ASTR 610
Theory of Galaxy Formation

Lecture 18: Disk Galaxies

Frank van den Bosch
Yale University, Spring 2024

The Structure & Formation of Disk Galaxies

In this lecture we discuss the structure and formation of disk galaxies. After a very brief overview of some of the main observational properties of disk galaxies, we discuss the `standard model' for disk galaxy formation. We discuss some successes and failures of this model, and the implications.

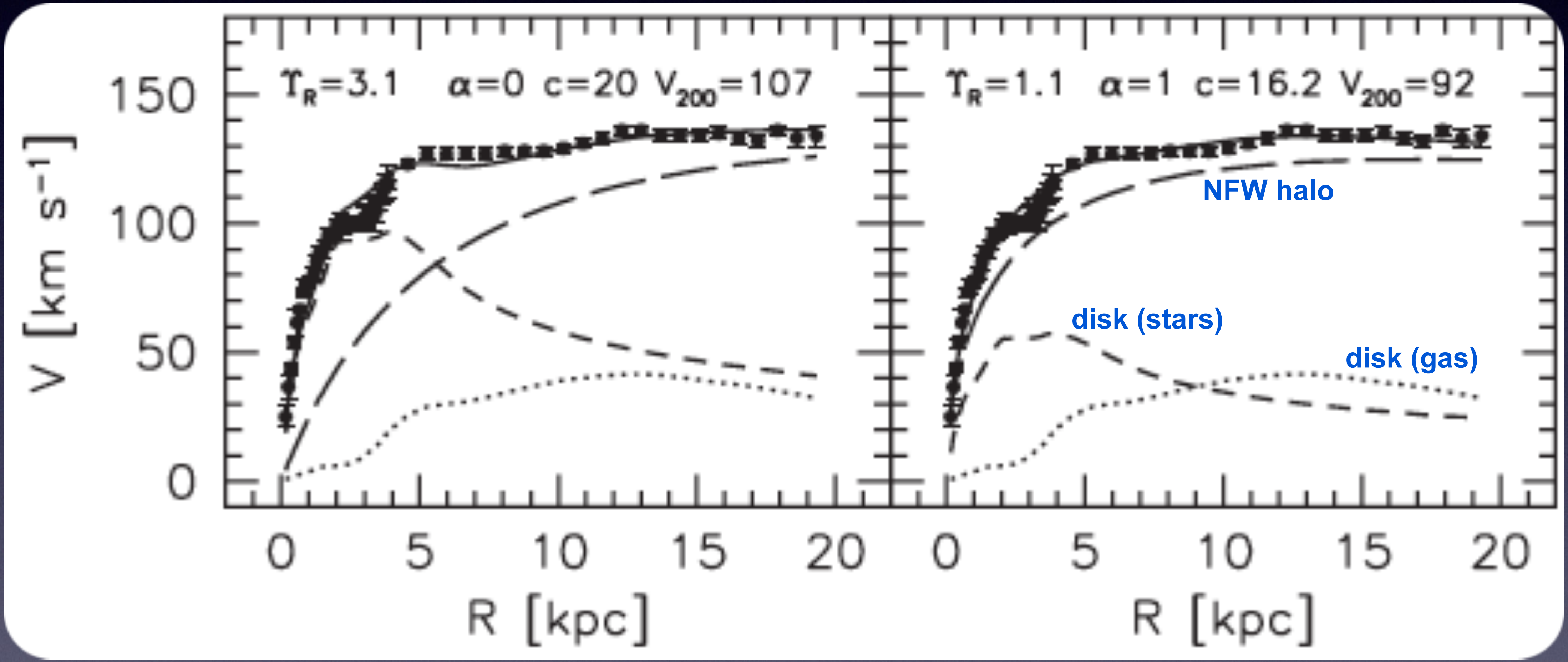
Topics that will be covered include:

- Exponential Disks
- Tully-Fisher relation
- Angular Momemtum Catastrophe
- Viscosity & Resonant scattering
- Adiabatic Contraction
- Angular Momentum Distributions

Observational Facts

Flat Rotation Curves

Disk galaxies have flat rotation curves. Unfortunately, it is difficult to obtain unique disk-halo(-bulge) decompositions....

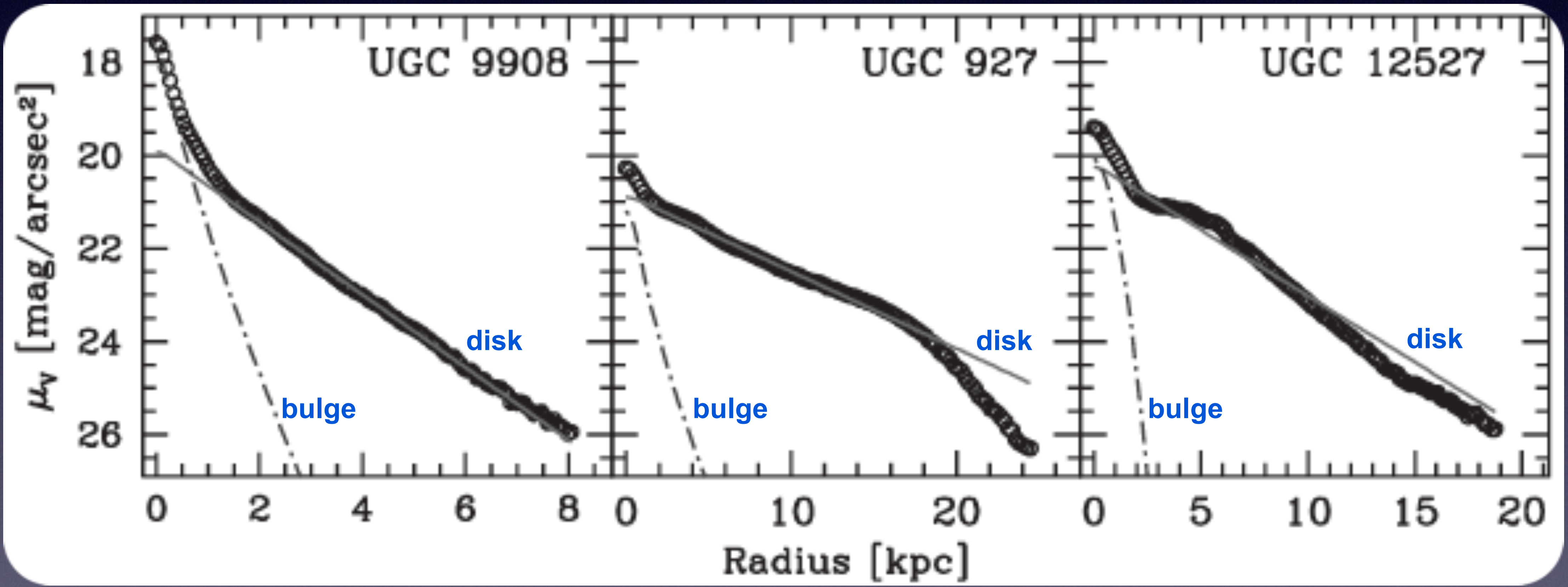


A frequently used decomposition is the one that maximizes the contribution due to the stellar disk (i.e., with a maximum M/L ratio for the stars). This is called a `maximal disk' decomposition....

Observational Facts

Exponential Surface Brightness Profiles

- Disk galaxies have surface brightness profiles that often are close to exponential.
- Deviations from exponential at small radii are attributed to bulge and/or bar.
- Deviations from exponential at large radii are attributed to star formation thresholds, radial migration, and/or maximum angular momentum...



Assuming that the disk is intrinsically round, and infinitesimally flat, the inclination angle follows from $\cos i = b/a$, with a and b the semi-major and semi-minor axes. Hence, face-on and edge-on correspond to $i=0^{o}$ and $i=90^{o}$, respectively.

Observational Facts

Exponential Surface Brightness Profiles

Because of their close-to-exponential appearance, disk galaxies are often modelled as infinitesimally thin, exponential disks:

surface brightness $I(R) = I_0\, \mathrm{e}^{-R/R_\mathrm{d}}$ $\quad L_\mathrm{d} = 2\pi \int_0^\infty I(R)\, R\, \mathrm{d}R = 2\pi\, I_0\, R_\mathrm{d}^2$

surface mass density $\Sigma(R) = \Sigma_0\, \mathrm{e}^{-R/R_\mathrm{d}}$ $\quad M_\mathrm{d} = 2\pi \int_0^\infty \Sigma(R)\, R\, \mathrm{d}R = 2\pi\, \Sigma_0\, R_\mathrm{d}^2$

circular velocity $V_{\mathrm{c,d}}^2(R) = -4\pi G \Sigma_0\, R_\mathrm{d}^2\, y\, [I_0(y)\, K_0(y) - I_1(y)\, K_1(y)]$

disk scale length R_d stellar mass-to-light ratio $M_\mathrm{d}/L_\mathrm{d} = \Sigma_0/I_0$

$y \equiv R/(2\, R_\mathrm{d})$ modified Bessel functions $I_n(x)$ $K_n(x)$

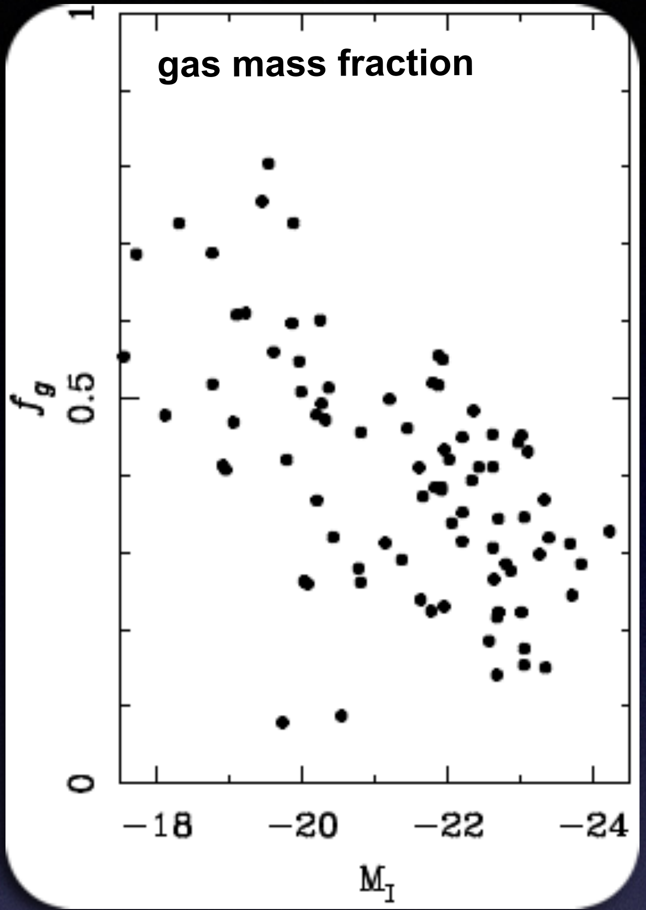
The circular velocity curve reaches a maximum at $R \simeq 2.16\, R_\mathrm{d}$

For more realistic models, with non-zero thickness, see MBW §11.1.1...

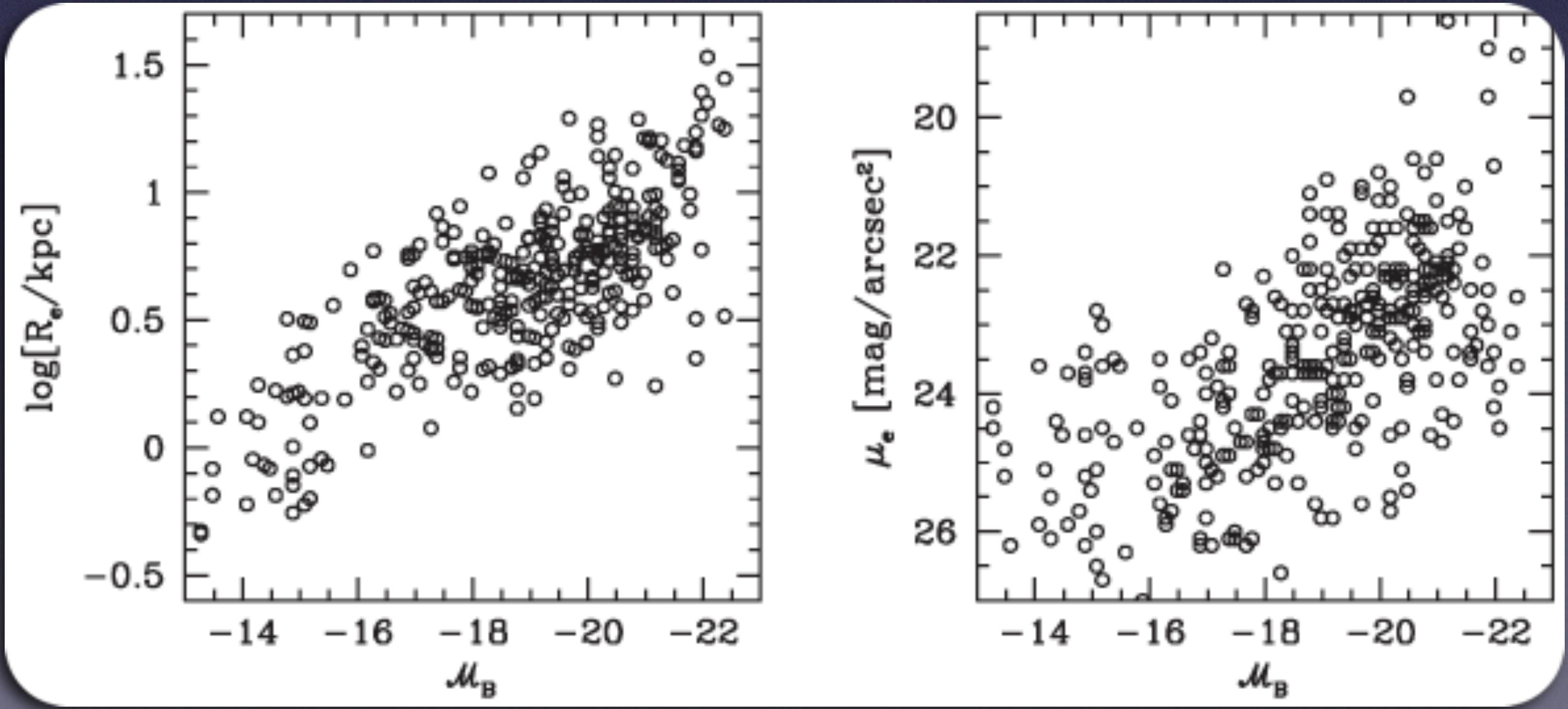
Scaling Relations

Brighter disks

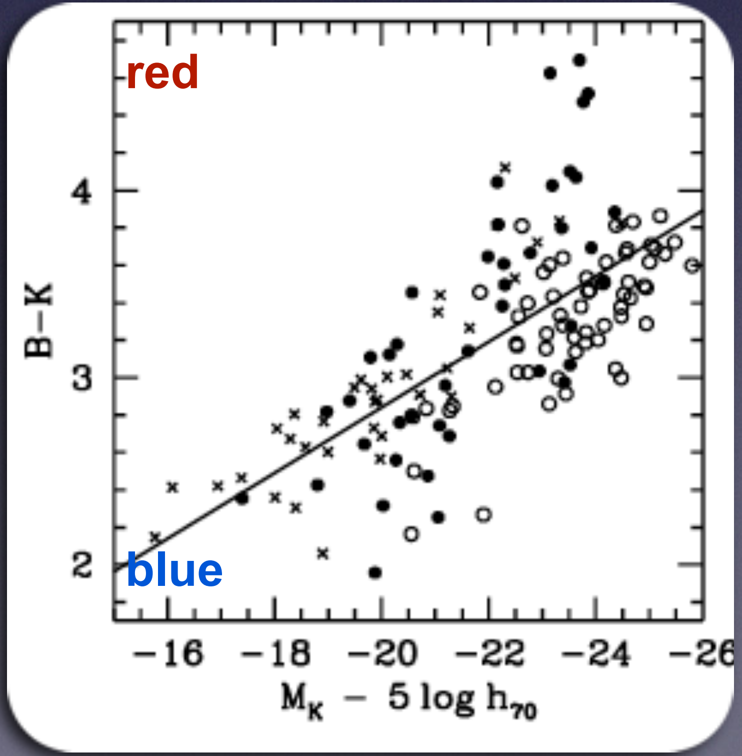
- are larger
- are redder
- have higher central SB
- have smaller gas mass fractions
- rotate faster (Tully-Fisher relation)



McGaugh & de Blok, 1997, ApJ, 534, 146



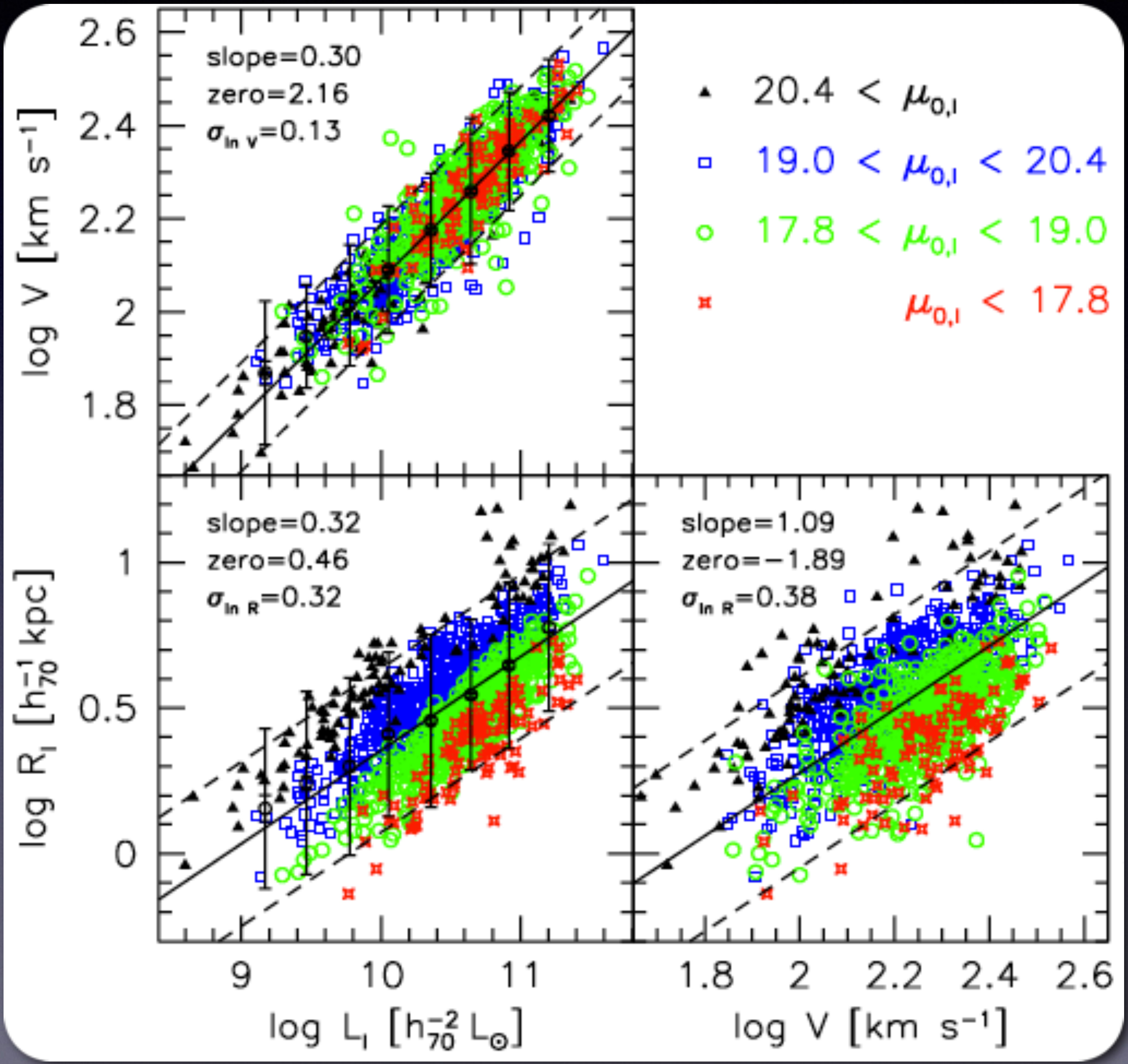
MBW, Fig. 2.20



vdBosch & Dalcanton, 2000, ApJ, 534, 146

Observational Facts

Scaling Relations



Source: Courteau et al. 2007, ApJ, 671. 203

Tully-Fisher (TF) relation:

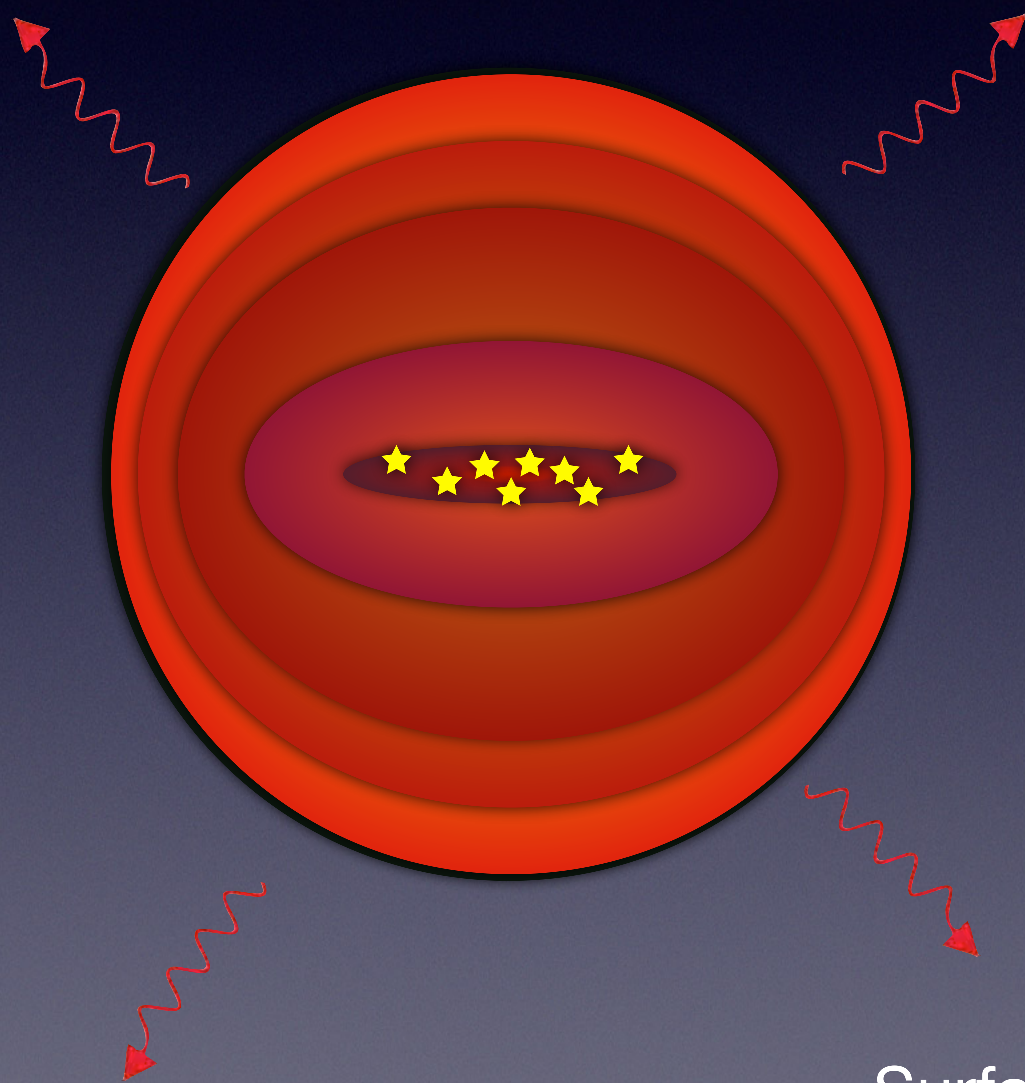
$$V \propto L_I^{0.30}$$

The slope of the TF relation depends on photometric band. It typically gets larger for bluer bands.

Getting models and simulations to reproduce the zero-point of the TF relation is a challenging problem that is still not entirely solved...

The scatter in TF relation is NOT correlated with surface brightness. As we will see, this gives important insight into the origin of the TF relation.

The Formation of Disk Galaxies

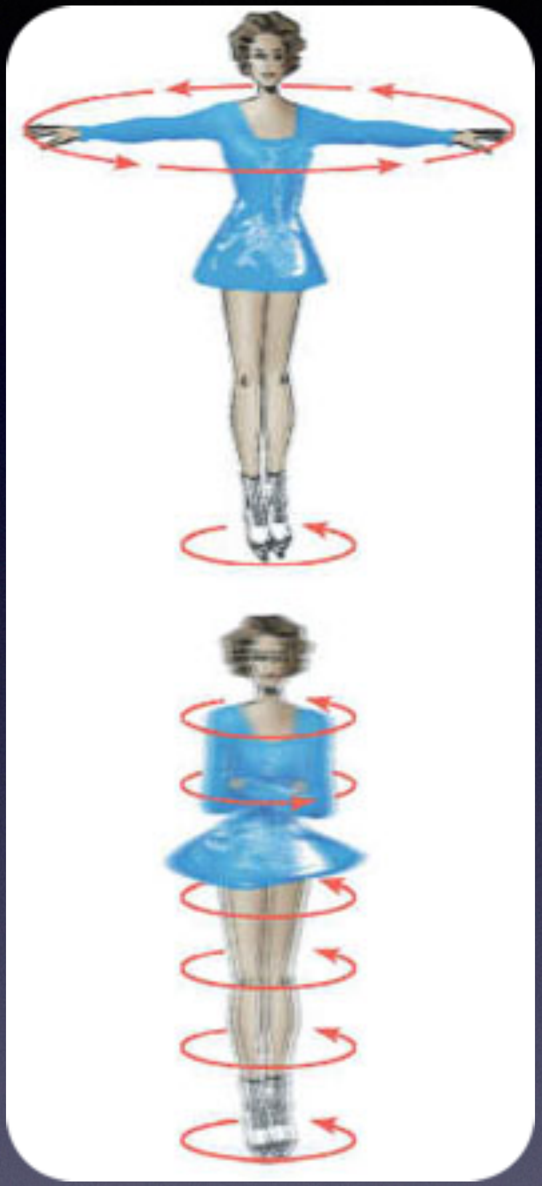


Hot (shock-heated) gas inside extended dark matter halo cools radiatively,

As gas cools, its pressure decreases causing the gas to contract

Since emission of photons is isotropic, angular momentum of cooling gas is conserved.

As gas sphere contracts, it spins up, and flattens



Surface density of disk increases, `triggering' star formation; a disk galaxy is born...

The Formation of Disk Galaxies

The Standard Picture

Disk galaxies are in centrifugal equilibrium; their structure is therefore governed by their (specific) angular momentum distribution.

The standard picture of disk formation, which was introduced in a seminal paper by Fall & Efstathiou (1980), is based on the following three "assumptions":

- the angular momentum originates from cosmological torques
- baryons & dark matter acquire identical specific angular momentum distributions
- baryons conserve their specific angular momentum while cooling

This standard picture was `modernized' in another seminal paper, by Mo, Mao & White (1998), and has subsequently been extended/revised by numerous studies (e.g., van den Bosch 1998, 2000, 2002; Dutton et al .2007; Dutton & van den Bosch 2012).

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 given by

$$J = \delta M \, (G M R)^{1/2}$$

where M is the mass of the black hole.

Clearly, this is very different from a realistic disk galaxy, whose mass distribution is close to exponential.....

Reason for this paradox is that although lowest energy state is preferred, its realization requires very efficient transport of angular momentum from inside out.

There are several mechanisms that can cause such angular momentum transport:

- **secular evolution** → viscosity (gas only) (Lin & Pringle 1987); resonant scattering (stars & gas) (Sellwood & Binney 2002)
- **hierarchical formation** → dynamical friction & ram pressure acting on gas clouds → **"angular momentum catastrophe"**

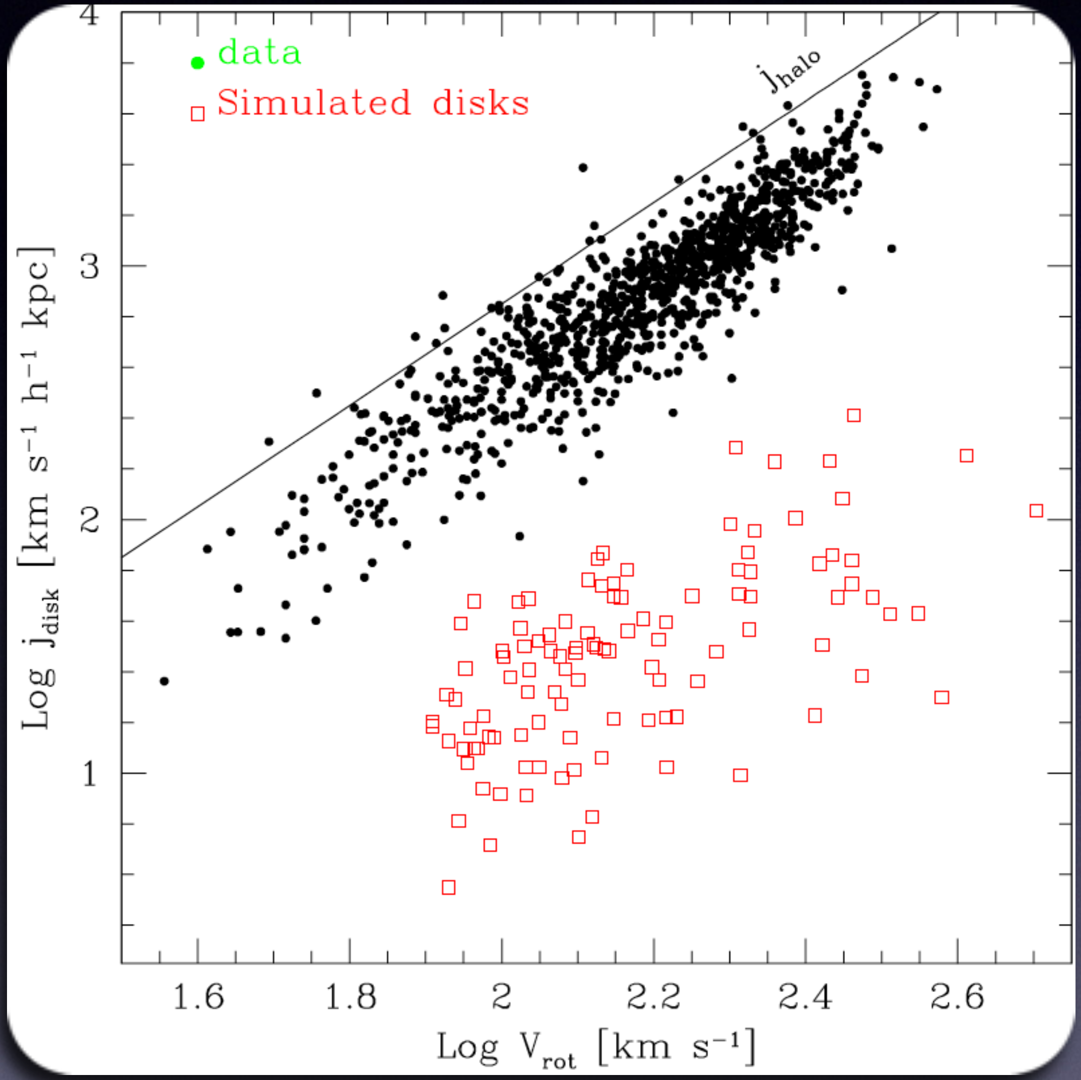
Angular Momentum Catastrophe

The fact that disk galaxies have a structure that deviates strongly from their minimal energy state indicates that these processes are not very efficient.

In simulations, however, efficient cooling causes much gas to condense into clumps at centers of subhaloes. This gas is delivered to disk via dynamical friction, transferring (orbital) angular momentum of gas to dark matter ➡ disks end-up much too small. This problem is known as the

angular momentum catastrophe

and requires feedback to prevent most of the gas from cooling in small haloes.



Source: Steinmetz & Navarro, 1999, ApJ, 513, 555

Although the secular processes are not very efficient, they cannot be neglected. Radial migration of stars due to resonant scattering of spirals (and bars) can cause significant redistribution of angular momentum within disk (e.g., Roškar et al. 2008)

Disk Formation: The Standard Picture

As a starting point, consider the following idealized case:

- self-gravity of the disk can be ignored
- dark matter halo is a singular, isothermal sphere $\rho(r) = V_{\rm vir}^2/(4\pi G r^2)$

Using the virial scaling relations between virial mass, virial radius and virial velocity of dark matter haloes (see lecture 11), and assuming that the mass that settles in the disk is a fraction m_d of that of the total virial mass, we have that

$$M_{\rm d} = m_{\rm d}\, M_{\rm vir} \simeq 1.3 \times 10^{11} h^{-1} M_\odot \left(\frac{m_{\rm d}}{0.05}\right) \left(\frac{V_{\rm vir}}{200\,{\rm km/s}}\right)^3 \mathcal{D}^{-1}(z)$$

where $\mathcal{D}(z) = \left[\frac{\Delta_{\rm vir}(z)}{100}\right]^{1/2} \left[\frac{H(z)}{H_0}\right]$ accounts for the redshift dependence.

Under the assumption that the disk is an infinitesimally thin, exponential,

$$J_{\rm d} = 2\pi \int_0^\infty \Sigma(R)\, V_{\rm c}(R)\, R^2\, {\rm d}R = 2 M_{\rm d}\, R_{\rm d}\, V_{\rm vir}$$

Disk Formation: The Standard Picture

If we define the parameter $j_{\rm d}$ via $J_{\rm d} \equiv j_{\rm d}\, J_{\rm vir}$, where $J_{\rm vir}$ is the angular momentum of the dark matter halo, then we can related $J_{\rm d}$ to the spin parameter of the dark matter halo, according to

$$\lambda = \frac{J_{\rm vir}\, |E|^{1/2}}{G\, M_{\rm vir}^{5/2}} = \frac{1}{j_{\rm d}} \frac{J_{\rm d}\, |E|^{1/2}}{G\, M_{\rm vir}^{5/2}}$$

Using the virial theorem, according to which $E = -K = -\frac{M_{\rm vir}\, V_{\rm vir}^2}{2}$, we have that

$$J_{\rm d} = \sqrt{2}\, j_{\rm d}\, \lambda\, M_{\rm vir}\, R_{\rm vir}\, V_{\rm vir}$$

$$J_{\rm d} = 2 M_{\rm d}\, R_{\rm d}\, V_{\rm vir}$$

$$M_{\rm d} = m_{\rm d}\, M_{\rm vir}$$

$$R_{\rm d} = \frac{1}{\sqrt{2}}\, \lambda \left(\frac{j_{\rm d}}{m_{\rm d}}\right) R_{\rm vir}$$

Using the virial scaling relations for dark matter haloes this yields

$$R_{\rm d} \simeq 10 h^{-1}{\rm kpc} \left(\frac{j_{\rm d}}{m_{\rm d}}\right) \left(\frac{\lambda}{0.05}\right) \left(\frac{V_{\rm vir}}{200\,{\rm km/s}}\right) \mathcal{D}^{-1}(z)$$

Disk Formation: The Standard Picture

$$R_{\rm d} \simeq 10 h^{-1}{\rm kpc} \left(\frac{j_{\rm d}}{m_{\rm d}}\right) \left(\frac{\lambda}{0.05}\right) \left(\frac{V_{\rm vir}}{200\,{\rm km/s}}\right) \mathcal{D}^{-1}(z)$$

Consider the MW: Assume that $V_{\rm vir} = V_{\rm rot} \simeq 220\,{\rm km/s}$ and that $j_{\rm d} = m_{\rm d}$
Then, using that $M_{\rm d} \simeq 5 \times 10^{10}\, M_{\odot}$ and $R_{\rm d} \simeq 3.5\,{\rm kpc}$ the above relation implies

$m_{\rm d} \sim 0.01$	⇒	$f_{\rm bar} \simeq 0.17$	only ~6% of baryons are in disk
$\lambda \sim 0.011$		$P(\lambda < 0.011) \simeq 3\%$	MW halo is rare...

Alternatively, if we assume that $\lambda = \bar{\lambda} \simeq 0.04$ we infer that $j_{\rm d} \simeq 0.3 m_{\rm d}$, which implies that disk has to grow preferentially out of low angular momentum material.

However, before drawing any conclusions, let's first consider a more realistic case:

- assume dark matter haloes follow NFW profile
- take self-gravity of disk into account

Disk Formation: The Standard Picture

As shown by Mo, Mao & White (1998; hereafter MMW), in this case one has that

$$R_{\rm d} = \frac{1}{\sqrt{2}} \lambda \left(\frac{j_{\rm d}}{m_{\rm d}}\right) R_{\rm vir} F_R^{-1} F_E^{-1/2}$$

new addition

F_E is defined by $$E = -\frac{M_{\rm vir} V_{\rm vir}^2}{2} F_E$$

F_R is defined by $$F_R = \frac{1}{2} \int_0^{R_{\rm vir}/R_{\rm d}} u^2 \, {\rm e}^{-u} \frac{V_{\rm c}(uR_{\rm d})}{V_{\rm vir}} {\rm d}u$$

see MMW for simple fitting functions for both F_E and F_R....

$$V_{\rm c}^2(R) = V_{\rm c,d}^2(R) + V_{\rm c,h}^2(R) = V_{\rm c,d}^2(R) + \frac{G\, M_{\rm h,ac}(R)}{R}$$

contribution to rotation curve due to self-gravity of disk

circular velocity of the dark matter halo, corrected for adiabatic contraction due to disk formation

With this modification, the MW can be reproduced with $\lambda \simeq m_{\rm d} \simeq 0.05$ in much better agreement with expectations

Adiabatic Contraction

When baryons cool and concentrate in the center of a dark matter halo, the halo structure will be modified due to the gravitational action of the baryons.

In general, it is difficult to model this action of disk on halo accurately, because it will depend on the detailed formation history of the disk-halo system.

However, if growth of disk is slow compared to dynamical time of dark matter particles in (center) of halo, the system adjusts itself adiabatically (reversable): in this case the final state is independent of the path taken.

If (i) the system is spherically symmetric, and (ii) all particles are on circular orbits, adiabatic invariance implies that the quantity r M(r) is conserved

$$r_{\rm f}\, M_{\rm f}(r_{\rm f}) = r_{\rm i}\, M_{\rm i}(r_{\rm i})$$

Here $M_{\rm i}(r)$ and $M_{\rm f}(r)$ are the initial and final mass profiles

In order to model adiabatic contraction, the above equation is often used, assuming that initially baryons and dark matter follow the same NFW profile, while after disk formation

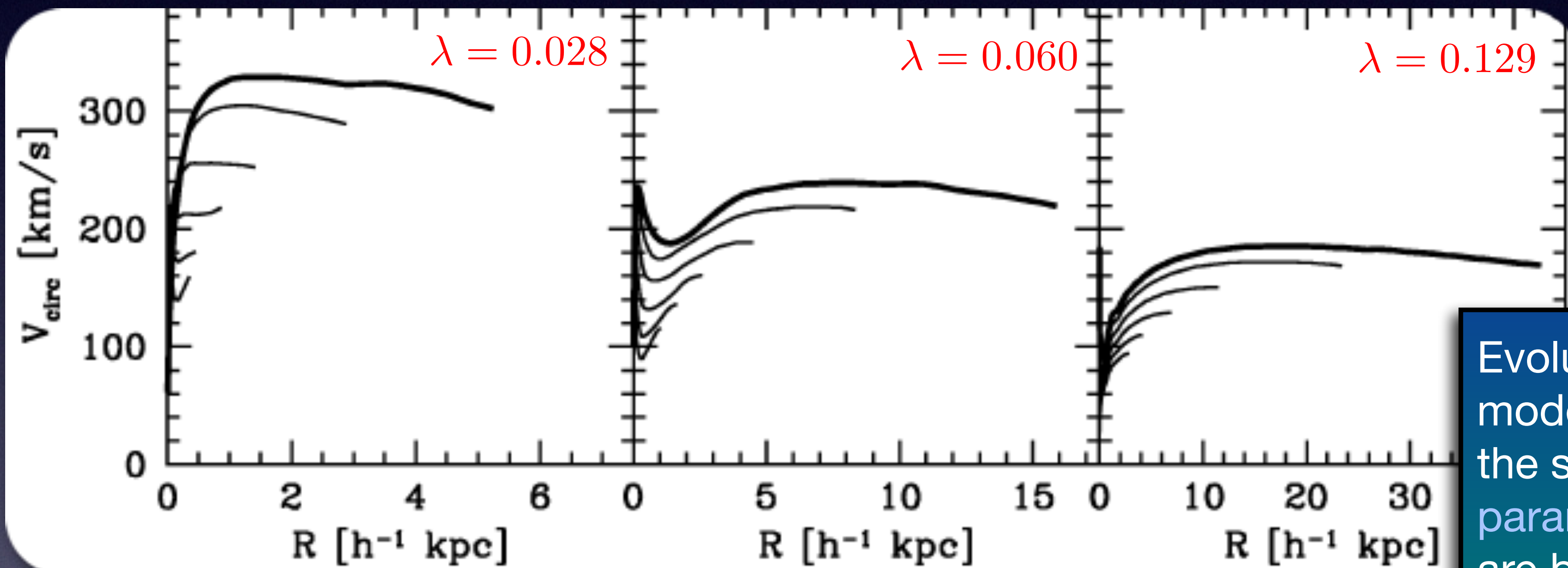
$$M_{\rm f}(r_{\rm f}) = M_{\rm d}(r_{\rm f}) + [1 - m_{\rm d}]\, M_{\rm i}(r_{\rm i})$$

Adiabatic Contractoin

For given m_d and $M_d(r)$ the above equations can be solved (iteratively) for r_f given r_i.

For realistic values of m_d, the effect of AC can be very substantial, resulting in $V_{rot}/V_{vir} >> 1$, where V_{rot} is the flat part of the rotation curve.
In fact, since smaller spin parameters result in more compact disks, the actual ratio V_{rot}/V_{vir} is a strong function of λ: i.e., V_{rot} is a poor indicator of V_{vir}!!!

van den Bosch, 2002, MNRAS, 332, 456



Evolution of RCs in disk formation models. All three disks reside in haloes of the same mass, but with different spin parameters. Self-gravity of disk and AC are both taken into account

Many studies in the literature assume that $V_{rot} = V_{vir}$ or $V_{rot} = V_{max}$, where the latter is the maximum circular velocity of the NFW halo (typically $V_{max}/V_{vir} \sim 1.2$)

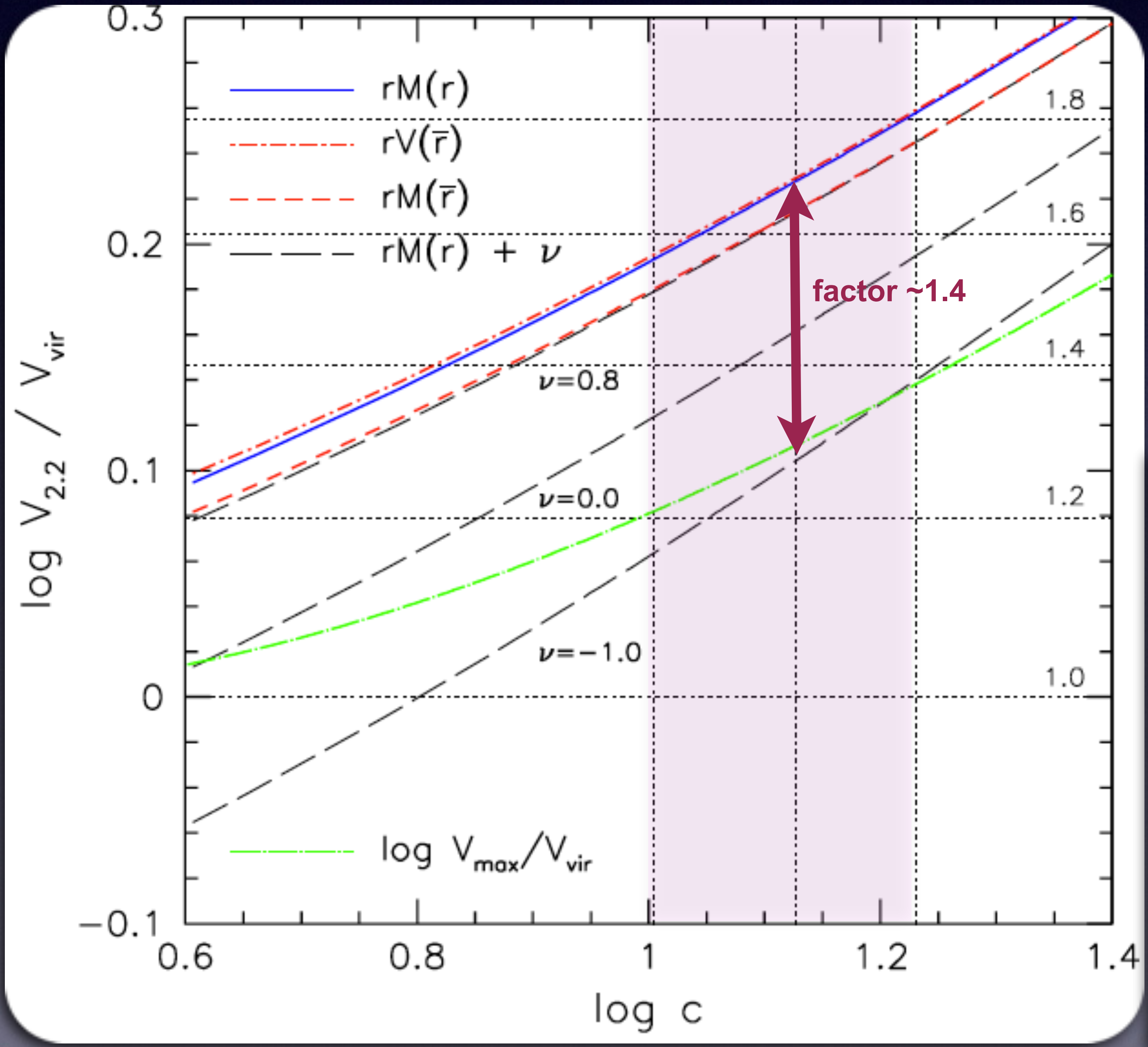
However, with self-gravity of disk and AC, one typically expects $V_{rot}/V_{vir} \sim 1.4$-1.8)

Adiabatic Contraction

Note: r M(r) is only an adiabatic invariant for spherical systems with circular orbits. In reality disk growth results in halo contraction that is slightly different. This can be parameterized by stating that $r_{\rm f} = \Gamma^{\nu} r_{\rm i}$, where $\Gamma = r_{\rm f}/r_{\rm i}$ in the simplified AC case discussed above, and ν is a `free parameter'.

Simulations suggest that $\nu \simeq 0.8$

- $\nu = 1$ standard AC
- $\nu = 0$ no adiabatic contraction
- $\nu < 0$ halo expansion



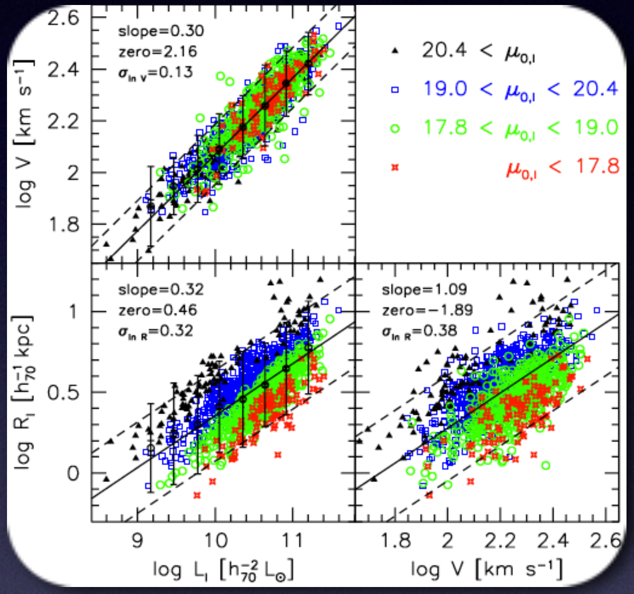
Dutton, vdBosch, Dekel & Courteau, 2007, ApJ, 654, 27

Note how even in the case without AC ($\nu = 0$) $V_{2.2} > V_{max}$, which is simply due to the self-gravity of the disk

Model predictions for the ratio $V_{2.2}/V_{vir}$ for disk galaxies embedded in NFW haloes. Results are shown as function of the halo concentration parameter, c. Here $V_{2.2}$ is the rotation velocity of the disk measured at 2.2 disk scale-lengths. Results are shown for different `forms' of adiabatic contraction (different values of ν). For comparison, the green curve shows the ratio V_{max}/V_{vir}. All models use $m_d = \lambda = 0.05$

Lessons Learned

Dutton et al. (2007) used an extended version of the MMW models to investigate what it takes for the models to reproduce the observational V-L-R scaling relations:



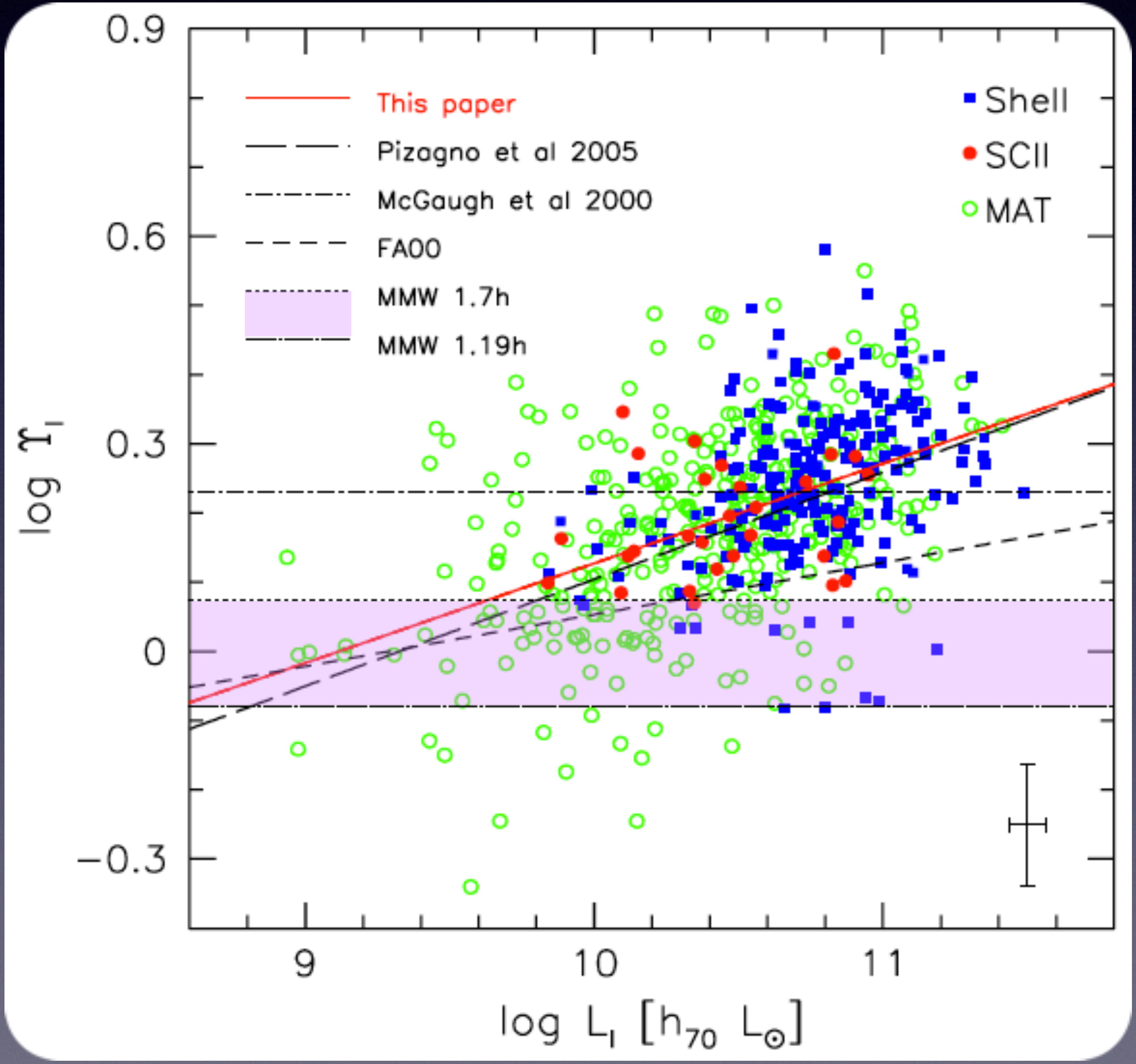
Model Ingredients

- Exponential disks in contracted NFW haloes.
- AC modeled, treating ν as a free parameter.
- Disk mass fraction modelled as $m_{\rm d} = m_{\rm d,0}\left(\frac{M_{\rm vir}}{10^{12}h^{-1}M_\odot}\right)^\alpha$ with α and $m_{\rm d,0}$ two additional free parameter
- Bulge formation included as described in MBW §11.2.4
- Disk split in stars and (cold) gas using SF threshold density; material with $\Sigma(R) > \Sigma_{\rm crit}(R)$ is turned into stars. This adds one more free parameter, the Toomre Q-parameter.
- The spin parameters of dark matter haloes follow a log-normal distribution with $\bar{\lambda}_{\rm DM} = 0.04$ and $\sigma_{\ln\lambda} = 0.5$
- We assume that $\lambda_{\rm gal} = \lambda_{\rm DM}$, i.e., that $j_{\rm d}/m_{\rm d} = 1$

Stellar Mass-to-Light ratios

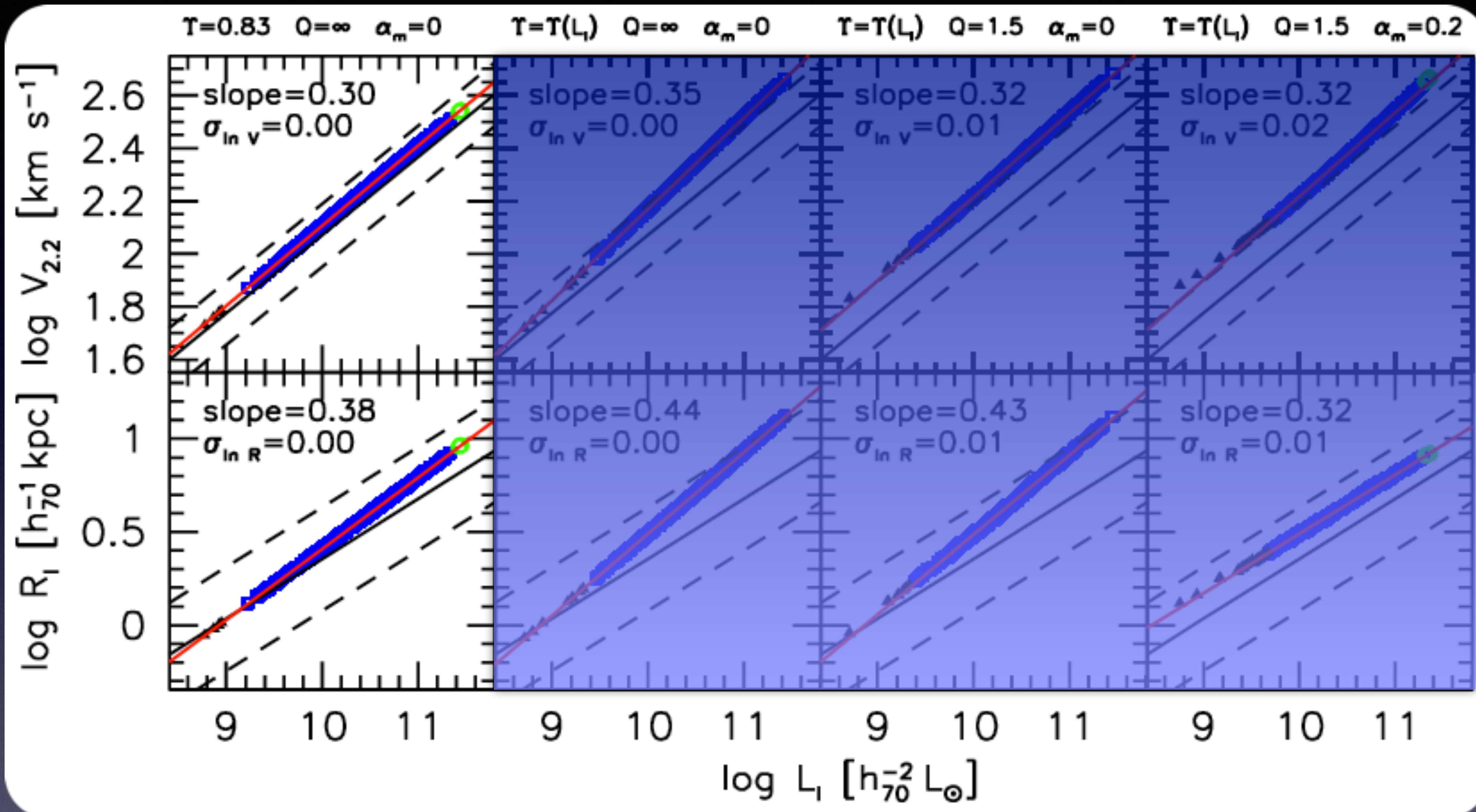
MMW models can fit V-L-R scaling relations, but they assumed unrealistically low values for stellar mass-to-light ratios, and assumed that disk was entirely stellar (no cold gas). Stellar population models show that stellar mass-to-light ratios of disk galaxies increase with luminosity, which will impact slopes of V-L-R relations....

Source: Dutton, vdBosch, Dekel & Courteau, 2007, ApJ, 654, 27



Dutton et al. (2007) took this into account by modelling the stellar mass-to-light ratios using the red line shown in the figure...

Models without scatter

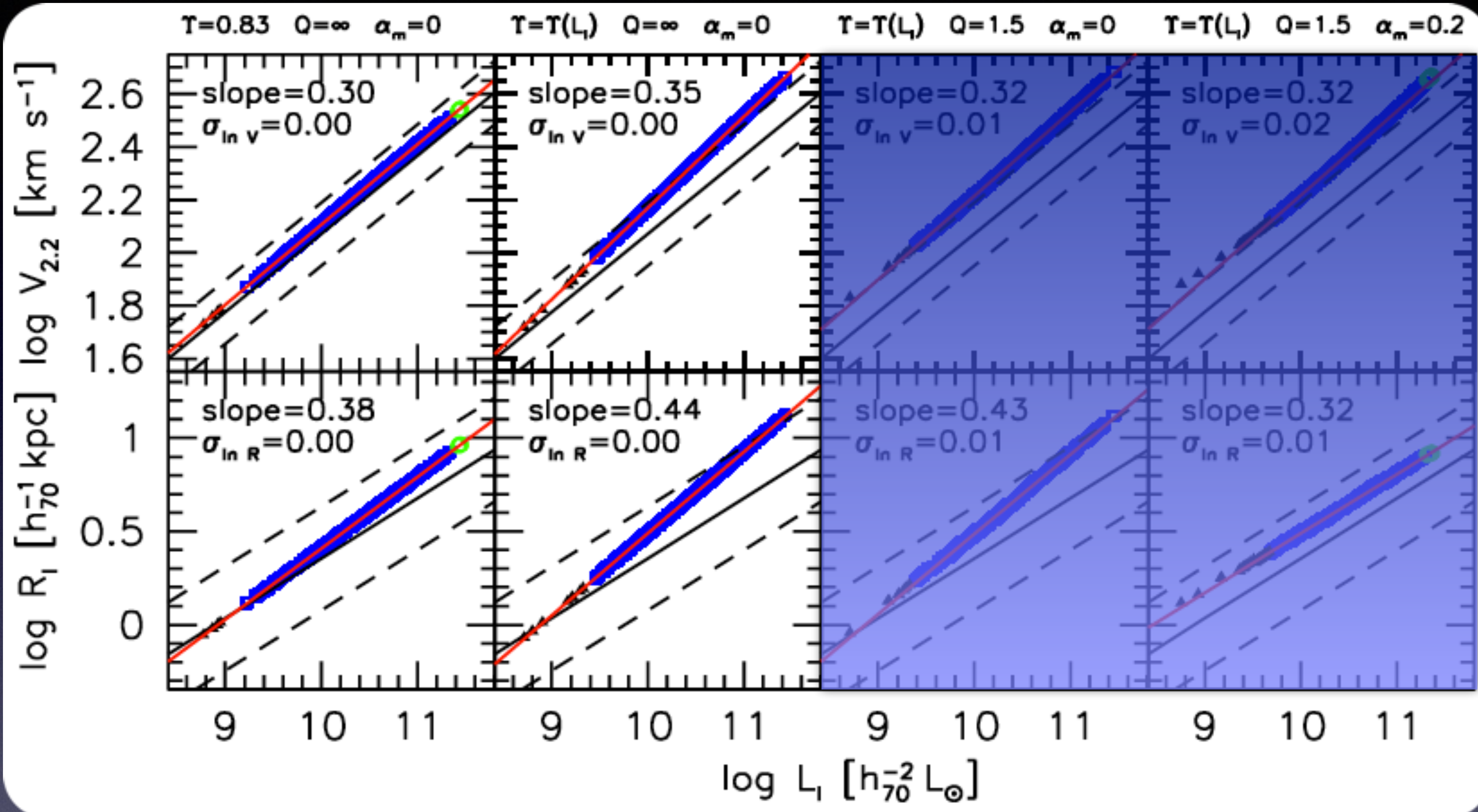


Source: Dutton, vdBosch, Dekel & Courteau, 2007, ApJ, 654, 27

The \`original' MMW model matches the slope and zeropoint of the TF relation, and even predicts roughly the correct size-luminosity relation.

However, it adopts unrealistic stellar mass-to-light ratios, and assumes that all the disk matter is in the form of stars (no cold gas).

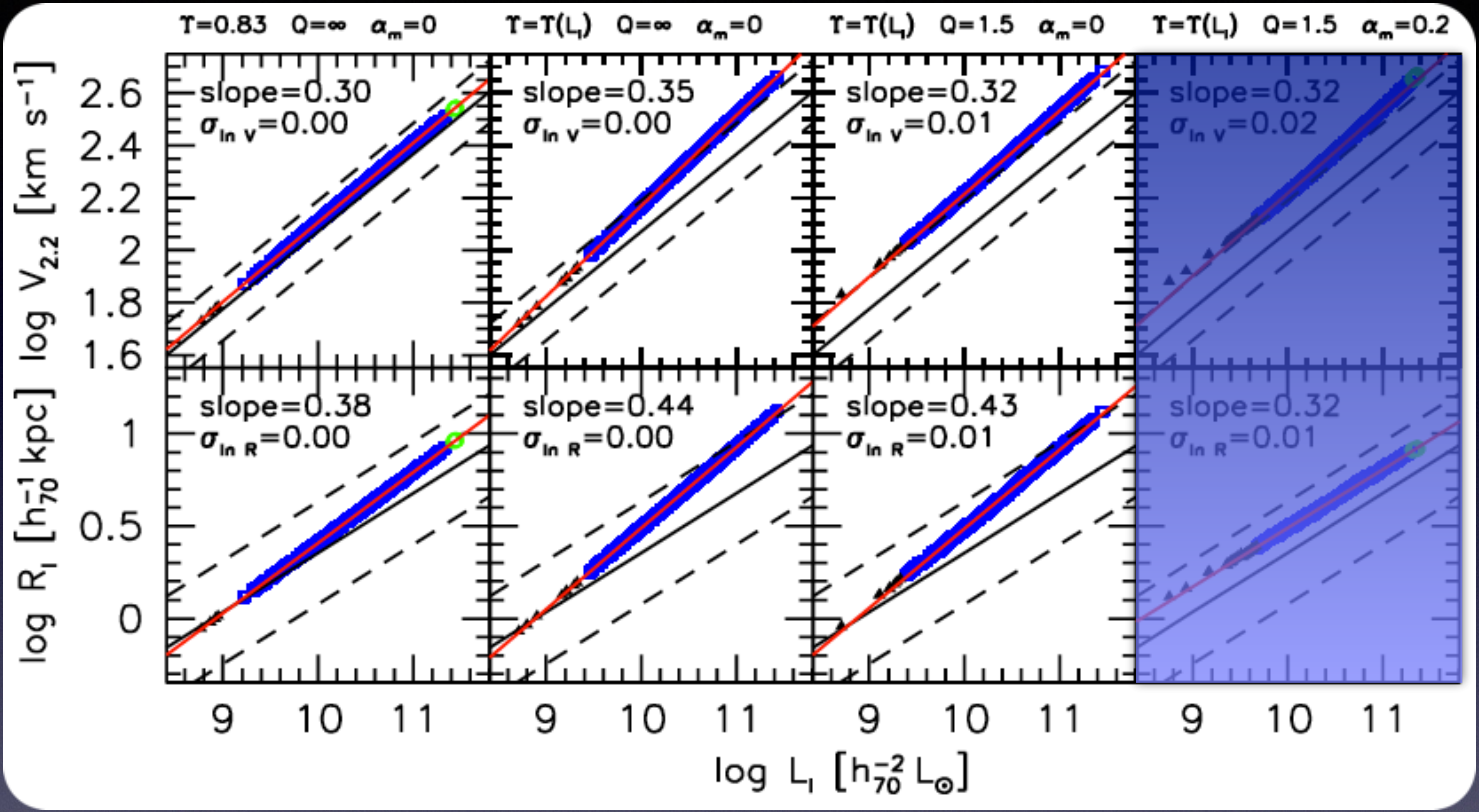
Models without scatter



Source: Dutton, vdBosch, Dekel & Courteau, 2007, ApJ, 654, 27

Including the empirical relation between stellar mass-to-light ratio and disk luminosity results in VL and RL relations that are too steep....

Models without scatter

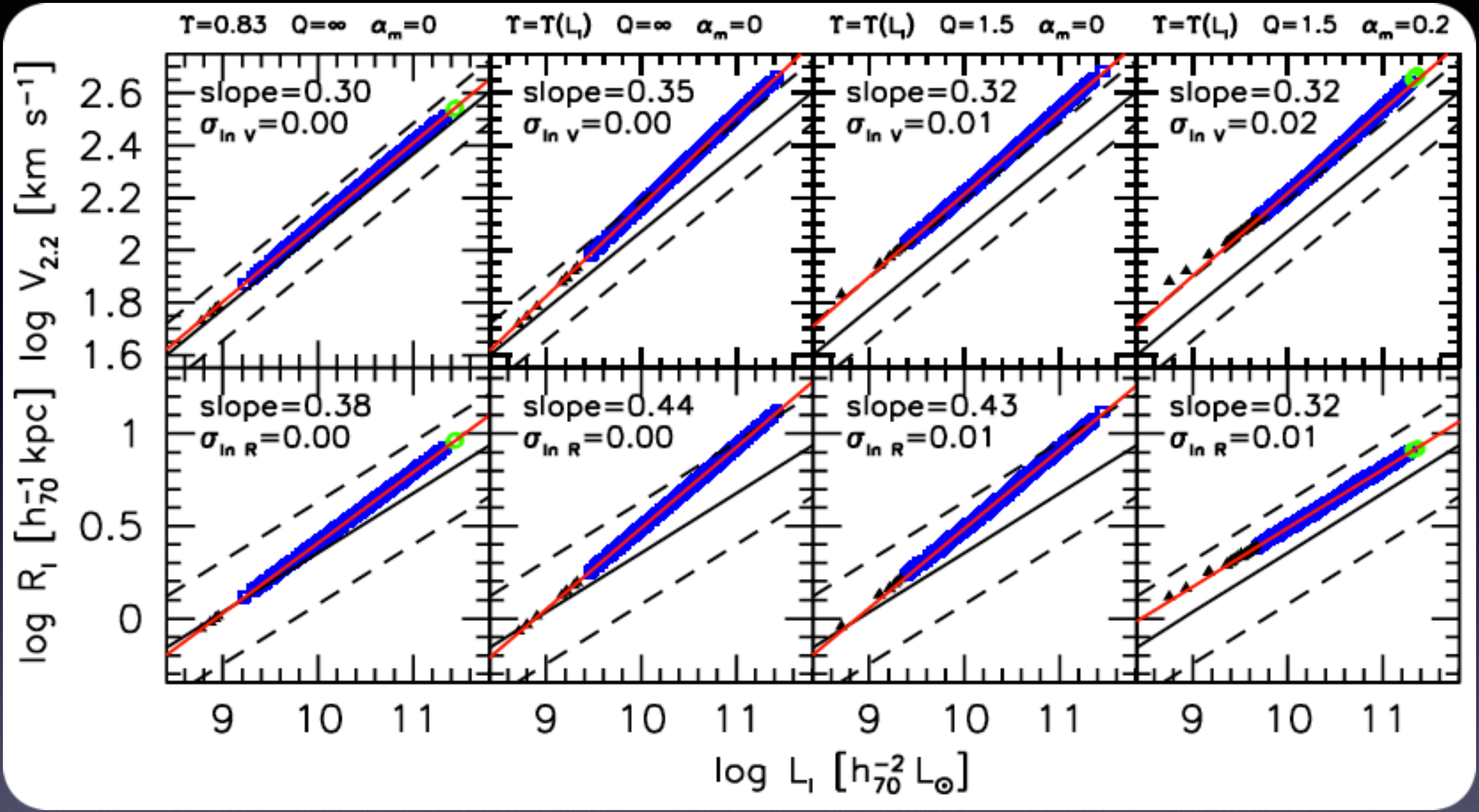


Source: Dutton, vdBosch, Dekel & Courteau, 2007, ApJ, 654, 27

Allowing only the disk matter with $\Sigma(R) > \Sigma_{\rm crit}(R)$ to form stars (the rest remains as cold gas), results in a TF relation with the correct slope, but with a zero-point that is ~2σ too high.

The RL relation still has a slope that is too steep...

Models without scatter



Source: Dutton, vdBosch, Dekel & Courteau, 2007, ApJ, 654, 27

One can adjust the disk-mass to halo-mass relation (i.e., α) in order to match the slope of the RL relation, but the zero-point of theTF relation is still ~2σ too high. Also, the disks are ~1σ too large...

The TF-zeropoint Problem

lower stellar mass-to-light ratios

required: $\log(M/L)_* \rightarrow \log(M/L)_* + \Delta_{\rm IMF}$ with $\Delta_{\rm IMF} \simeq -0.5$

problem: the most one can afford with changes in IMF is $\Delta_{\rm IMF} \sim -0.2$

lower halo concentrations

required: $c(M) \rightarrow \eta\, c(M)$ with $\eta \simeq 0.4$

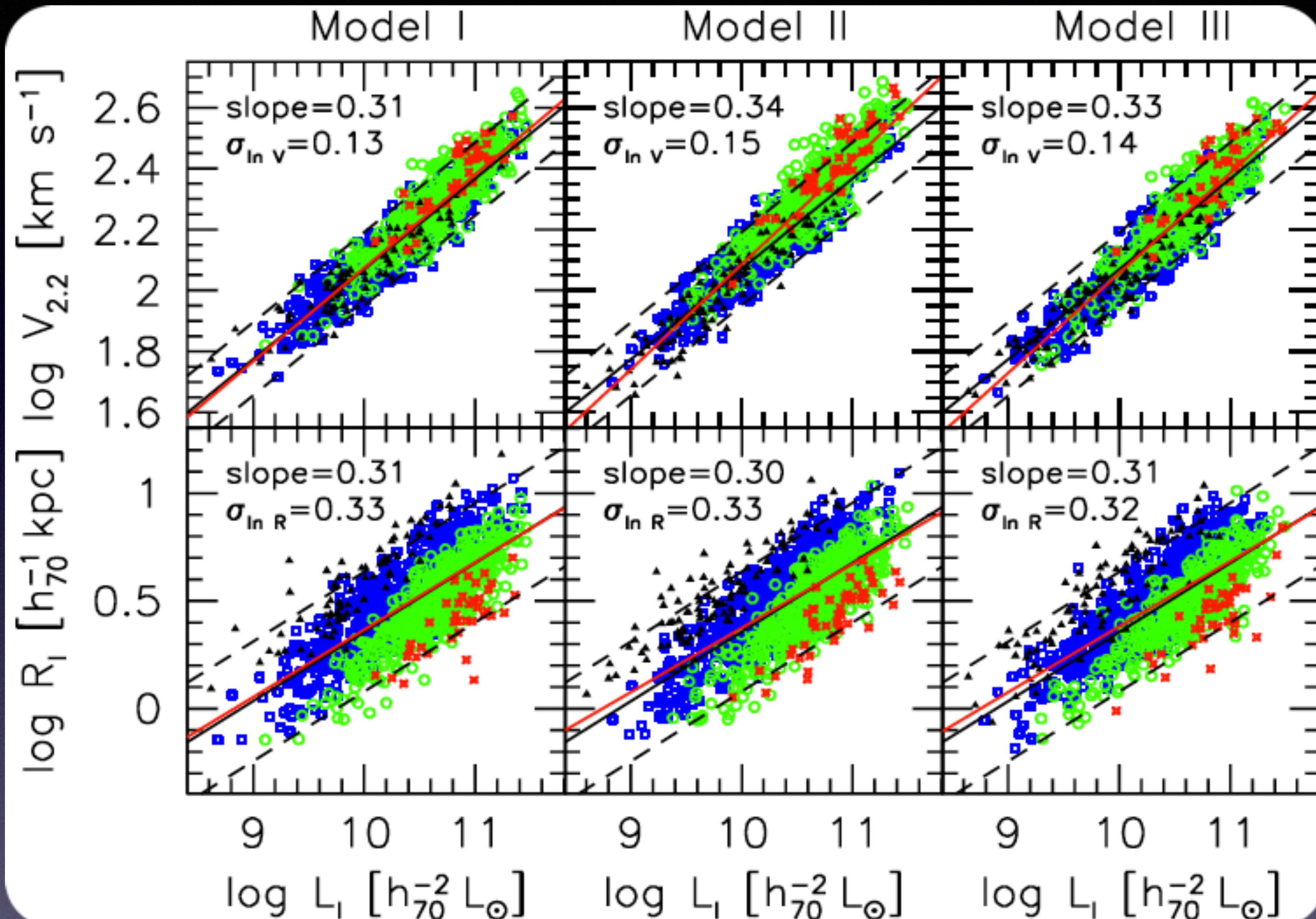
problem: halo concentrations depend on cosmology; the model of Dutton et al. assumed WMAP1 cosmology. For WMAP3 cosmology $\eta \simeq 0.75$. Having $\eta \simeq 0.4$ implies a cosmology that is violently inconsistent with current cosmological constraints.

modify adiabatic contraction

required: halo expansion, i.e., $\nu < 0$

problem: requires clumpy formation of disks; difficult to reconcile with their low-entropy nature. Solution may require mechanism(s) that can expell dark matter from center, such as the impulsive heating mechanism due to SN feedback discussed in Lecture 11.

Three Models with Scatter



Source: Dutton, vdBosch, Dekel & Courteau, 2007, ApJ, 654, 27

Model	ν	η	$\Delta_{\rm IMF}$
I	0.8	0.8	-0.4
II	1	0.5	-0.2
III	0	0.8	-0.2

AC parameter

c(M) parameter

All three models match slopes, zero-points and scatter of the observed TF and RL relations. However, the scatter in the observed RL relation requires $\sigma_{\ln\lambda} \leq 0.25$

For comparison, simulations predict that $\sigma_{\ln\lambda} \simeq 0.5$; Hence, disk galaxies may only form in a subset of all haloes, preferably those with smaller spin parameter (which are the ones that experienced a more quiescent formation history)...

The Origin of Exponential Disks

If there is no angular momentum loss (or redistribution) during the disk formation process, then the disk surface density profile is a direct reflection of the specific angular momentum distribution of the proto-galaxy.

$$\Sigma_{\rm d}(R) \longleftrightarrow M_{\rm bar}(j_{\rm bar}) \longleftrightarrow M_{\rm DM}(j_{\rm DM})$$

In Lecture 11 we have seen that the specific angular momentum distributions of dark matter haloes are well fit by the Universal profile:

$$\mathcal{P}(j) = \frac{\mu j_0}{(j + j_0)^2} \quad \Rightarrow \quad M(<j) = M_{\rm vir} \frac{\mu j}{(j + j_0)}$$

If picture of detailed specific angular momentum conservation is correct then

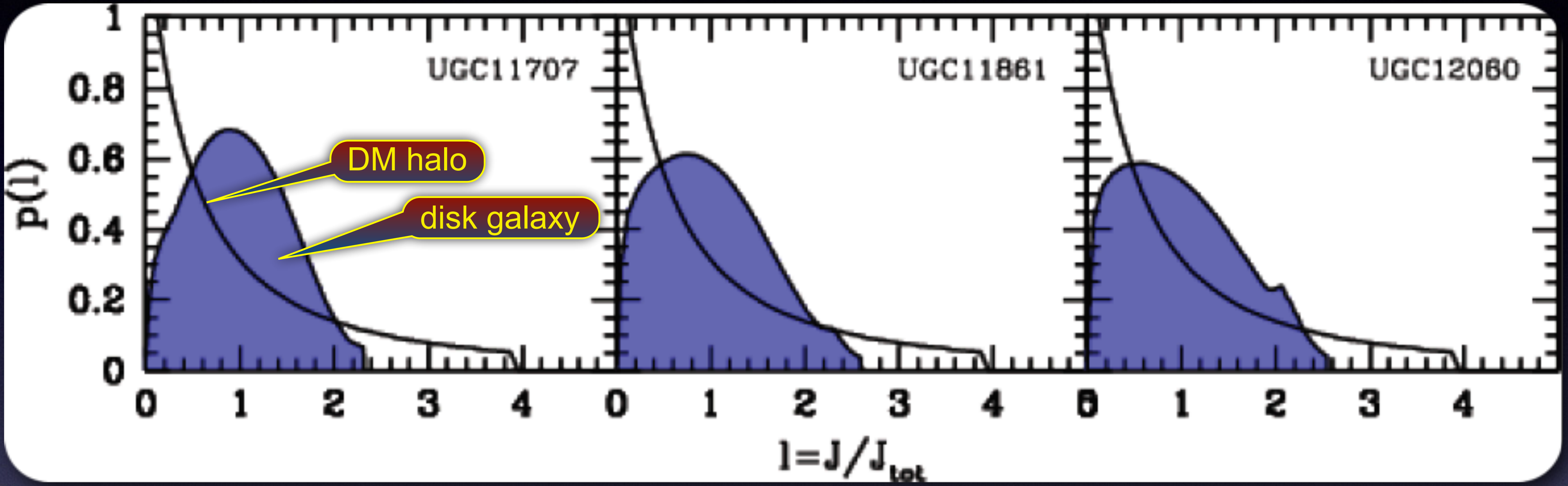
$$\frac{M_{\rm d}(<r)}{M_{\rm d}} = \frac{M_{\rm h}(<j)}{M_{\rm vir}}$$

where r and j are related according to $j = r\,V_{\rm c}(r)$

⇒ For given halo density and angular momentum profiles one can predict $\Sigma_{\rm d}(R)$

The Origin of Exponential Disks

Unfortunately, the disk surface density profiles thus predicted look nothing like the observed, exponential disks....



vdBosch et al. 2001, MNRAS, 326, 1205

Specific angular momentum distributions of dark matter haloes (solid lines) and observed disk galaxies (shaded curves). Former are averages obtained from numerical simulations, while latter derive from observed density profiles and rotation curves...

Observed disk galaxies lack both high and low specific angular momentum compared to predictions. This has important implications for disk formation:

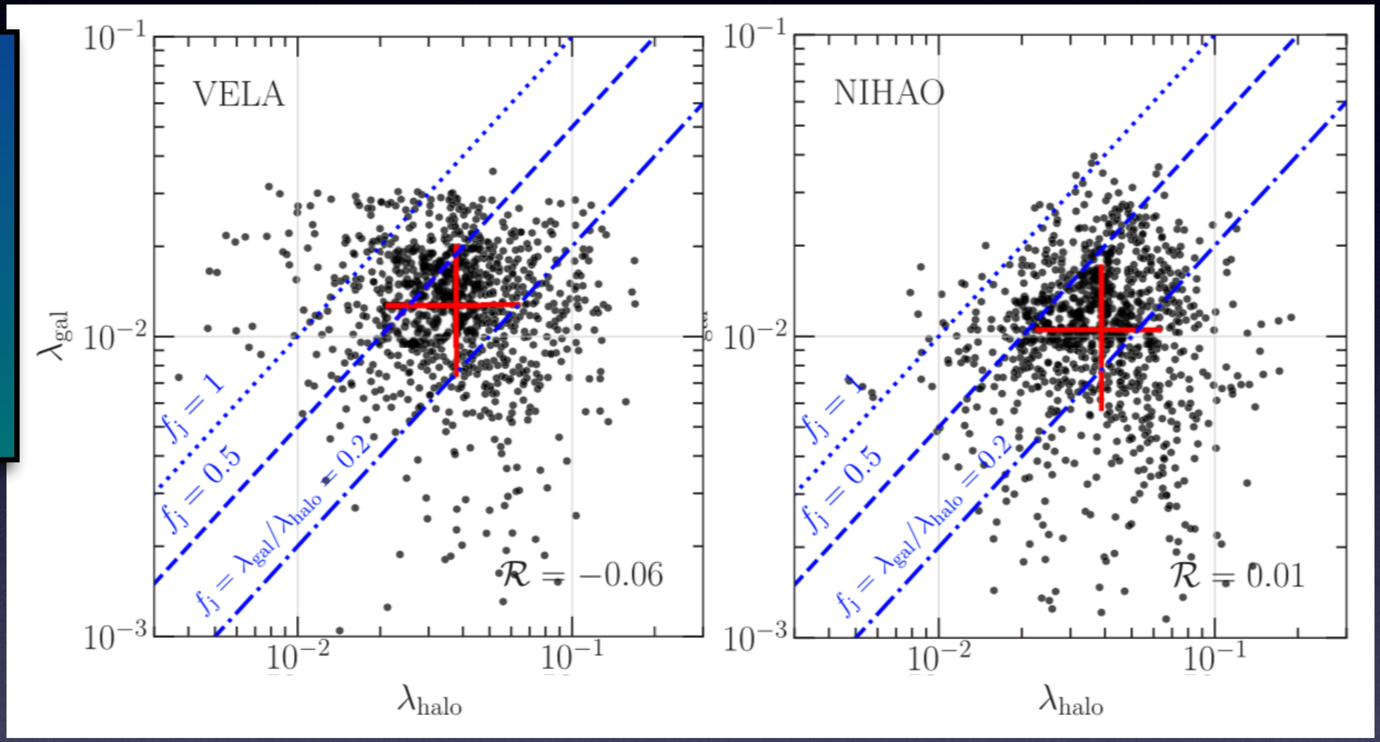
- inside-out formation leaves highest angular momentum material in halo
- feedback somehow preferentially ejects the low-angular momentum material

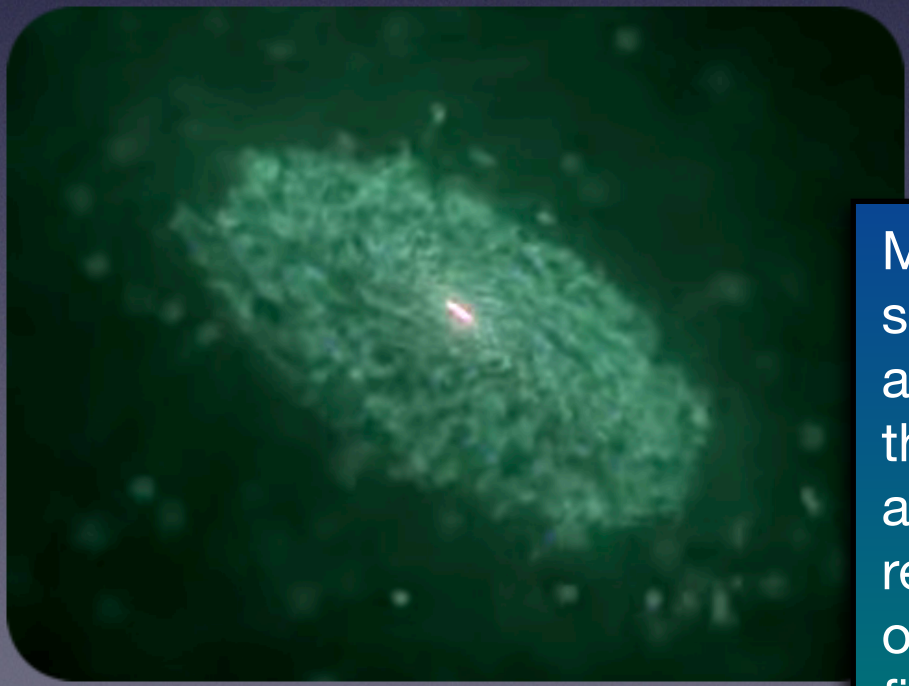
The Origin of Angular Momentum

Although the Fall & Efstathiou (1980) + MMW model for disk formation has proven able to explain observed galaxy scaling relations (albeit after considerable tuning as demonstrated in Dutton+07), recent advances suggest that the crucial assumption nunderlying these models (i.e., disk angular momentum ⟺ halo spin parameter) is likely to be incorrect.

Correlation between spin parameter of disk galaxies and that of their halos in two different suites of hydrodynamical zoom simulations. Note the utter lack of correlation, which is inconsistent with the `standard' picture of disk formation.

Source: Jiang et al. 2019, MNRAS, 488, 4801





https://apod.nasa.gov/apod/ap120717.html

Modern simulations seems to suggest that disk formation is a much more violent process than originally believed, with angular momentum being related to mergers and inflow of gas along streams and filaments….

…many challenges remain…

For an up-to-date review on theoretical challenges in galaxy formation, see Naab & Ostriker 2017, ARAA, 55, 59