Galaxy formation and evolution
53863, 5 op, autumn 2020

on Zoom

Lecture 10: Formation of disk galaxies, 13/11/2020

On this lecture we will discuss

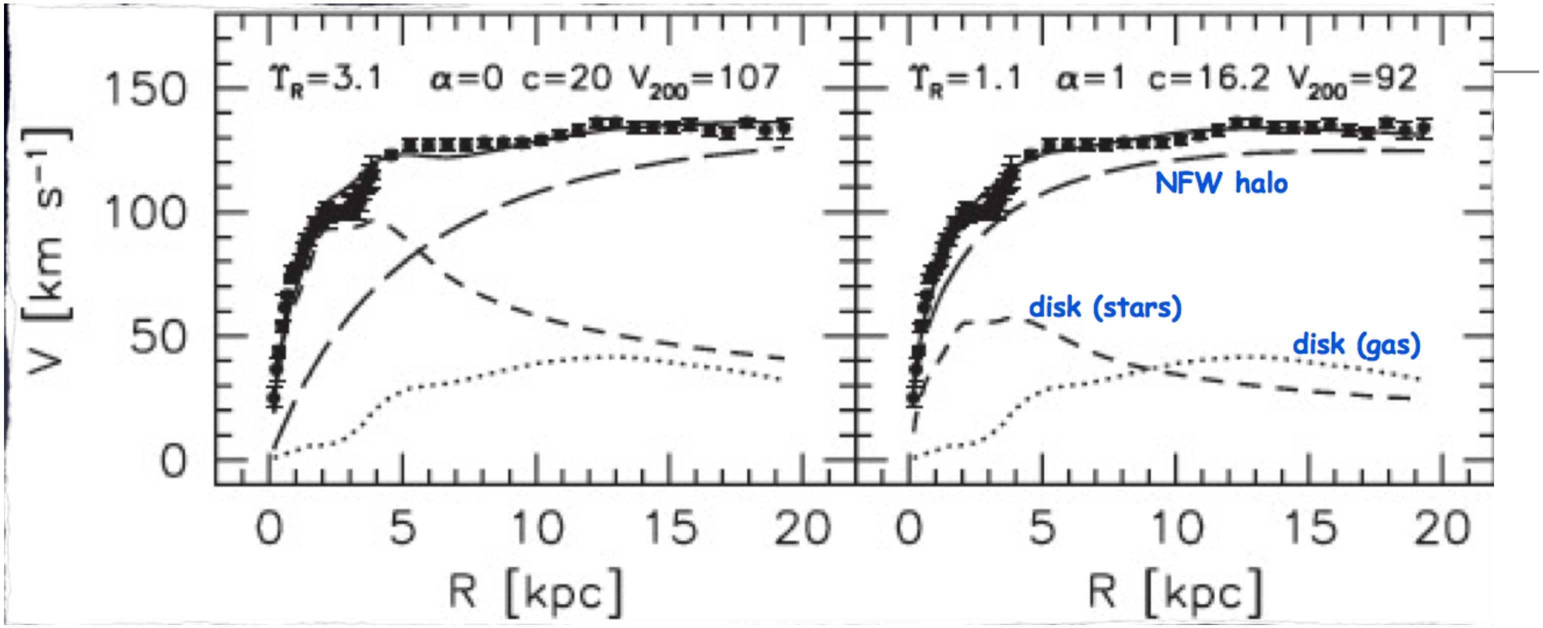
1. Basic observational results of disk galaxies.
2. Scaling relations and the Tully-Fisher relation of disk galaxies.
3. Formation of disk galaxies. Basic processes.
4. Transport of angular momentum and the angular momentum problem.
5. The standard model of disk galaxy formation. Adiabatic contraction.
6. The origin of exponential disks.
7. The lecture notes correspond to: MBW: pages 495-520 (§11.1-11.4)

10.1 Disk galaxies: Observational results

- A disk galaxy consists of a disk component made of stars and atomic and molecular gas, often with spiral arms and in about half of the cases a central bar component.
- A successful model of disk galaxy formation should be able to explain the observed structure of disk galaxies and the following observational facts:

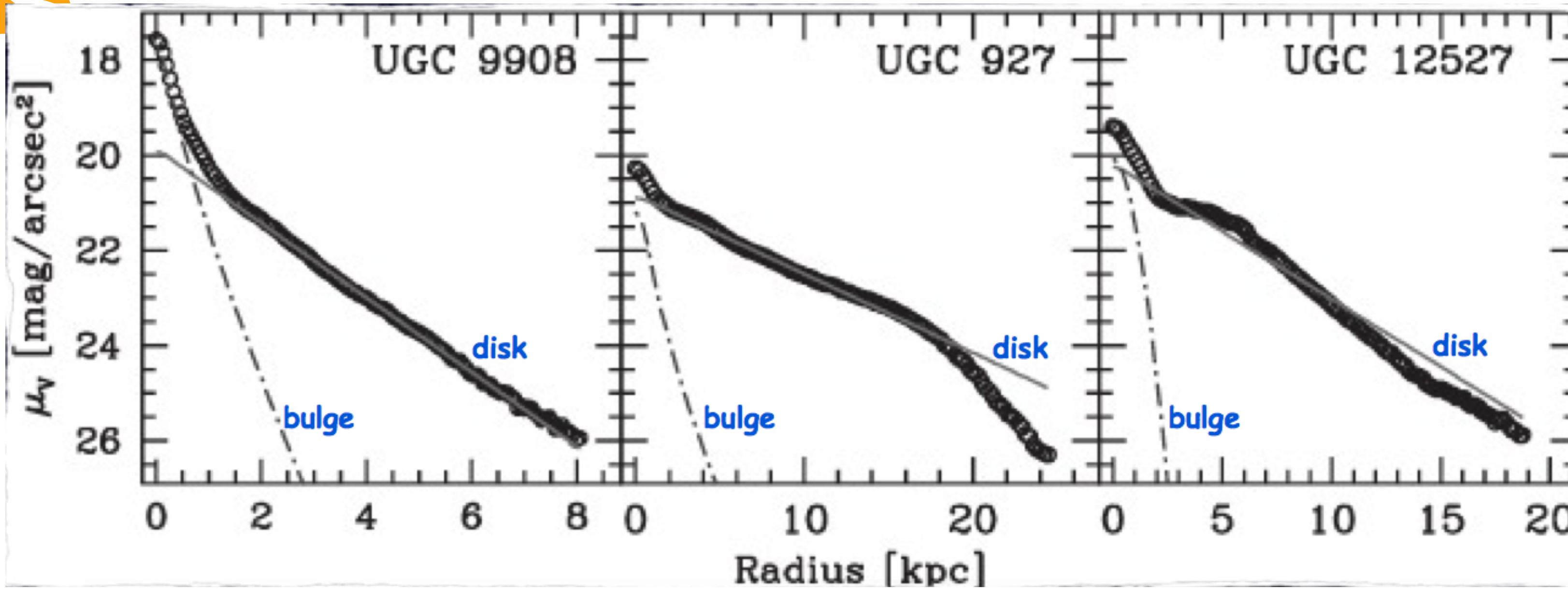
1. Brighter disks are, on average, larger, redder, rotate faster, and have a smaller gas mass fraction.
2. Disk galaxies have flat rotation curves.
3. The surface brightness profiles of disk galaxies are close to exponential.
4. The outer parts of disks are generally bluer, and have lower metallicity than the inner parts.

Rotation curves



- Disk galaxies have flat rotation curves that extend well beyond the observational stellar disk -> very strong evidence for dark matter.
- A unique disk-bulge-halo decomposition is very difficult to obtain and often a maximal disk decomposition is used, which maximizes the contribution of the stellar disk (maximum mass/light ratio for the stars).

Surface brightness profiles



- Surface brightness profiles of disk galaxies are close to exponential.
- Deviations from the exponential at small radii are attributed to the bulge and/or bar component.
- Deviations from the exponential at large radii are attributed to density thresholds for star formation, radial migration and/or the maximum angular momentum of the infalling gas.

Mathematical models of disks

- Disks are often modelled by infinitesimally thin, exponential disks for which we get the following surface mass density:

$$\Sigma(R) = \Sigma_0 e^{-R/R_d} \quad M_d = 2\pi \int_0^\infty \Sigma(R) R dR = 2\pi \Sigma_0 R_d^2$$

- The gravitational potential and circular velocity of an infinitesimally thin disk can be found with the help of Bessel functions (J_0 the cylindrical Bessel function of zero order. I_0 and K_0 are modified Bessel functions):

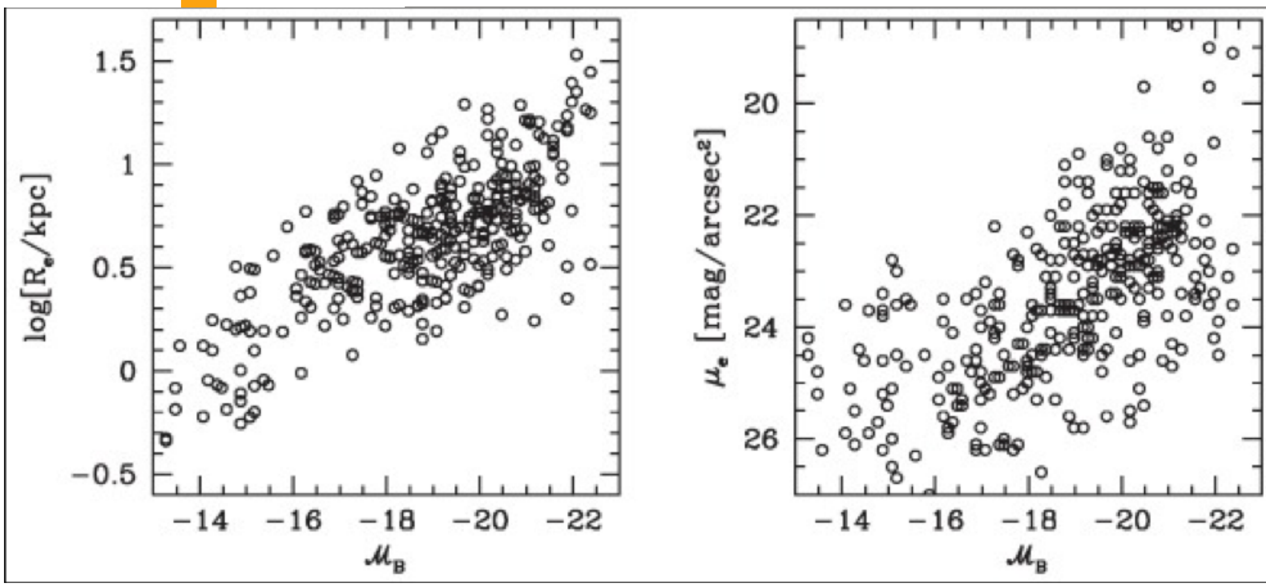
$$\Phi(R,z) = -2\pi G \int_0^\infty J_0(kR)\bar{\Sigma}(k) e^{-k|z|} dk$$

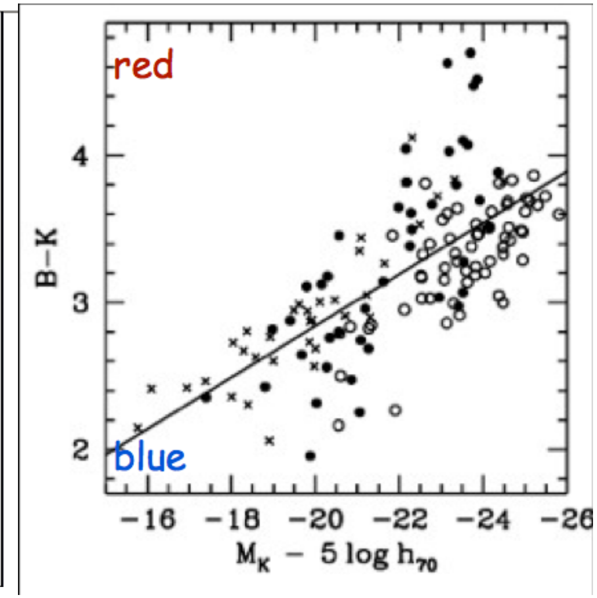
$$V_{c,d}^2(R) = -4\pi G \Sigma_0 R_d y^2 [I_0(y)K_0(y) - I_1(y)K_1(y)], \quad y = R/(2R_d)$$

- However, real disks have also a vertical thickness in the z-direction and often a simple isothermal sheet model is used:

$$\rho(R,z) = \rho(R,0)\mathrm{sech}^2(z/2z_d)$$

Observational correlations of disks





- Brighter disks are:

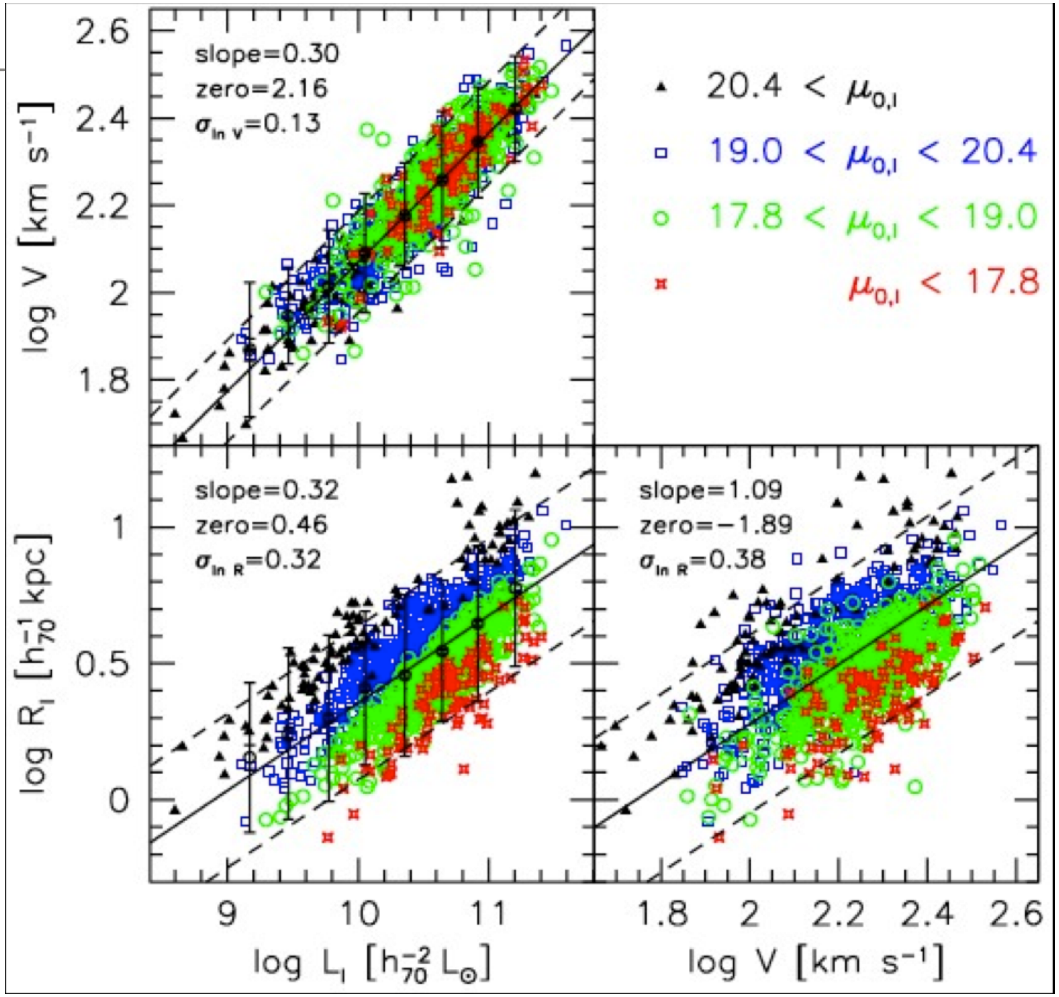
1. Larger and redder
2. Have higher central surface brightness values.
3. Have smaller gas fractions (not shown here directly).

The Tully-Fisher relation

- Brighter disks also rotate faster as manifested in the Tully-Fisher relation:

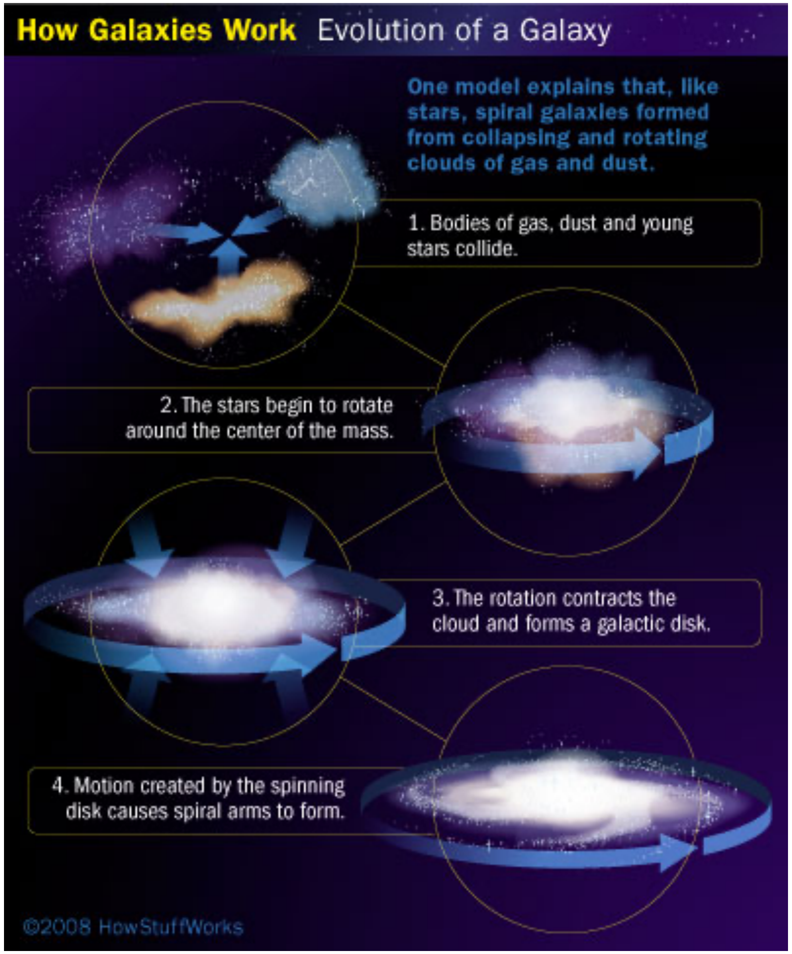
$$V \propto L_I^{0.30}$$

- The slope of the TF-relation depends on the photometric band. The scatter in the TF-relation is not correlated with the surface brightness.
- Theoretically and numerically it is very challenging to reproduce the TF zero point.



10.2 Formation of disk galaxies: Basic processes I

1. Hot shock-heated gas inside a dark matter halo cools radiatively.
2. As gas cools, its pressure decreases causing the gas to contract.
3. The emission of photons is isotropic and thus the angular momentum is conserved.
4. The gas sphere contracts, spins up and flattens.
5. The surface density of the disk increases, at some point the critical threshold for star formation is reached and a disk galaxy is born.



Basic processes II

- In the standard picture of disk galaxy formation developed by Fall&Efstathiou (1980) disk galaxies are in centrifugal equilibrium and therefore their structure is governed by their specific angular momentum distribution.
- The main assumptions in the standard picture are:

1. The angular momentum originates from cosmological torques between neighbouring dark matter haloes.
2. Baryons and dark matter acquire initially identical specific angular momentum distributions.
3. Baryons conserve their specific angular momentum while cooling and settling in the dark matter haloes.

Angular momentum transport

- For a given angular momentum J, the state of lowest energy and hence the state preferred by nature, is the one in which all mass except an infinitesimal fraction δM collapses into a black hole, while δM is on a Keplerian orbit with radius R and M is the mass of the BH:

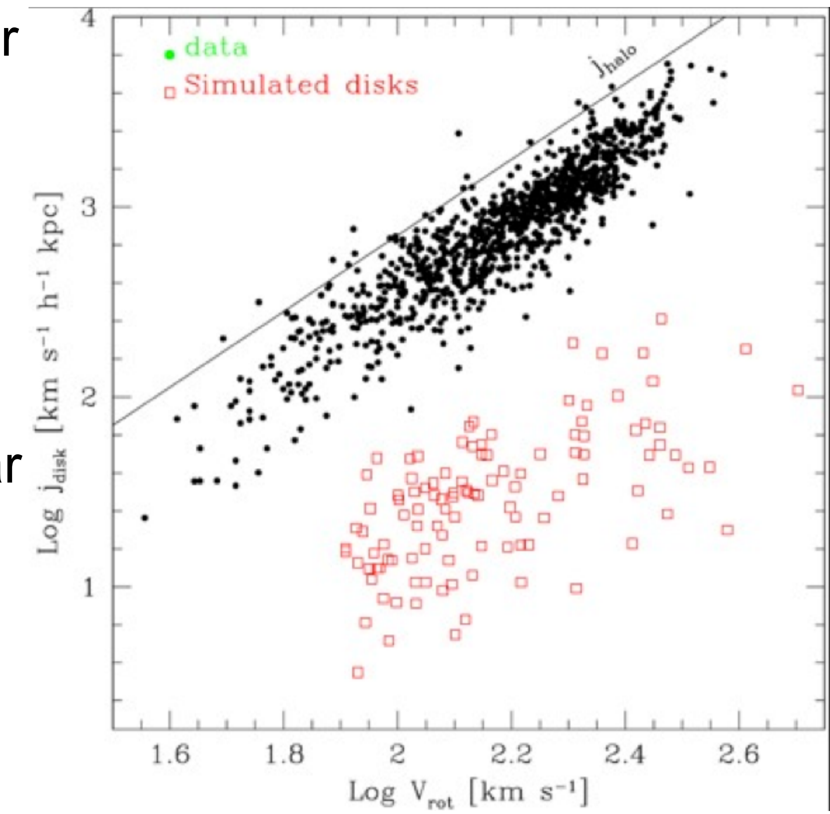
$$J = \delta M (GMR)^{1/2}$$

- This is clearly very different from realistic exponential disks.
- The reason for this discrepancy is that although the lowest energy state is preferred, its realisation requires efficient transport of angular momentum from inside out.
- Several mechanisms can transport angular momentum:

1. Secular evolution (gas viscosity or resonant scattering of stars&gas)
2. Hierarchical formation (dynamical friction & ram pressure scattering).

Angular momentum problem

- The fact that galaxies deviate from the minimum energy state indicates that angular momentum transport is not that efficient.
- However, in numerical simulations gas cooling causes much of the gas to condense into dense clumps.
- This clumpy gas is delivered to the disk via dynamical friction transferring orbital angular momentum of the gas to dark matter -> simulated disks end-up too small.
- This problem is known as the angular momentum catastrophe and strong super-nova feedback is required to prevent overcooling and loss of angular momentum.



10.3 The standard picture of disk galaxy formation: Disk mass

- In a first idealised approximation we ignore the self-gravity of the disk and assume that the dark matter is a singular, isothermal sphere:

$$\rho(r) = V_{\mathrm{vir}}^2/(4\pi G r^2)$$

- Next we assume that a mass fraction m_d of the total virial mass settles in the disk. The total virial mass can be calculated from the virial velocity together with a cosmological factor that accounts for the evolution of the virial density with redshift:

$$M_d = m_d M_{\mathrm{vir}} \simeq 1.3 \times 10^{11} h^{-1} M_\odot \left(\frac{m_d}{0.05}\right) \left(\frac{v_{\mathrm{vir}}}{200 \mathrm{\ km/s}}\right)^3 D^{-1}(z)$$

$$D(z) = \left[\frac{\Delta_{\mathrm{vir}}(z)}{100}\right]^{1/2} \left[\frac{H(z)}{H_0}\right]$$

Disk angular momentum I

- The disk angular momentum for an infinitesimally thin exponential disk:

$$J_d = 2\pi \int_0^\infty \Sigma(R) V_c(R) R^2 dR = 2 M_d R_d V_{\mathrm{vir}}$$

- Now we can define the parameter j_d as $J_d=j_d J_{vir}$, where J_{vir} is the angular momentum of the dark matter halo. The disk angular momentum can be related to the dimensionless spin parameter:

$$\lambda = \frac{J_{\mathrm{vir}} |E|^{1/2}}{G M_{\mathrm{vir}}^{5/2}} = \frac{1}{j_d} \frac{J_d |E|^{1/2}}{G M_{\mathrm{vir}}^{5/2}}$$

- Using the virial theorem, according to which we have:

$$E = -K = -\frac{M_{\mathrm{vir}} V_{\mathrm{vir}}^2}{2}$$

Disk angular momentum II

- Using the virial theorem we can also derive a disk scale radius:

$$J_d = \sqrt{2} j_d \lambda M_{\rm vir} R_{\rm vir} V_{\rm vir} \quad \& \quad J_d = 2 M_d R_d V_{\rm vir}$$

$$R_d = \frac{1}{\sqrt{2}} \lambda \left(\frac{j_d}{m_d} \right) R_{\rm vir}$$

- Combining with the cosmological expression for R_{vir}:

$$R_d \simeq 10 h^{-1}\ \mathrm{kpc} \left(\frac{j_d}{m_d} \right) \left(\frac{\lambda}{0.05} \right) \left(\frac{V_{\rm vir}}{200\ \mathrm{km/s}} \right) D^{-1}(z)$$

- Using the values for the Milky Way V_{vir}=220 km/s and j_d=m_d, M_d=5x10^{10} $M_\odot$ and R_d=3.5 kpc results in m_d~0.01 and λ~0.011. Cosmic baryon fraction f_{bar}~0.17, only 6% of the baryons in the disk and the low spin of the MW halo would make it relatively rare (mean λ~0.04).

Disk models with self-gravity

- Assuming that the disk has 1) self-gravity and is situated in a dark matter halo with an 2) NFW profile we have to do the following modifications:

$$R_d = \frac{1}{\sqrt{2}}\lambda\left(\frac{j_d}{m_d}\right)R_{\rm vir}F_R^{-1}F_E^{-1/2}$$

- F_E is defined by: $$E = -\frac{M_{\rm vir}V_{\rm vir}^2}{2}F_E$$
- F_R is defined by: $$F_R = \frac{1}{2}\int_0^{R_{\rm vir}/R_d} u^2 e^{-u}\frac{V_c(uR_d)}{V_{\rm vir}}du$$

- The total rotation curve is then due to a combination of the disk and halo masses and now a more probable λ~0.05 model is consistent with the observations: $$V_c^2(R) = V_{c,d}^2(R) + V_{c,h}^2(R) = V_{c,d}^2(R) + \frac{GM_{h,ac}(R)}{R}$$

Adiabatic contraction I

- When baryons cool and concentrate in the centres of dark matter haloes, they modify the halo structure by their gravitational action.
- In general, it is difficult to model this action of the disk on the halo accurately, except if the growth of the disk is slow compared to the dynamical time of the dark matter particles in the centre of the halo. In this case the system adjusts itself adiabatically (reversible) and the final state is independent of the path taken.
- For a spherically symmetric system with the particles on spherical orbits, adiabatic invariance implies that rM(r) is conserved:

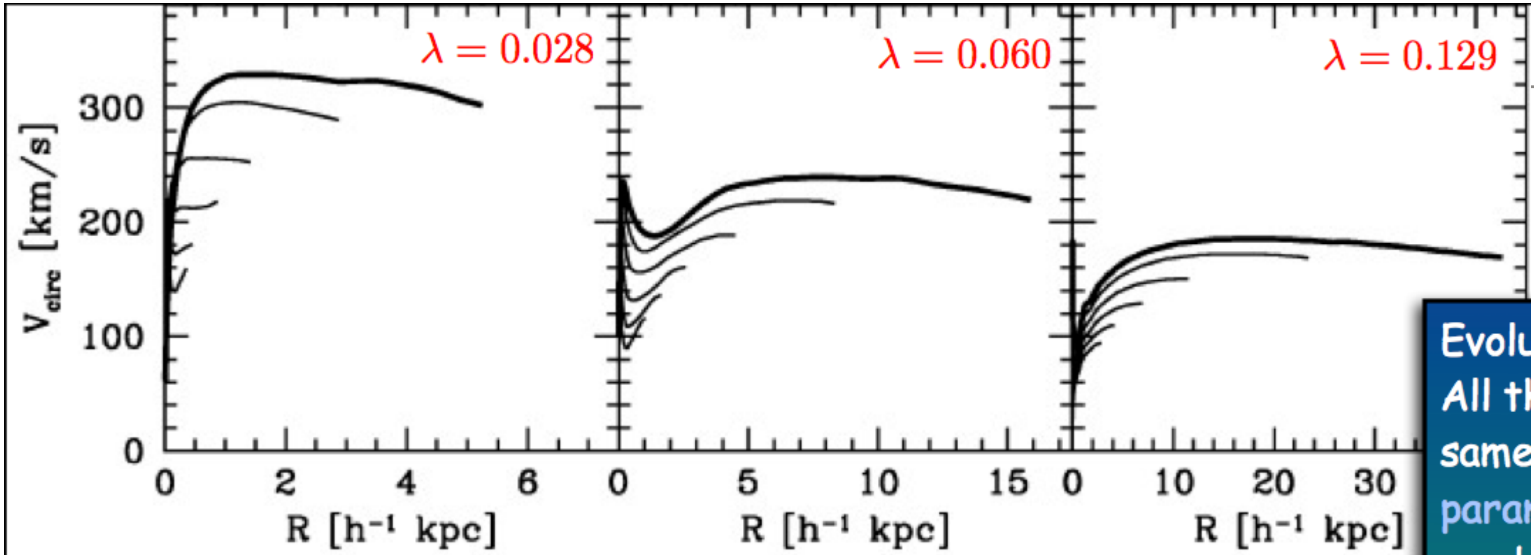
$$r_f M_f(r_f) = r_i M_i(r_i)$$

- Adiabatic contraction in a halo with an NFW DM profile:

$$M_f(r_f) = M_d(r_f) + [1 - m_d] M_i$$

Adiabatic contraction II

Thick lines at z=0, thinner lines at z=1, 2, 3, 4, 5 from top to bottom.



- For a given m_d and $M_d(r)$ the adiabatic equations can be solved iteratively for a given r_f and given r_i.
- For realistic disk masses m_d the effect of adiabatic contraction can be substantial. The effect is stronger for compact disks with small λ and V_{rot}/V_{vir}~1.4-1.8 instead of V_{rot}/V_{vir}~1, which is assumed in most galaxy formation models.

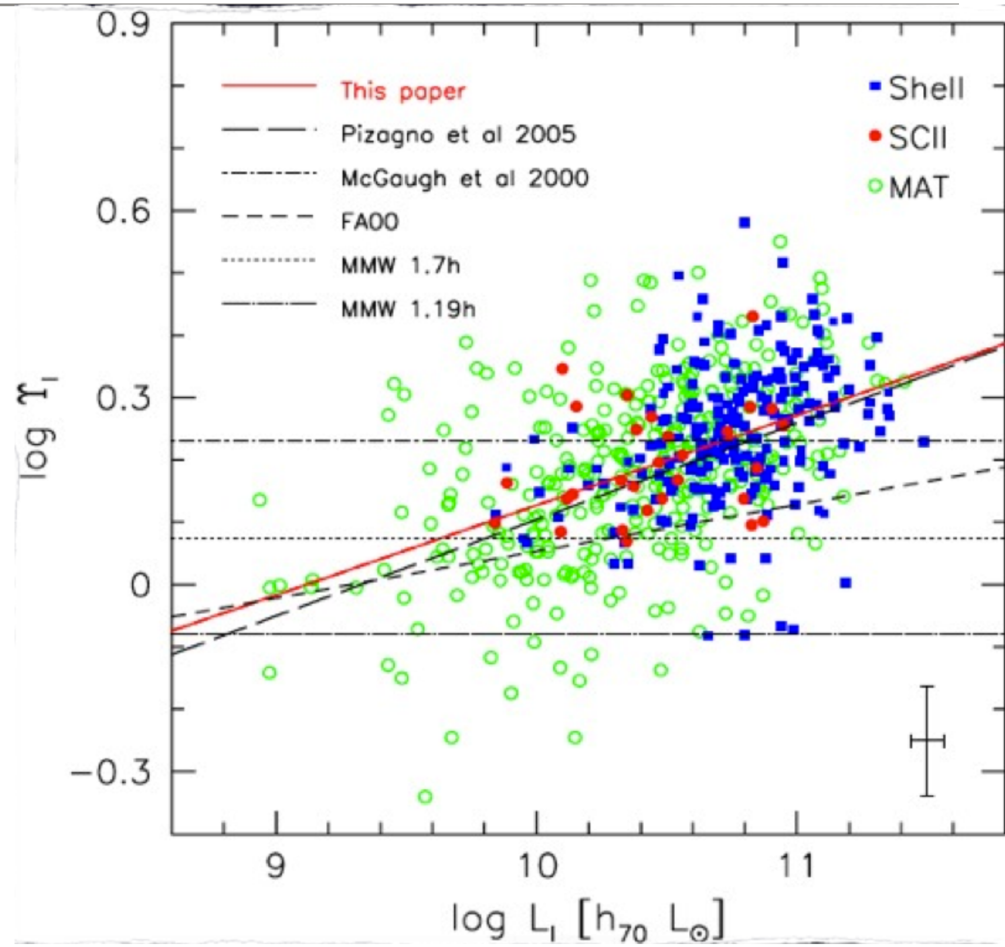
Formation of the bulge

- The bulges in disk galaxies have essentially properties that are similar of mini-elliptical galaxies, i.e. they are slowly rotating peanut-shaped objects with primarily very old stars. There are two main formation scenarios for bulges in disk galaxies:

1. The bulge formation could be the result of the merger of gas-rich galaxies, in which substantial amount of gas is funnelled through tidal torques to the centres of galaxies, where they result in a starburst (bulge formation) and feed the supermassive black hole. Large bulges are called classical bulges.
2. Smaller mass bulges might also form in internal secular processes through instabilities in the galactic disk. In this picture bulge formation is closely related to bar-formation and these lower mass bulges are often called pseudobulges to distinguish them from the more massive classical bulges.

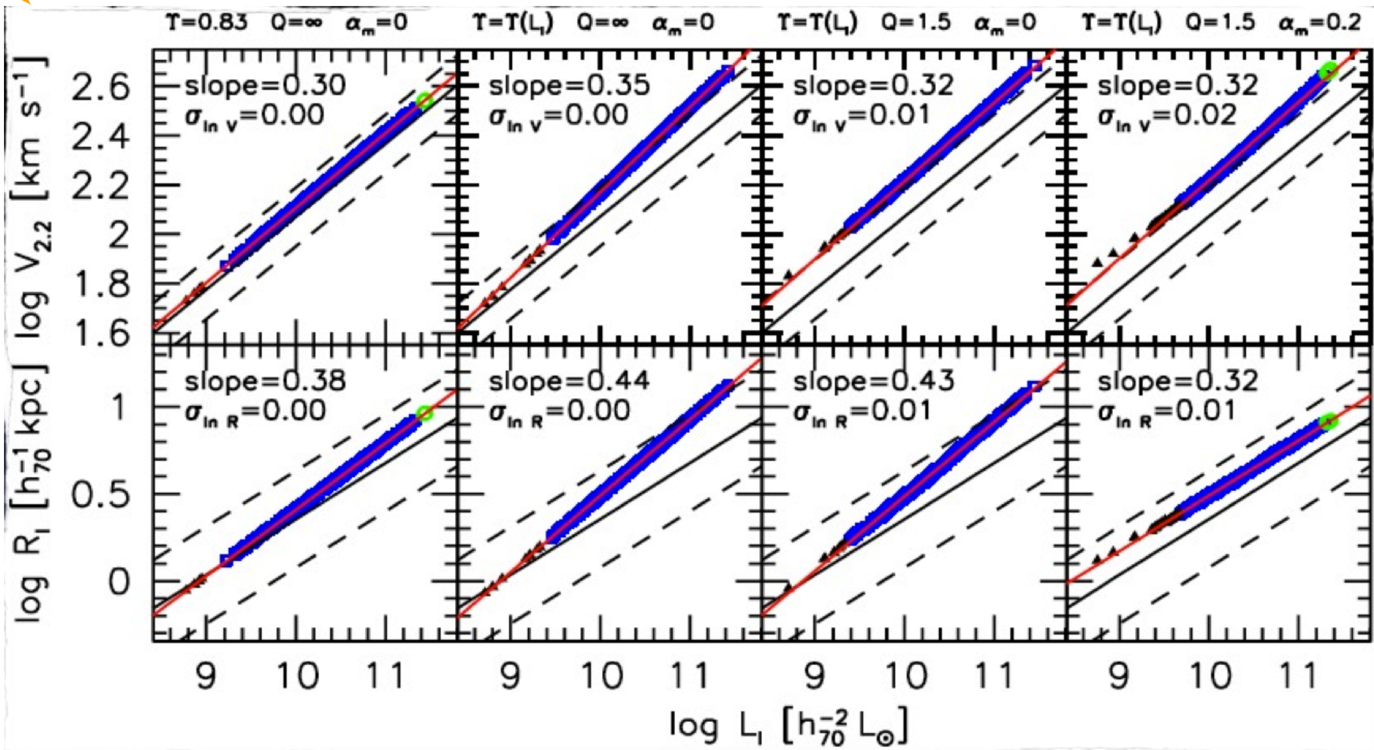
10.4 Constraints from observations: Stellar mass-to-light ratios

Mass-to-light ratio as function of luminosity



- The standard formation picture presented here can fit the observed disk scaling relations and the Tully-Fisher relation. However, this only works for unrealistically low stellar mass-to-light ratios and for disk models without gas.
- In reality the stellar mass-to-light ratio of disk galaxies increase with luminosity and this should be taken into account in the models.

Tully Fisher relation I



- Including the correction for the stellar M/L ratio (second column), allowing only disk matter with $\Sigma(R)>\Sigma_{crit}(R)$ to form stars (third column) and adjusting the disk-mass to halo-relation (m_d) (fourth column) improves the fits to the observed scaling relations.

Tully Fisher relation II

- The slope of the Tully-Fisher relation can be well fitted in this way, but the zero-point of the TF relation is still too high and also the disks are too large.
- Solutions to this problem include:

1. Lower stellar mass-to-light ratios: The M/L ratio can be lowered by increasing the relative weight of more luminous stars, but typically the model would require a large IMF shift, which is directly violated by observations.
2. Lower halo concentrations: Lowering the halo concentrations by a factor of two would reduce the velocities and produce a lower TF zero-point. Halo concentrations are determined by cosmology and thus cannot be changed too much.
3. Modify the adiabatic contraction model: Require halo expansion, i.e. from impulsive supernova feedback violating the adiabatic assumption and thus lowering the central velocities.

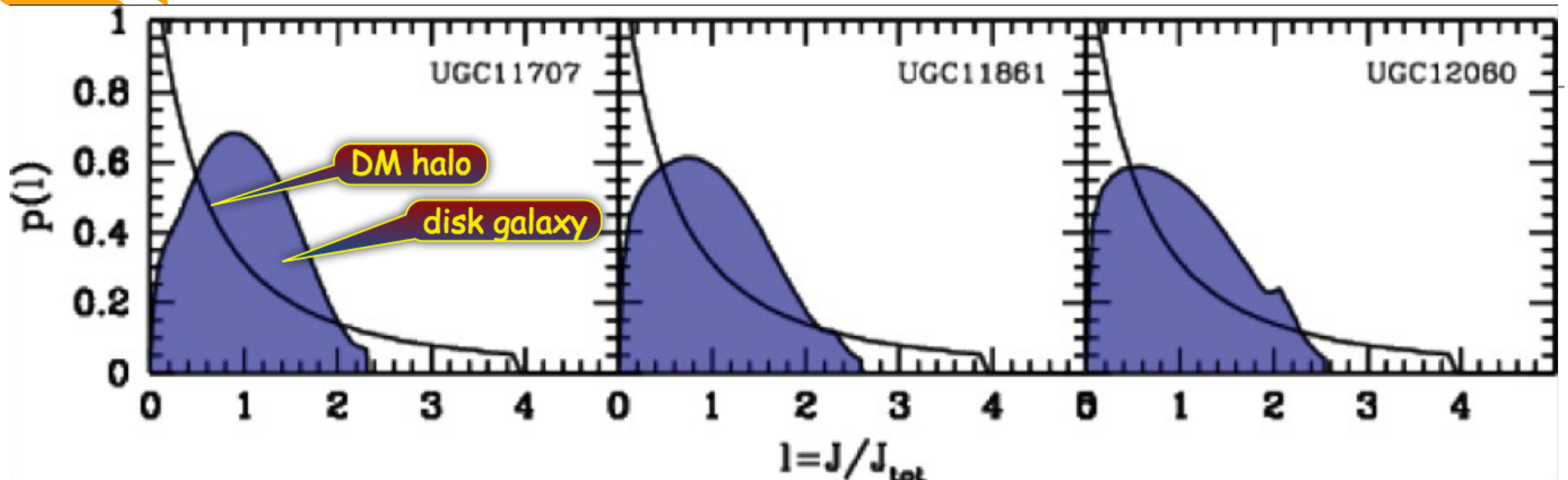
The origin of exponential disks I

- Assuming there is no angular momentum loss or redistribution during the disk formation process, the disk surface density profile is a direct reflection of the specific angular momentum distribution of the proto-galaxy: $\Sigma_d(R)$ <-> $M_{bar}(j_{bar})$ <-> $M_{DM}(j_{DM})$
- If the picture of detailed specific angular momentum conservation is correct then the following equation should be valid, where $j=rV_c(r)$:

$$\frac{M_d(<r)}{M_d} = \frac{M_h(<j)}{M_{\rm vir}}$$

- This means that for a given halo dark matter density and angular momentum profile one can predict the disk surface mass $\Sigma_d(R)$.

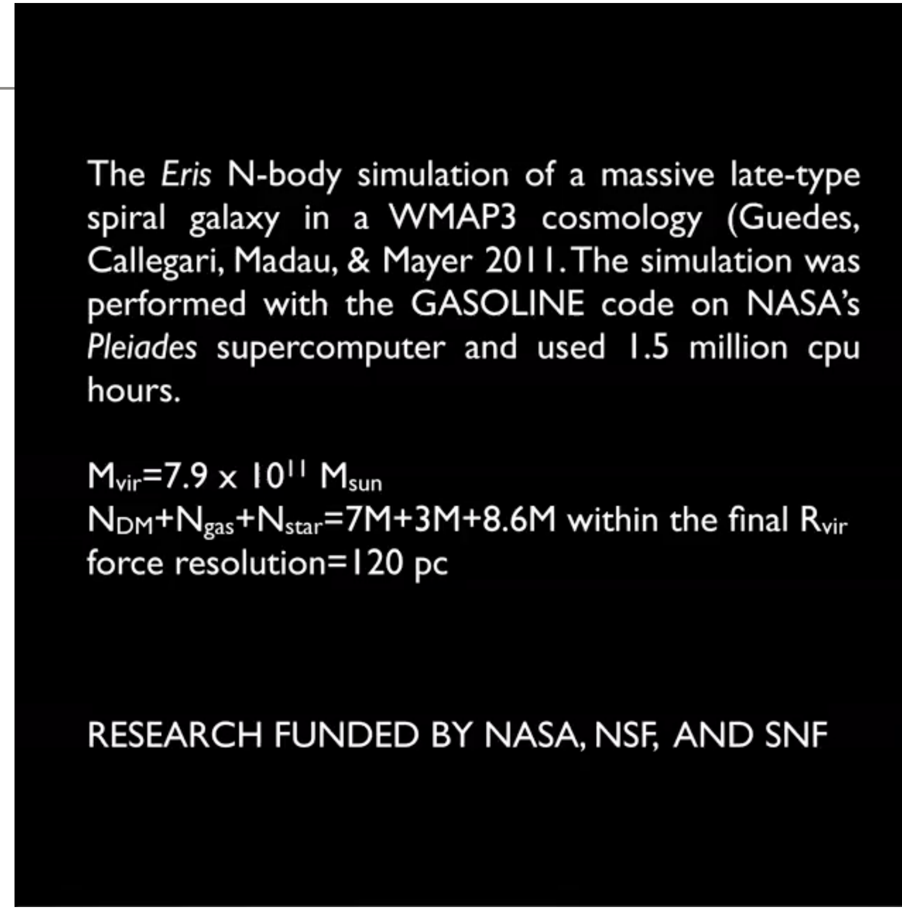
The origin of exponential disks II



- Unfortunately the disk surface density profiles predicted in this way look nothing like exponential disks.
- Observed galaxies lack both high and low specific angular momentum material compared to the predictions. The inside-out formation of disk galaxies leaves the highest angular momentum material in the halo and feedback somehow preferentially ejects low-angular momentum material.

Numerical simulations of disk galaxies

- A key requirement in numerical disk galaxy simulation is a model of strong supernova feedback that prevents overcooling and ensures prolonged star formation allowing time to build a disk.
- High numerical resolution is also a necessity in order to resolve individual star-forming regions and allow for a detailed spiral structure.



Eris N-body simulation (Guedes et al. 2011)

What have we learned?

1. Disk galaxies have flat rotation curves and exponential surface brightness profiles, which a successful model of galaxy formation should explain.
2. The key physical reason for the formation of disk galaxies is that radiative cooling is an isotropic process, thus the gas contracts while approximately conserving angular momentum resulting in extended disks.
3. Adiabatic contraction steepens significantly the inner density profiles resulting in peak rotation velocities, well above the virial velocity.
4. In order to produce exponential disks, the feedback processes must somehow leave the highest angular momentum disk material in the halo and eject preferentially low angular momentum material.