sqrt{\frac{\lambda}{\lambda_{\rm crit}} \left(1 - \frac{\lambda}{\lambda_{\rm crit}} \right)} \tag{78}$$

A disk with Q value larger than $Q(\lambda)$ is stable for the perturbation mode with wavelength λ, and vice versa.

The function $Q(\lambda)$ has a maximum $Q_{\rm max} = 1$ at $\lambda = \lambda_{\rm crit}/2$. Thus, a disk is stable against all local perturbations if $Q > 1$. If Q decreases, the first mode to become unstable is the one with $\lambda = \lambda_{\rm crit}/2$.

Final note. The derivation here was done for a gaseous disk. A similar derivation can be done for stellar disk, in which the effective pressure is due to the random motion of the stars. The result is very similar, with stable disk for

$$Q \equiv \frac{\sigma_R \kappa}{3.36 G\Sigma_0} > 1 \tag{79}$$

where σ_R is the radial velocity dispersion.

The most reliable estimate of the value of Q comes from measurements near the solar neighborhood. The values obtained are $Q_\star = 2.7 \pm 0.4$ and $Q_g = 1.5$. Thus, our galaxy is stable, but not too far from being unstable. In practice, simulations show that a spiral pattern will grow idf $Q \lesssim 1.2$ or so (see Figure 7). This is very close to the observed value.

6. The spiral arms

One of the basic questions in study of disk galaxies is the origin of the spiral arms. Observationally, spiral arms are seen in disks that contain gas, but not in gas-poor (S0) galaxy disks. When the sense of the galactic rotation is known, the spiral arms almost always trail the rotation.

The first ingredient in producing spiral arms is differential rotation. For galaxy with flat rotation curve, $v(R) = const$, its angular velocity is $\Omega(R) = v/R \propto R^{-1}$. This means that

any feature in the disk will be wrapped into a trailing spiral pattern due to the differential rotation.

However, differential rotation is not enough to explain the observed spiral structure. If we again assume flat rotation curve, $\Omega = v/R$ then

$$\frac{d\Omega}{dR} = -\frac{v}{R^2} \tag{80}$$

Thus, two points on the disk that are separated by radius ΔR that are initially at the same azimuth, will be sheared apart over time. After time t, they will be separated by an angle

$$\left|\frac{d\Omega}{dR}\right| \Delta R t$$

and they will return to the same azimuth after a time given by

$$\left|\frac{d\Omega}{dR}\right| \Delta R t = 2\pi \tag{81}$$

Therefore, a spiral pattern will be wrapped up on a radial scale ΔR after a time t given by

$$\frac{\Delta R}{R} = \frac{2\pi R}{vt} \tag{82}$$

If we take values from the Milky Way, e.g., $R = 8.5$ kpc and $v = 200$ km/s, we obtain

$$\frac{\Delta R}{R} = 0.25 t_{Gyr}^{-1} \tag{83}$$

This is a very tightly wrapped spiral; spiral arms like this would make an angle of ~ 2 degrees to the tangent direction. As opposed to that, we observed the arms not to be so tight: in Sa spiral galaxies, they are ~ 5 degrees, and in Sc spirals, they are around 10 - 30 degrees.

Rather, the leading idea is that of **spiral density waves**, proposed by Lin & Shu. The orbits in spiral galaxies are not quite circulars, but rather ellipticals. These ellipses are slightly tilted with respect to each other (see Figure 8).

Thus, there are regions of slightly higher density then the surroundings. Higher density also implies higher gravity. Thus, objects, such as gas clouds will be attracted to these regions and will drift towards them.

The properties of the spiral arms can be explained if they are **not** rotating with the stars, but rather are **density waves**.

- Spiral arms are locations where the stellar orbits are such that stars are more densely packed.
- Gas is also compressed, possibly triggering star formation and generating population of young stars. Indeed, most of a galaxy's star formation takes place in the spiral arms. Furthermore, the brightest - bluest - shortest lived stars will never be too far from the spiral arm where they were born.
- Arms rotate with **pattern speed**, which is not equal to the circular velocity, namely, long lived stars enter and leave spiral arms repeatedly.
- The pattern speed is less than the circular velocity, partially alleviating the winding up problem.

In isolated disk, creation of a density wave requires an instability. Such instability may result from self gravity of the stars and /or gas. If the disk is massive (strong self gravity), small perturbations which slightly compress part of the disk will form clumps, which will be further compressed by their self gravity, however- the extra pressure will resist compression. If the disk is unstable (Toomre's Q parameter is less than unity), the spiral arms will spontaneously develop within a few rotation periods. During the process, if the disk heats up, Q is raised and the spiral pattern will die.

REFERENCES

[1] H. Mo, F. van den Bosch and S. White, *Galaxy Formation and Evolution* (Cambridge), chapter 11.

[2] J. Binney and S. Tremaine, *Galactic Dynamics* (Princeton), chapter 6

[3] L.S. Sparke and J.S. Gallagher, *Galaxies in the Universe* (Cambridge), chapter 2

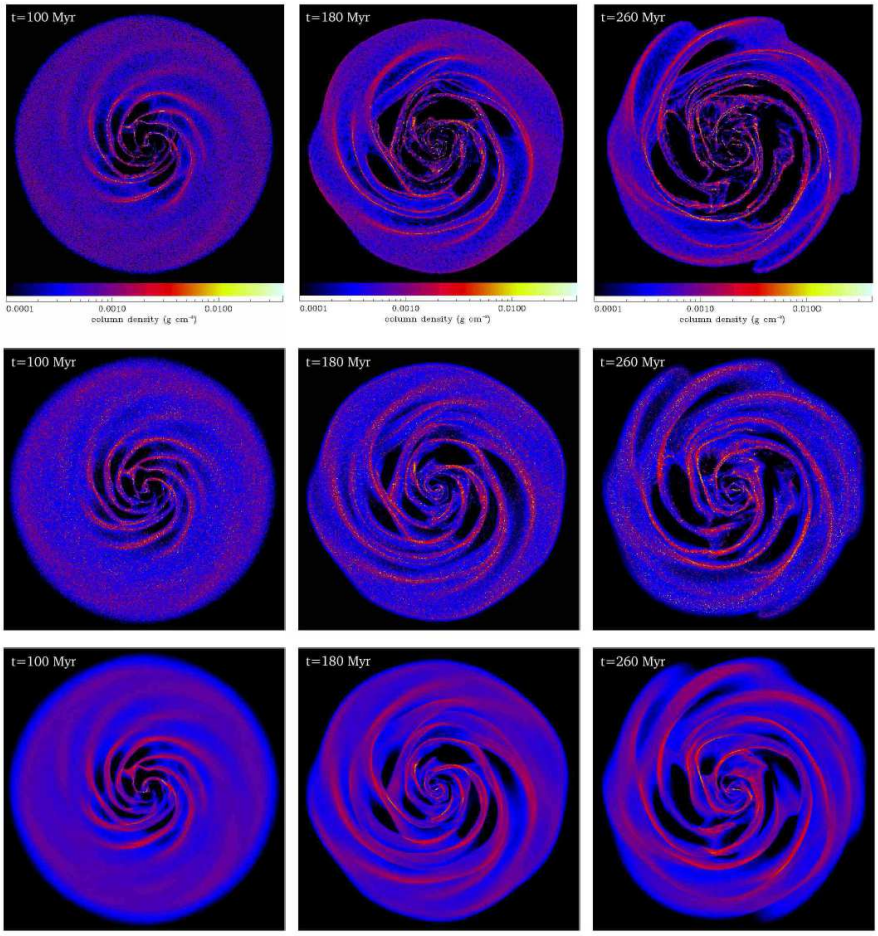

Fig. 7.— Simulation showing the formation and evolution of a spiral pattern in gaseous disk, for different gas temperatures. Disk size in these simulations is 11 kpc. Taken from Dobbs & Bonnell, arXiv:0801.3562

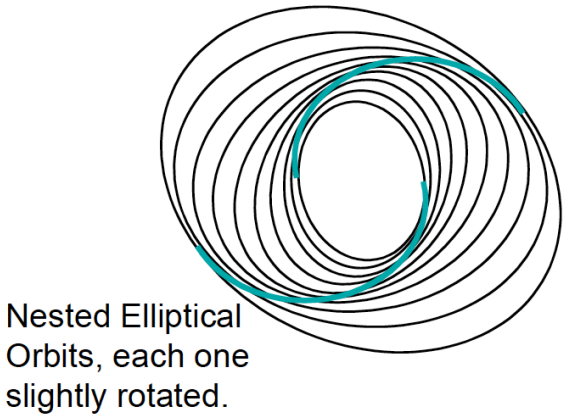

Fig. 8.— Self consistent constitution of density perturbations in a disk.