

Galaxy formation and evolution

PAP 318, 5 op, autumn 2020
on Zoom

Lecture 14: Summary Lecture, 09/12/2020



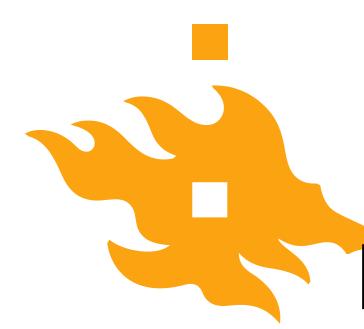
1. Galaxy formation as a process

- In order to understand galaxy formation we need to cover the following subjects:
 1. **Cosmology:** Galaxies evolve on cosmological time and length scales, thus we need to understand the space-time structure on large scales.
 2. **Initial conditions:** These were set by physical processes in the very early Universe and they are beyond direct astronomical observations. Strong link to particle physics based cosmology.
 3. **Physical processes:** The messy astrophysics that is required to understand galaxy formation include general relativity, hydrodynamics, dynamics of collisionless systems, plasma physics, thermodynamics, electrodynamics, atomic and nuclear physics and the theory of radiative processes.



Diversity of the galaxy population I

- Galaxies are diverse objects, meaning a large number of parameters is required to characterize any given galaxy:
1. **Morphology:** To first approximation there exists only two basic types of galaxies: Elliptical galaxies are mildly flattened, ellipsoidal systems that are mainly supported by random motions of their stars. Spiral galaxies have highly flattened disks that are mainly supported by rotation. The faintest galaxies do not typically fall on the Hubble sequence and they are divided into irregulars and dwarf spheroidals.
 2. **Luminosity and stellar mass:** Luminosities of galaxies can range between $\sim 10^3 L_\odot$ - $10^{12} L_\odot$. Total luminosity is related to the total number of stars in the galaxy, but with a large scatter due to different stellar populations in different galaxies. Most light in the Universe is contributed by Milky-Way type L^* galaxies, which are bright and relatively common.



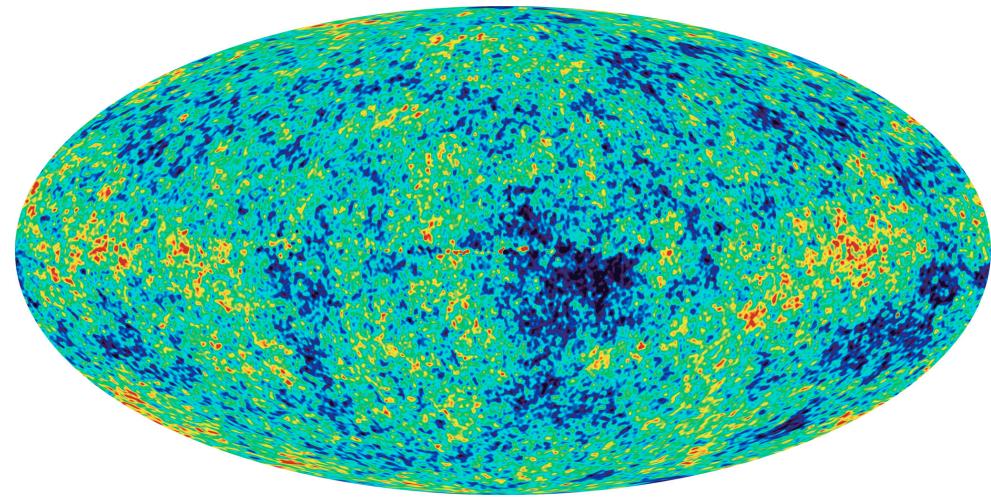
Diversity of the galaxy population II

3. **Size and surface brightness:** Galaxies do not have well defined boundaries, use half-light radius, which encloses half of the total light. For disk galaxies size is measure of their specific angular momenta and for ellipticals a measure of dissipation during their formation.
4. **Gas mass fraction:** The amount of gas: $f_{\text{gas}} = M_{\text{cold}} / [M_{\text{cold}} + M_*]$ is an important parameter. Typically ellipticals have negligibly small gas fractions, disk galaxies have typically $f_{\text{gas}} \sim 0.1$ in the local Universe with this fraction approaching unity for very high redshifts.
5. **Colour:** The colour of the galaxy reflects the underlying stellar population in the galaxies. Ellipticals, which have old stellar populations are typically red and disks with younger stars are bluer. Also higher metallicities make galaxies redder and extinction by dust also make a galaxy appear red.



Initial conditions

- If the Universe was perfectly uniform and isotropic, there would be no structure formation.
- In order to explain galaxies, we need some deviations from perfect uniformity.
- The standard model does not explain the origin of these perturbations.
- Currently the leading theory for the origin of perturbations is quantum fluctuations in the inflaton field in the very early Universe.

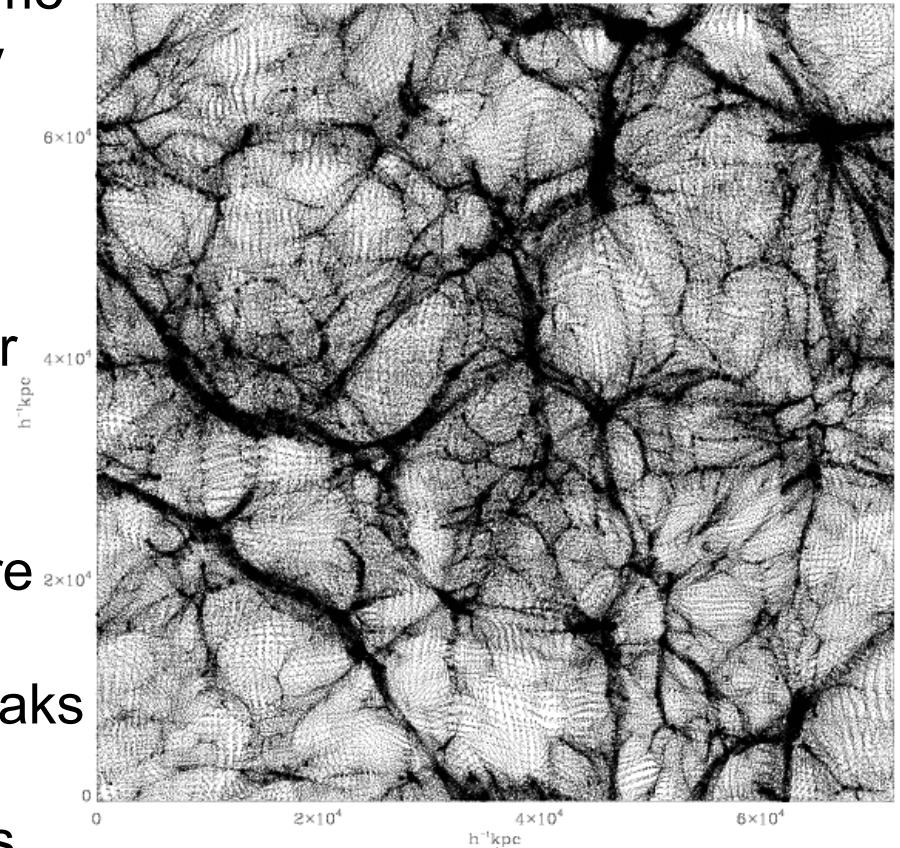


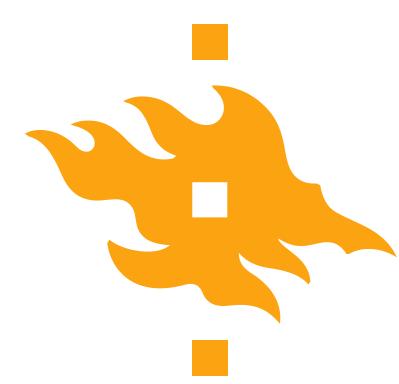
The cosmic microwave background (CMB) was formed $t=380,000$ years after the Big Bang.



Gravitational instability and structure formation

- Density perturbations grow with time in an expanding Universe. Slightly overdense regions will attract its surroundings more.
- In a static universe $\delta\rho/\rho$ grows exponentially (Jeans collapse, star formation), but in an expanding universe growth is slower $\delta\rho/\rho \propto t^\alpha$.
- At early times the perturbations are still linear $\delta\rho/\rho \ll 1$. Once the perturbation reaches $\delta\rho/\rho \sim 1$ it breaks away from the expansion and collapses -> numerical simulations.

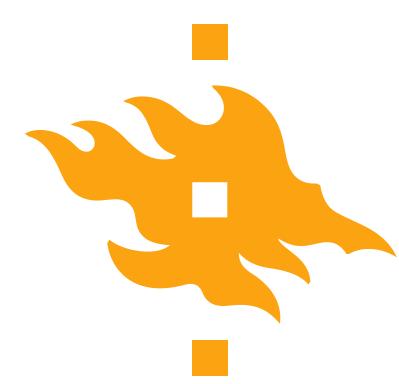




Timescales I

- The galaxy formation process is a competition among different processes, such as the collapse of dark matter, gas cooling, star formation etc. The relative importance of the various processes at various stages in a galaxies life is determined by whichever timescale is shortest:
- **Hubble time:** Time scale at which the Universe as a whole evolves, $t=H_0^{-1}$. This is the time scale on which substantial evolution of the galaxy population takes place.
- **Dynamical time:** This is the time required to orbit across an equilibrium dynamical system: For a system of mass M and radius R :

$$t_{\text{dyn}} = \sqrt{\frac{3\pi}{16G\bar{\rho}}}, \quad t_{\text{ff}} = t_{\text{dyn}}/\sqrt{2}$$

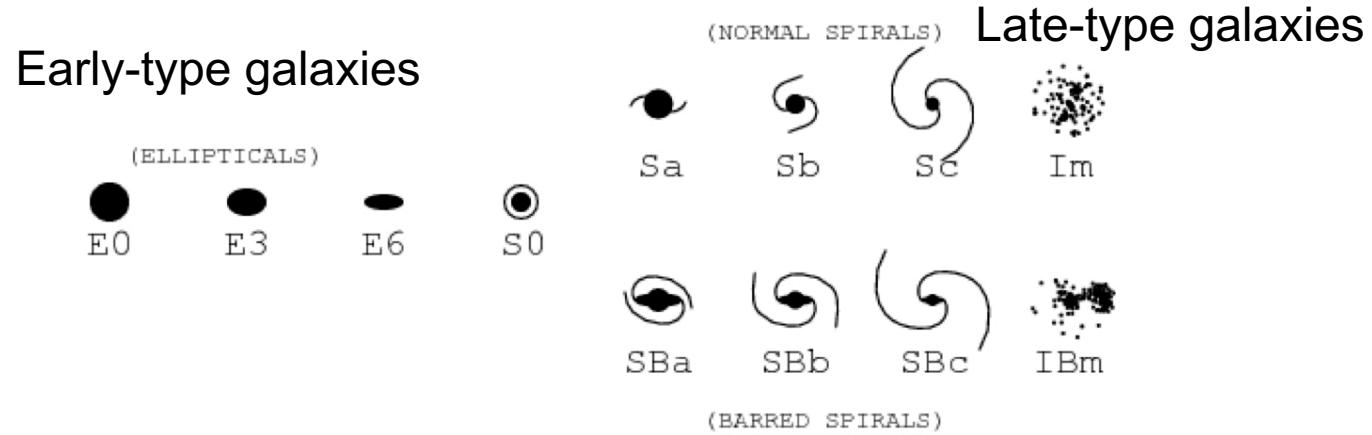


Timescales II

- **Cooling time:** This is the ratio between the thermal energy content and the energy loss rate through cooling for the gas component.
- **Star-formation timescale:** This is the ratio between the cold gas content of a galaxy and the star-formation rate.
- **Chemical enrichment time:** This is a measure for the time scale on which the gas is enriched by heavy elements. In general different for different elements.
- **Merger time:** The typical time a galaxy must wait before experiencing a merger with an object of similar mass.
- **Dynamical friction time:** The timescale on which a satellite object in a large halo loses its orbital energy and spirals to the centre. It is proportional to $M_{\text{sat}}/M_{\text{main}}$.



2. The classification of galaxies I



- The classic Hubble tuning-fork diagram still forms the basis for modern morphological galaxy classification, works best for bright galaxies.
- 1. Elliptical galaxies, have smooth, almost elliptical isophotes and are divided into subtypes E0, E1,...E7, where the integer is the one closest to $10(1-b/a)$, where a is the major and minor axis.
- 2. Spirals have thin disks with spiral arm structures and they are divided into two branches, barred and normal spirals.



Elliptical galaxies: Surface brightness

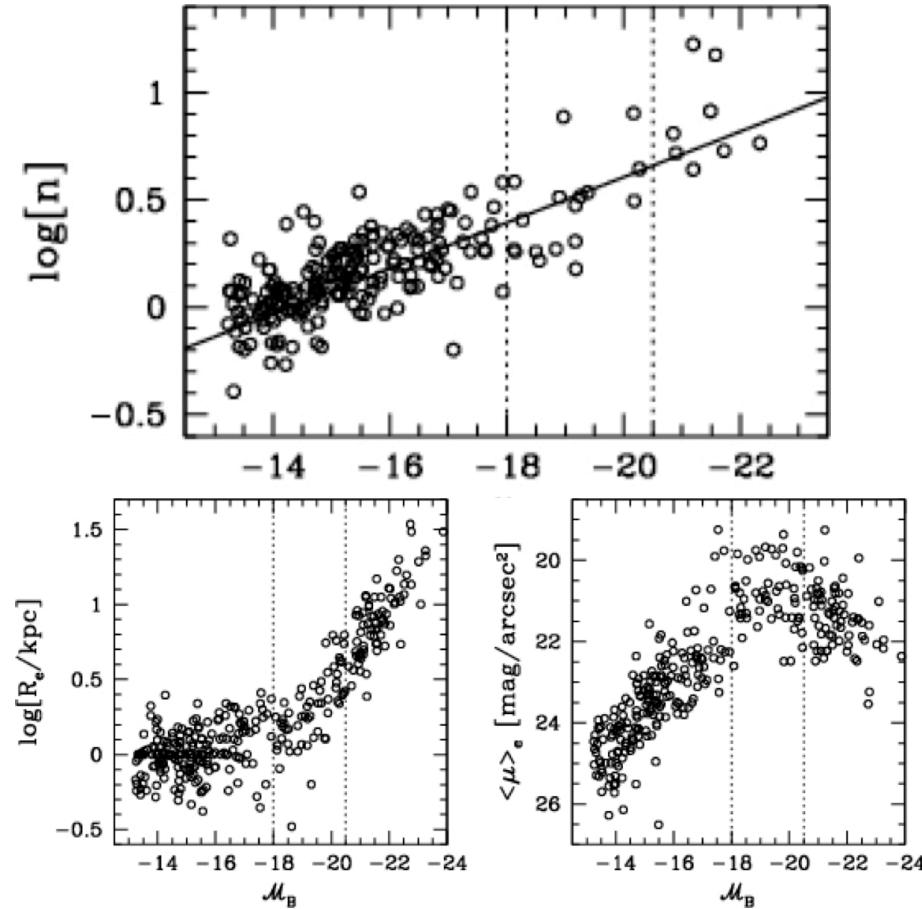
- The surface brightness profiles of ellipticals can be well fitted by the general Sérsic profile:

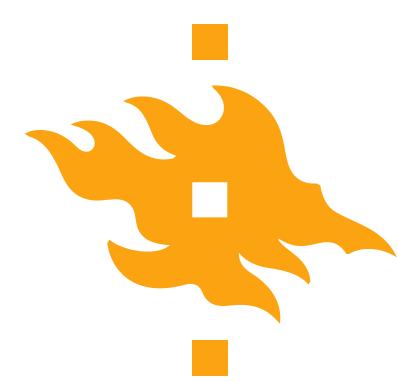
$$I(R) = I_e e^{-\beta_n [(R/R_e)^{1/n} - 1]}$$

- R_e is the half-light radius, $I_e = I(R_e)$, n is the Sérsic index ($n=1$ exponential and $n=4$ de Vaucouleurs' profile) and β_n follows from the definition of R_e .

- Total luminosity:

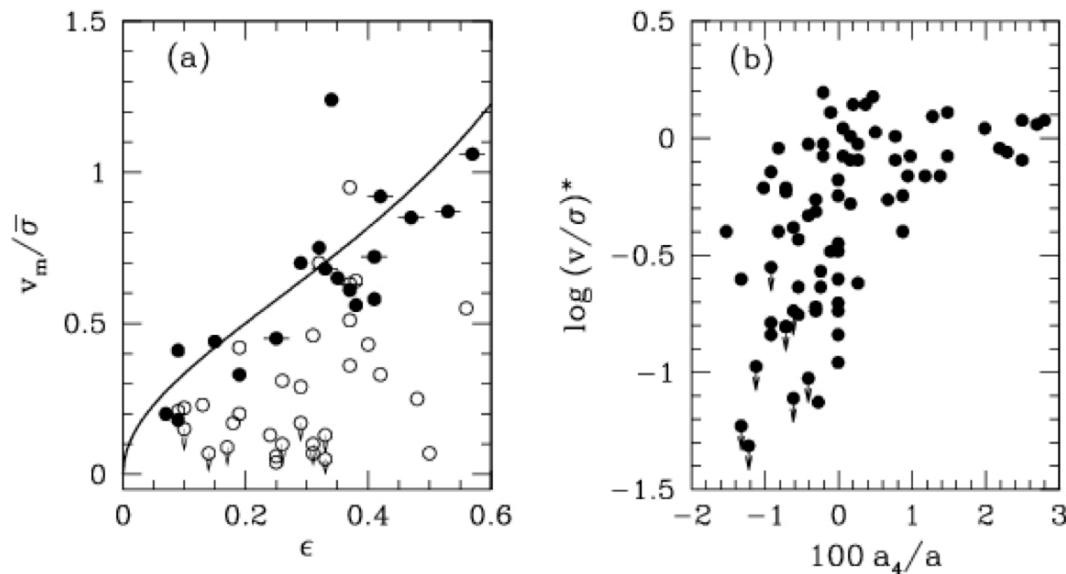
$$L = 2\pi \int_0^\infty I(R) R dR = \frac{2\pi n \Gamma(2n)}{(\beta_n)^{2n}} I_0 R_e^2$$



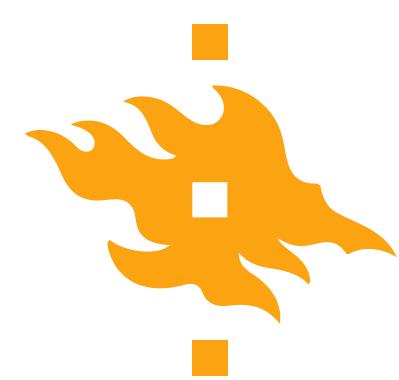


Kinematic properties

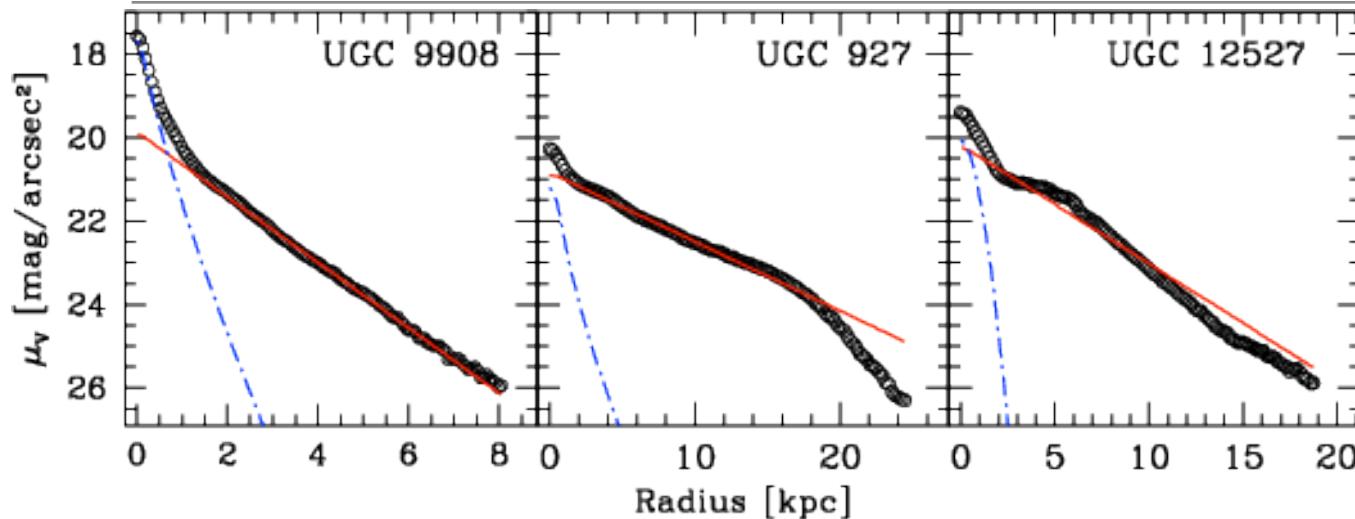
- Giant ellipticals are rotating very slowly and their very low v_{rot}/σ values cannot explain the observed flattening and it must be due to velocity anisotropy, rather than rotation.
- In contrast ellipticals at intermediate L show v_{rot}/σ values that are consistent with rotational flattening.
- Kinematic and photometric properties are linked!



The solid line gives the expected flattening (ϵ) as a function of the rotation measure v_{rot}/σ . Note that although low-L ellipticals have $v_{\text{rot}}/\sigma \sim 1$, disk galaxies have $v_{\text{rot}}/\sigma \sim 20$.



Disk galaxies: Surface brightness



- The surface brightness profiles can be fitted by exponential profiles:
$$I(R) = I_0 e^{-R/R_d}, \quad I_0 = \frac{L}{2\pi R_d^2}$$
- The inner parts of typical disk galaxies can be fitted by a Sérsic function and this describes the bulge-component, which is essentially a ‘minielliptical’.

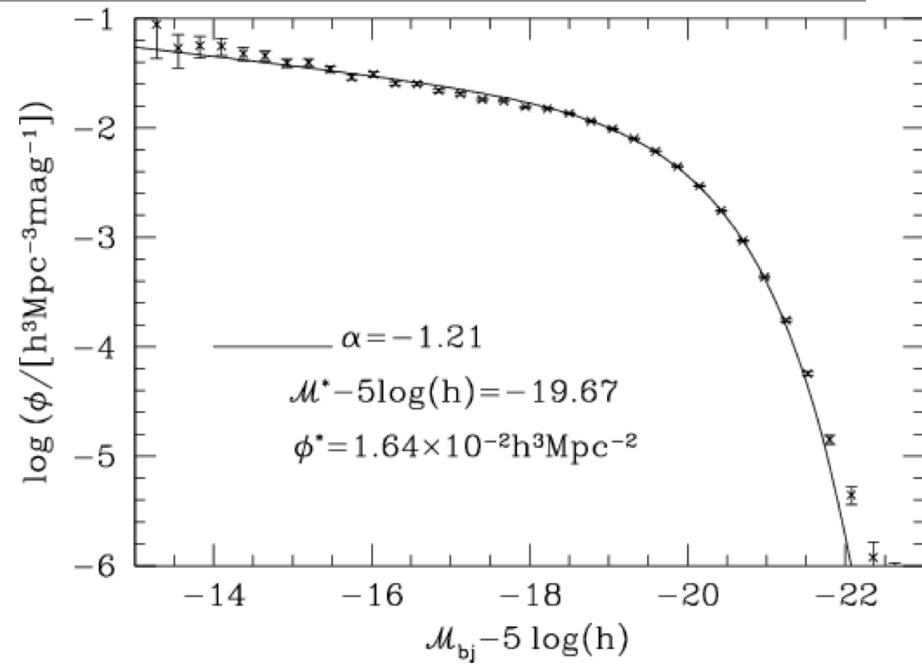


Statistical properties: Luminosity function

- The luminosity function $\phi(L)dL$ describes the number density of galaxies with luminosities in the range $L \pm dL/2$ and can be fitted by the Schechter function:

$$\phi(L)dL = \phi^* \left(\frac{L}{L^*} \right)^\alpha e^{-(L/L^*)} \frac{dL}{L^*}$$

- L^* is the characteristic luminosity, α is the faint-end slope, and ϕ^* is an overall normalization.
- The mean number density, n_g , and mean luminosity density L_g can be derived by integration:



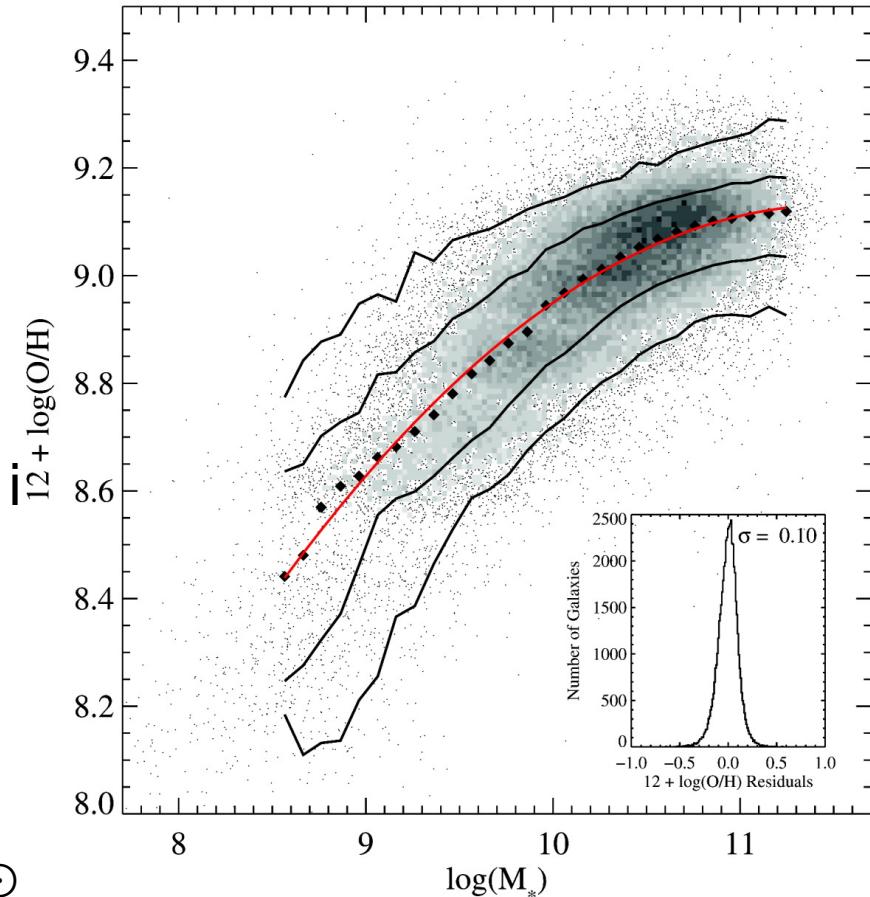
$$n_g = \int_0^\infty \phi(L)dL = \phi^* \Gamma(\alpha + 1)$$

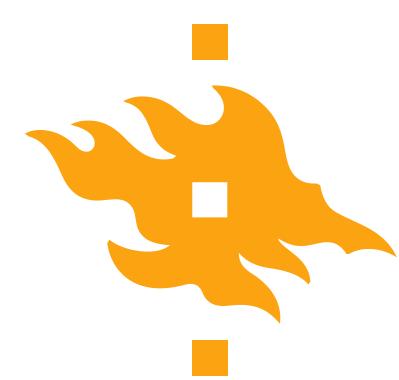
$$L_g = \int_0^\infty \phi(L)LdL = \phi^* L^* \Gamma(\alpha + 2)$$



Mass-metallicity relation

- The average galaxy metallicity is an important parameter and it can be different for the stars and the gas in the galaxy depending on the inflow/outflow properties.
- The Figure shows the average gas oxygen content relative to hydrogen in units of $12 + \log(\text{O/H})$.
- The correlation is remarkably tight with more massive galaxies having larger metallicities with a clear flattening above a few times $10^{10} M_{\odot}$





3. Robertson-Walker metric I

- In an isotropic and homogeneous Universe there exists a three-dimensional surface in space time, on which the density, temperature and expansion rate are uniform and evolve according to a universally agreed time, called the *cosmic time*.
- The Universe is maximally symmetric and can be described by Robertson-Walker metric: ([+---] Here general relativity notation.)

$$ds^2 = c^2 dt^2 - dl^2 = c^2 dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2(d\vartheta^2 + \sin^2 \vartheta d\phi^2) \right]$$

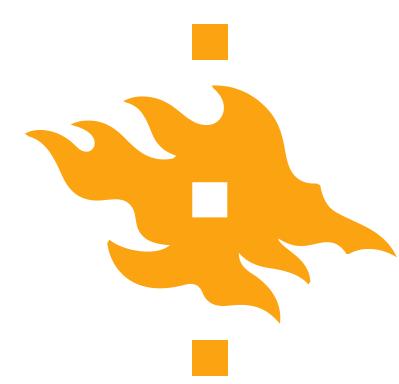
- $a(t)$ is the cosmic scale factor that describes the overall size of the Universe. K describes the curvature of the Universe ($K=+1$ closed, spherical geometry, $K=0$ flat and $K=-1$, open, hyperbolic geometry).



Angular diameter distance

- The comoving distance $\chi(r)$ and the proper distance $l=a(t)\chi(r)$ are not observables, because the light from a distant source observed at the present time was emitted at an earlier time.
- We can define an angular distance d_A that relates the observable angular size (ϑ) to the size of the object (D):
$$\vartheta = \frac{D}{d_A}$$
- The proper size D can be considered as the proper distance between two light signals, sent from two points with the same radial coordinate r_e at a given cosmic time t_e . Thus D is just the integral of dl in the Robertson-Walker metric and we get an expression for the angular diameter distance:

$$D = a_e r_e \int d\vartheta = \frac{a_0 r_e}{1+z} \vartheta \quad d_A = \frac{a_0 r_e}{1+z_e} = a_e r_e$$



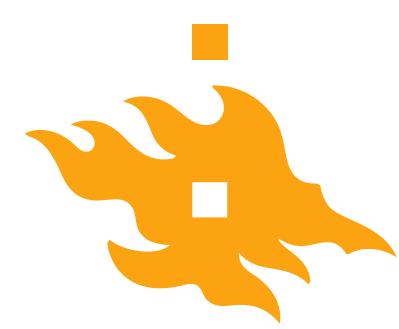
Luminosity distance

- Correspondingly the luminosity distance can be defined: $F = \frac{L}{4\pi d_L^2}$
- Let us consider an area A , which covers the solid angle ω at the distance of the observed object, corresponding to an area of ωd_A^2 . Because of the expansion of the Universe the corresponding area at the origin is larger:

$$A = \omega d_A^2 (a_0/a_e)^2 = (a_0 r_e)^2 \omega$$

- If the same number of photons pass through the area A in a time interval δt_0 we have:
$$\frac{L \delta t_e \omega}{4\pi h_P \nu_e} = \frac{F \delta t_0 A}{h_P \nu_0}$$

$$F = \frac{\omega}{4\pi} \frac{L}{A} \left(\frac{a_e}{a_0} \right)^2 = \frac{L}{4\pi [a_0 r_e (1+z)]^2} \quad d_L = D(1+z) = d_A(1+z)^2$$



Friedmann equations IV

- Now, finally the Friedmann equations can be derived by inserting the values for the Ricci tensor and the energy momentum tensor: We get two separate equations for the time-time (0,0) component and for the space-space (1:3,1:3) components:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + 3\frac{P}{c^2} \right) + \frac{\Lambda c^2}{3} \quad \text{FRW1: Time-time components}$$

$$\frac{\ddot{a}}{a} + \frac{2\dot{a}^2}{a^2} + \frac{2Kc^2}{a^2} = 4\pi G \left(\rho - \frac{P}{c^2} \right) + \Lambda c^2 \quad \text{Space-space components}$$

- Finally inserting the first equation into the second we get FRW2:

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3}$$



Age of the Universe I

- The age of the Universe in a homogeneous expanding Universe can be derived from the equation:

$$t(z) = \int_0^{a(z)} \frac{da}{a} = \frac{1}{H_0} \int_z^\infty \frac{dz}{(1+z)E(z)}$$

- This equation can be integrated numerically for any cosmology and in special cases analytically.
- For the EdS model $\Omega_{m,0}=1$ ja $\Omega_{\Lambda,0}=0$:

$$t = \frac{1}{H_0} \frac{2}{3} (1+z)^{-3/2}$$



Distances in the Universe II

- Finally we can derive the angular diameter distance in comoving coordinates:

$$r = f_K \left[\frac{c}{H_0 a_0} \int_0^z \frac{dz}{E(z)} \right]$$

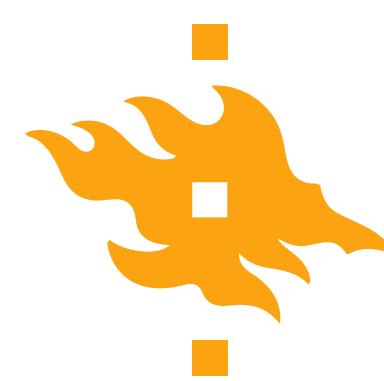
$$f_K(\chi) = \sin \chi \ (K = +1); \ f_K(\chi) = \chi \ (K = 0); \ f_K(\chi) = \sinh \chi \ (K = -1)$$

- When $z \ll z_{\text{eq}}$ and $\Omega_{\Lambda,0}=0$ a closed expression exists for all three values of K (Mattig's formula):

$$a_0 r = \frac{2c}{H_0} \frac{\Omega_0 z + (2 - \Omega_0)[1 - (\Omega_0 z + 1)^{1/2}]}{\Omega_0^2(1 + z)}$$

- For a flat ($\Omega_{m,0} + \Omega_{\Lambda,0} = 1$) universe $r = \chi$:

$$a_0 r = \frac{c}{H_0} \int_0^z \frac{dz}{[\Omega_{\Lambda,0} + \Omega_{m,0}(1 + z)^3]^{1/2}}$$



The growth of small perturbations I

- In what follows we present a Newtonian analysis for the growth of small perturbations. The more correct approach would be to use a full relativistic analysis, which is far from trivial. However, using the classic Newtonian analysis important results can be derived.
- The equations of gas dynamics for a fluid in a gravitational field can be written down using the Lagrangian derivative:
 1. Equation of continuity: $\frac{d\rho}{dt} = \frac{\partial}{\partial t} + (\bar{v} \cdot \nabla)$
 2. Equation of motion: $\frac{d\bar{v}}{dt} = -\frac{1}{\rho} \nabla p - \nabla \phi$
 3. Gravitational potential: $\nabla^2 \phi = 4\pi G \rho$



Final equation for growth of perturbations

- Combining the previous equations we get:

$$\frac{d^2\Delta}{dt^2} + 2 \left(\frac{\dot{a}}{a} \right) \frac{d\Delta}{dt} = \frac{c_s^2}{\rho_0 a^2} \nabla_c^2 \delta\rho + 4\pi G \delta\rho$$

- We get the final equation by seeking wave solution for Δ of the form
$$\Delta \propto e^{i\vec{k}_c \cdot \vec{r} - \omega t}$$
$$\frac{d^2\Delta}{dt^2} + 2 \left(\frac{\dot{a}}{a} \right) \frac{d\Delta}{dt} = \Delta(4\pi G \rho_0 - k^2 c_s^2)$$
- This is a second order differential equation and describes the general evolution of small density perturbations $\Delta = \delta\rho/\rho$ in the Newtonian non-relativistic case.



4. Expanding medium

- Let us then return to the full version of the equation. The second term, which describes expansion modifies the Jeans' analysis and thus the resulting growth rate will be different. We begin by studying the long-wavelength limit $\lambda \gg \lambda_J$, in which case the pressure term $c_s^2 k^2$ can be neglected:

$$\frac{d^2 \Delta}{dt^2} + 2 \left(\frac{\dot{a}}{a} \right) \frac{d\Delta}{dt} = 4\pi G \rho \Delta$$

- Before considering the general solution, let us first solve this equation for two special cases $\Omega_0=1$ and for $\Omega_0=0$, for which the scale factors evolve as:

$$\Omega_m = 1, a = \left(\frac{3}{2} H_0 t\right)^{2/3} \quad \Omega_m = 0, a = H_0 t$$



Expanding medium $\Omega_m=1$

- The equation can be rewritten using the following substitutions as:

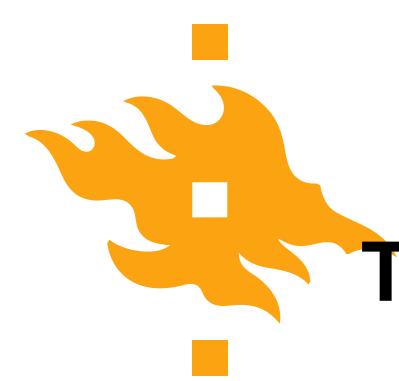
$$\Omega_m = 1, a = \left(\frac{3}{2}H_0t\right)^{2/3}, \frac{\dot{a}}{a} = \frac{2}{3t}, 4\pi G\rho = \frac{2}{3t^2}$$
$$\frac{d^2\Delta}{dt^2} + \frac{4}{3t}\frac{d\Delta}{dt} - \frac{2}{3t^2}\Delta = 0$$

- By inspection, it can be seen that power-law solution of the form $\Delta=at^n$ must exist:

$$n(n-1) + \frac{4}{3}n - \frac{2}{3} = 0 \Rightarrow n_1 = \frac{2}{3}, n_2 = -1$$

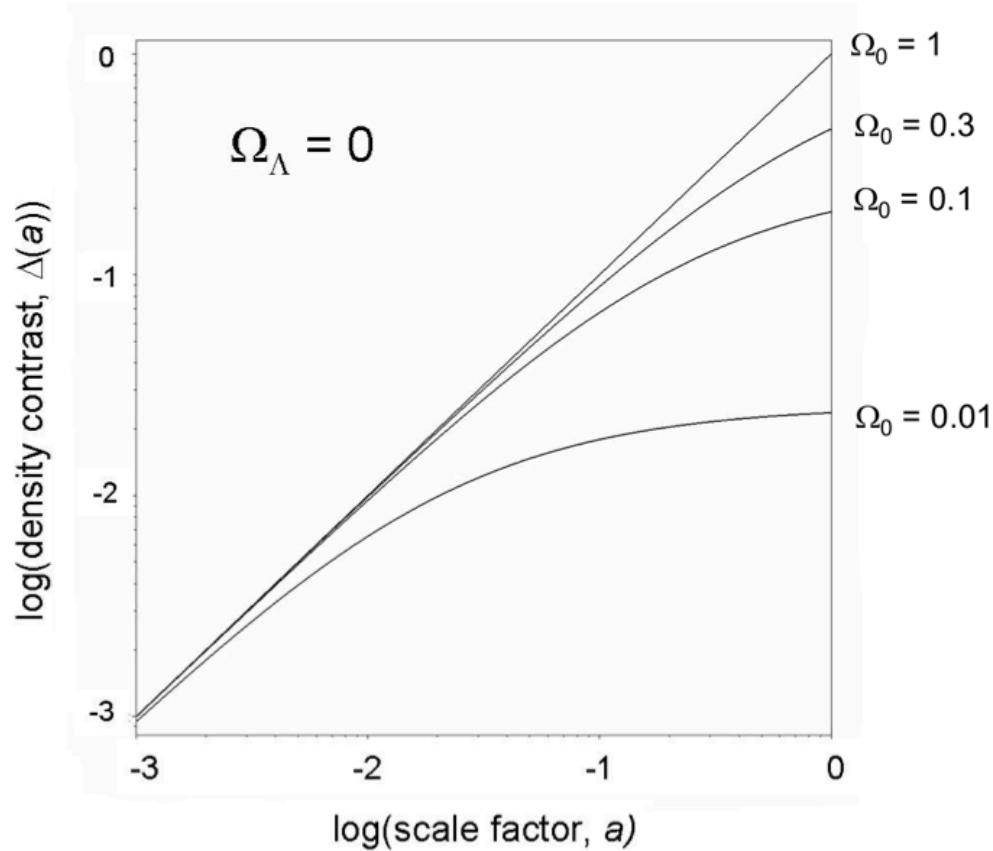
The latter solution corresponds to a decaying mode. The $n=2/3$ solution corresponds to a growing mode: $\Delta \propto t^{2/3} \propto a = (1+z)^{-1}$:

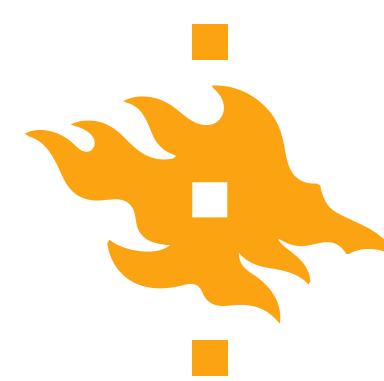
$$\Delta = \frac{\delta\rho}{\rho} \propto (1+z)^{-1}$$



The general solution: Models with $\Omega_\Lambda=0$

- The constants have been chosen, so that the $\Delta=1$ at the present epoch $a=1$ for the standard $\Omega_{0,m}=1$, $\Omega_\Lambda=0$ model.
- The plot shows Δ from $a=10^{-3}$ to $a=1$ and is consistent with the analytical result that the amplitudes of the density perturbations vary as $\Delta \propto a$ so long as $\Omega_0 z \gg 1$, but the growth essentially stops at smaller redshifts.





Instabilities in the relativistic case III

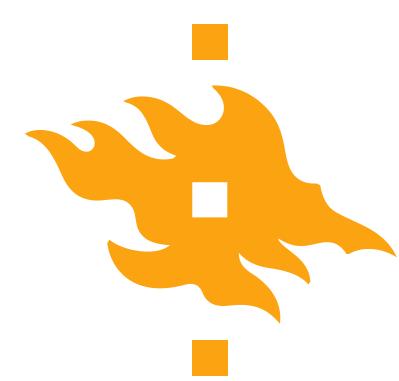
- Neglecting the pressure gradient terms in the equation we get:

$$\frac{d^2\Delta}{dt^2} + 2 \left(\frac{\dot{a}}{a} \right) \frac{d\Delta}{dt} = \frac{32\pi G\rho}{3} \Delta$$

- We once again seem powerlaw solution of the form $\Delta=at^n$, recalling that in the radiation-dominated era, the scale factor evolves with cosmic time as $a\propto t^{1/2}$.
- We find solutions $n=\pm 1$. Hence for long wavelengths $\lambda>>\lambda_J$, the growing mode corresponds to:

$$\Delta \propto t \propto a^2 \propto (1+z)^{-2}$$

- Thus again perturbations grow only algebraically with cosmic time instead of exponential growth.



Basic problem of galaxy formation

- In summary, throughout the matter-dominated era, the growth rate of perturbations on physical scales much greater than the Jeans' length is: $\Delta = (\delta\rho/\rho) \propto a = (1+z)^{-1}$.
- Since galaxies exist at the present-day ($z=0 \Delta \geq 1$), it follows that at the last scattering surface (CMB) at $z \sim 1000$, fluctuations must have been present with amplitudes of at least $\Delta = (\delta\rho/\rho) \geq 10^{-3}$.
- The slow growth of density perturbations is the source of a fundamental problem in understanding the origin of galaxies. The large-scale structure could not have condensed out of the primordial plasma by exponential growth of infinitesimal statistical perturbations.
- Because of the slow development of the density perturbations, we have the opportunity of studying the formation of structure directly at redshifts of below $z \sim 1000$.



The gravitational potential on super-horizon scales

- By studying the conformal Newtonian gauge we can crudely estimate the form of the Newtonian potential, which can be used on superhorizon scales in both the radiation- and matter-dominated eras.
- Matter era: $\nabla^2 \delta\phi = 4\pi G \delta\rho \Rightarrow \frac{\delta\phi}{L^2} = 4\pi G \rho \Delta$
- Radiation era: $\nabla^2 \delta\phi = 8\pi G \delta\rho \Rightarrow \frac{\delta\phi}{L^2} = 8\pi G \rho \Delta$
- Inserting the densities and referring to the scale L as the comoving scale, $L=aL_0$, we see that in both cases we get:
$$\delta\phi = 4\pi G \rho_0 \Delta_0 L_0^2 \quad \& \quad \delta\phi = 8\pi G \rho_0 \Delta_0 L_0^2$$
- The superhorizon perturbations in the gravitational potential are independent of the scale factor. These are the potentials that need to be inserted in the metric and they are frozen on superhorizon scales. There cannot be causal connection on the superhorizon scale L.



5. The sound speed as a function of cosmic epoch I

- The sound speed is proportional to the square root of the ratio of the pressure which provides the restoring force to the inertial mass density of the medium (subscript S, constant entropy, adiabatic sound waves):
$$c_S^2 = \left(\frac{\partial p}{\partial \rho} \right)_S$$
- The sound speed in the early Universe changes as the dominant contributors to the pressure, p and density ρ change. The sound speed can then be written as:

$$c_S^2 = \frac{(\partial p / \partial T)_r}{(\partial \rho / \partial T)_r + (\partial \rho / \partial T)_m} = \frac{c^2}{3} \frac{4\rho_r}{4\rho_r + 3\rho_m}$$



Adiabatic fluctuations in the matter-dominated era

- The variation of the baryonic mass within the particle horizon can be worked out as using $r_H \propto a^{3/2}$.

$$M_{b,\text{hor}} = \left(\frac{\pi r_H^3}{6} \right) \rho_b = \frac{3.0 \times 10^{22}}{(\Omega_0 h^2)^{1/2}} a^{3/2} M_\odot$$

- The Jeans' mass is now with the appropriate sound speed of $c_s = c(4\rho_{\text{rad}}/9\rho_b)^{1/2} \propto a^{-1/2}$, $\lambda_J \propto c_s \rho_b^{-1/2} \propto a$. Thus the Jeans mass will be independent of the scale factor a :

$$M_J = \frac{3.75 \times 10^{15}}{(\Omega_b h^2)^2} M_\odot$$

- Note that perturbations that are larger than this mass grew according to the standard result $\Delta \propto a$, when they came through the horizon.
- At recombination the pressure from the radiation disappears and the appropriate sound speed is the sound speed of gas with $T=3000$ K.

$$M_J = (\pi \lambda_J^3 / 6) \rho_B = 1.6 \times 10^5 (\Omega_0 h^2)^{-1/2} M_\odot$$



The damping of sound waves I

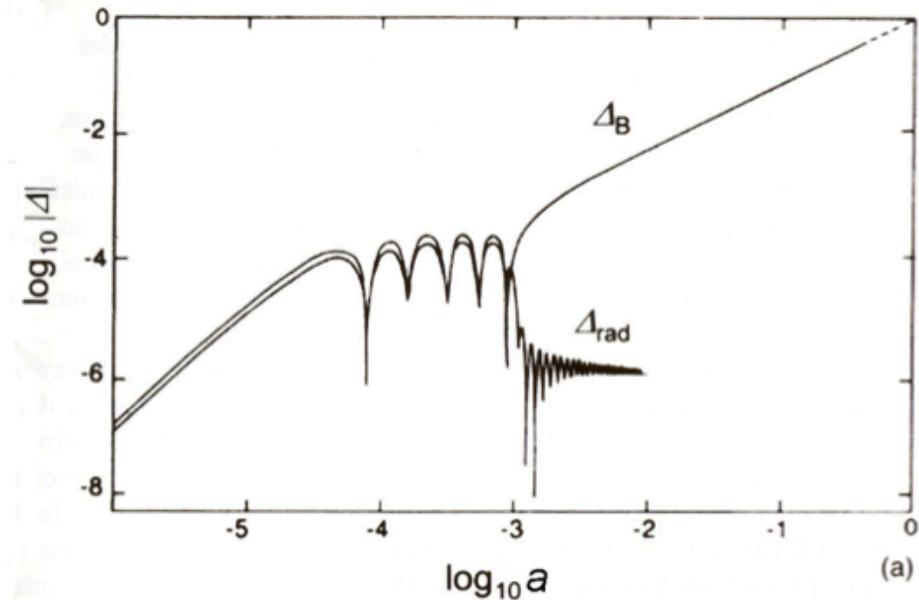
- Although the matter and radiation are closely coupled through the pre-recombination era, the coupling is not perfect and radiation can diffuse out of the density perturbations. Since the radiation provides the restoring force for support of the perturbation, the perturbation is damped out if the radiation has time to diffuse out of it. This process is referred to as Silk damping.
- At any epoch the free path of scattering of photons by electrons is $\lambda = (N_e \sigma_T)^{-1}$, where $\sigma_T = 6.665 \times 10^{-29} \text{ m}^2$ is the Thomson cross-section. According to kinetic theory the photons can diffuse over the distance:

$$r_D \approx (Dt)^{1/2} = \left(\frac{1}{3} \lambda c t\right)^{1/2}$$

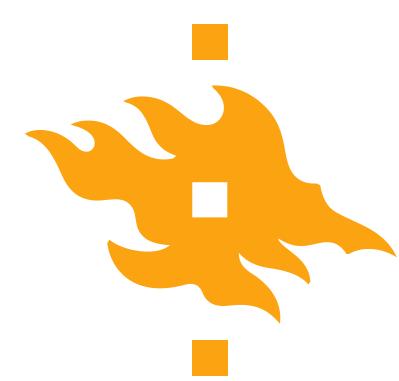


Problems with the baryonic models of galaxy formation

- A major problem with baryonic models of galaxy formation is that constraints from primordial nucleosynthesis show that the majority of the matter density must be in non-baryonic form.
- In purely baryonic theories, the fluctuations in the CMB had to have large amplitudes of $\Delta \gtrsim 3 \times 10^{-3}$, well above the observed constraints.



The diagram shows schematically how perturbations develop in purely baryonic structure formation models.



Instabilities in the presence of dark matter I

- Neglecting the internal pressure of the fluctuations, the expressions for the density contrasts in the baryons and the dark matter Δ_B and Δ_D respectively, can be written as a pair of coupled equations:

$$\begin{cases} \ddot{\Delta}_B + 2\left(\frac{\dot{a}}{a}\right)\dot{\Delta}_B = 4\pi G\rho_B\Delta_B + 4\pi G\rho_D\Delta_D \\ \ddot{\Delta}_D + 2\left(\frac{\dot{a}}{a}\right)\dot{\Delta}_D = 4\pi G\rho_B\Delta_B + 4\pi G\rho_D\Delta_D \end{cases}$$

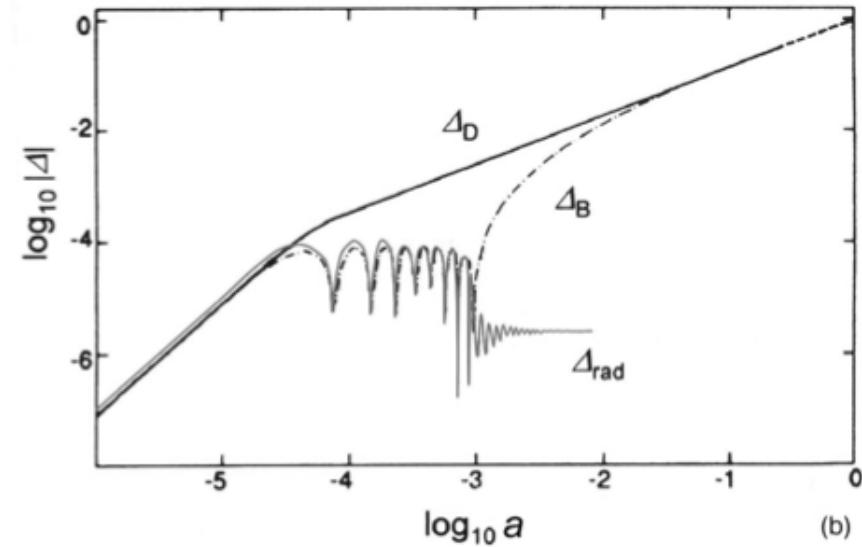
- Let us find the solution for the case in which the dark matter $\Omega_0=1$ and the baryon density is negligible compared with that of the dark matter. Then the second equation above reduces to the equation for which we have already found the solution $\Delta_D=Ba$, where B is a constant. Therefore the equation for the evolution of the baryon perturbation becomes:

$$\ddot{\Delta}_B + 2\left(\frac{\dot{a}}{a}\right)\dot{\Delta}_B = 4\pi G\rho_DBa$$

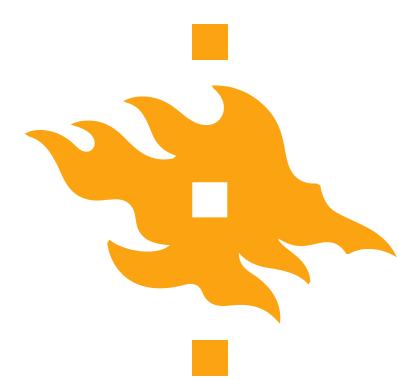


The standard cold dark matter (CDM) scenario I

- Separate evolution of the CDM perturbation and the coupled photon-baryon plasma from when the perturbation entered the horizon until recombination.
- Regeneration of large-amplitude perturbations in baryons after the epoch of recombination.
- The decay of perturbations in the photon gas after the epoch of recombination.
- The growth of perturbations on superhorizon scales at $a < 3 \times 10^{-5}$ as the potential fluctuations were frozen-in on super-horizon scales.



Structure development in the standard CDM picture. The observed $\Delta \sim 10^{-5}$ fluctuations in the CMB can only be reconciled with the existence of galaxies in models with cold dark matter.



6. The two-point correlation function for galaxies I

- To make a quantitative comparison between theories of galaxy formation and the observed distribution of galaxies, we need to quantify the spectrum of density perturbations in the Universe.
- The simplest quantitative description of the statistical distribution of galaxies on large scale is provided by the two-point correlation function, which describes the excess probability of finding a galaxy at distance r from a galaxy selected at random over that expected in a uniform, random distribution:

$$dN(r) = N_0[1 + \xi(r)]dV$$

- N_0 is the background number density and the correlation function $\xi(r)$ can also be written in terms of finding pairs of galaxies as:

$$dN_{\text{pair}} = N_0^2[1 + \xi(r)]dV_1 dV_2$$



The perturbation spectrum

III

- Performing the integral over an isotropic probability distribution of angles θ on a sphere, that is we integrate $\cos(kr \cos \theta)$ over $(\frac{1}{2})\sin \theta d\theta$ gives the final result:

$$\xi(r) = \frac{V}{2\pi^2} \int |\Delta_k|^2 \frac{\sin kr}{kr} k^2 dk = \frac{V}{2\pi^2} \int P(k) \frac{\sin kr}{kr} k^2 dk$$

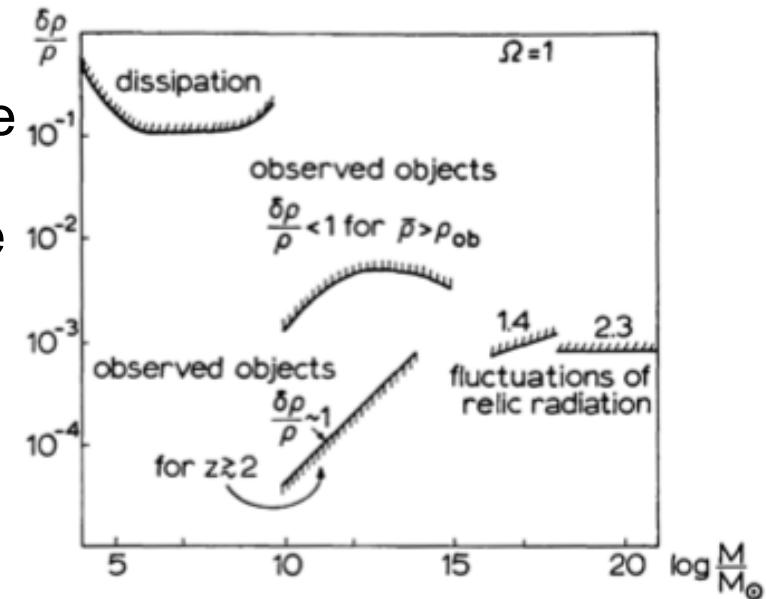
- The function $(\sin kr)/kr$ allows only wavenumbers $k \leq r^{-1}$ to contribute to the amplitude of fluctuations on the scale r . Fluctuations with larger wavenumbers corresponding to smaller scales, average out to zero on the scale r .
- The inverse relation for the power spectrum can also be derived:

$$P(k) = \frac{1}{V} \int_0^\infty \xi(r) \frac{\sin kr}{kr} 4\pi r^2 dr$$



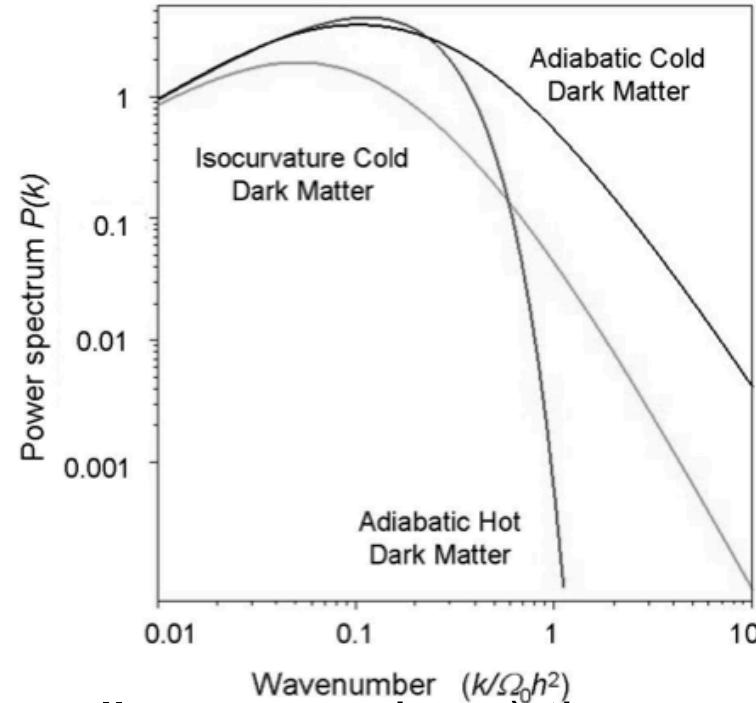
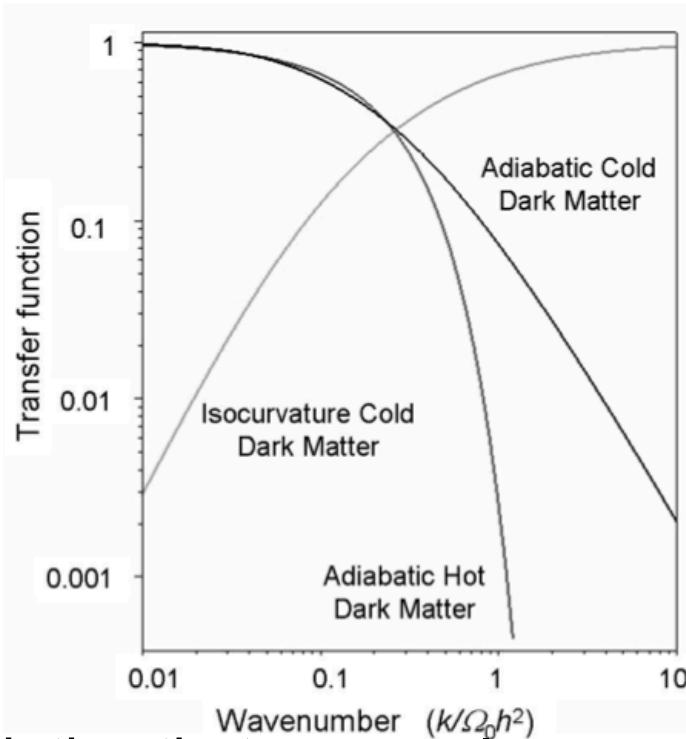
The Harrison-Zeldovich power spectrum

- The rather special value of $n=1$ is known as the Harrison-Zeldovich spectrum. In particular Sunyaev and Zeldovich found that in order to produce the observed structure today, the initial fluctuations had to have a scale-invariant spectrum $\delta\rho/\rho=10^{-4}$ on mass scales from 10^5 to $10^{20} M_\odot$.
- Harrison studied the form the primordial spectrum must have in order to prevent the overproduction of excessively large amplitude perturbations on small and large scales. A power spectrum of the form $P(k)\propto k$ does not diverge on large physical scales and so is consistent with the observed large-scale isotropy of the Universe.

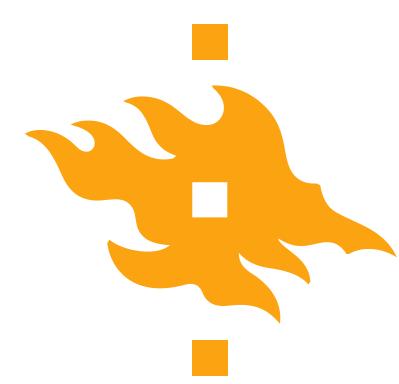




Transfer functions and the resulting power spectra



- Notice that on very large scales (small wavenumbers) the spectrum is unprocessed. On the scales of galaxies and clusters, the spectrum has been strongly modified.



Biassing I

- So far we have implicitly assumed that the visible parts of galaxies trace the distribution of dark matter. However this might not always be the case and the generic term for this is biassing:

$$\xi_{\text{gal}}(r) = b^2 \xi_D(r)$$

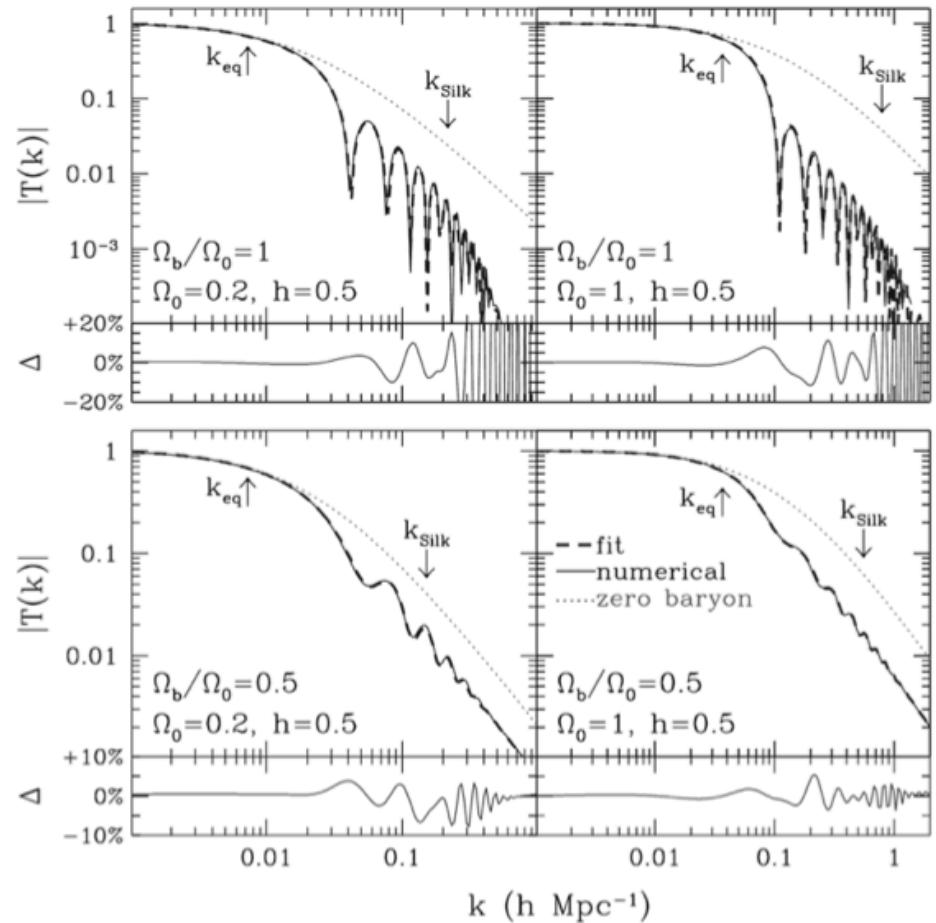
$$P_{\text{gal}}(k) = b^2 P_D(k) \quad \left(\frac{\delta\rho}{\rho}\right)_{\text{gal}} = b \left(\frac{\delta\rho}{\rho}\right)_D \quad \Delta_{\text{gal}} = b \Delta_D$$

- Here b is the bias-parameter, if the $b > 1$ galaxies are a biased representation of the underlying dark matter.
- This parameter was especially important in early cosmological models, where the observations of galaxies had to be reconciled with $\Omega_0=1$. Now that we know that $\Omega_0=0.3$ recent analysis have shown that $b=1.04 \pm 0.11$ (consistent with $b=1$).



The role of baryon perturbations

- Four examples of the transfer function for models with baryons only (top pair of diagrams) and with mixed cold and baryonic models (bottom pair of diagrams).
- The numerical results are shown as solid lines and their fitting functions by dashed lines.
- The baryons, which make up ~20% of the total mass leave imprints upon the galaxy power spectrum.



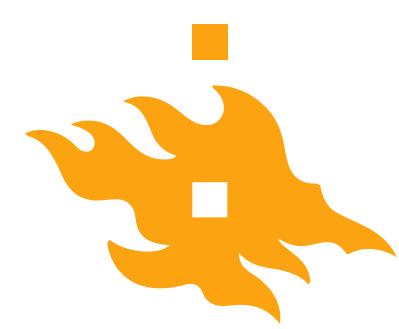


7. Isotropic top-hat collapse I

- The collapse of a uniform spherical density perturbation in an otherwise uniform Universe can be worked out exactly analytically and this model is usually referred to as spherical top-hat collapse.
- The dynamics are the same as those of a closed Universe with $\Omega_0 > 1$ and the variation of the scale factor of the perturbation a_p is given by the parametric solution (see also Lecture 4):

$$a_p = A(1 - \cos \theta) \quad t = B(\theta - \sin \theta)$$
$$A = \frac{\Omega_0}{2(\Omega_0 - 1)} \quad B = \frac{\Omega_0}{2H_0(\Omega_0 - 1)^{3/2}}$$

- The perturbation reached maximum radius at $\theta = \pi$ and then collapsed to infinite density at $\theta = 2\pi$.



Final overdensity in the spherical top-hat collapse model I

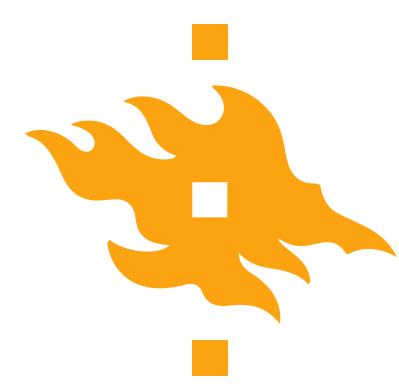
- The virialization happens when the perturbation has collapsed by a factor of 2 and the density of the perturbation has increased by a factor of $2^3=8$.
- The scale factor of the perturbation reached value $a_{\max}/2$ at time $t=(1.5+\pi^{-1})t_{\max}=1.81t_{\max}$. The background density was then:

$$\left(\frac{a_{\max}}{a_{\text{vir}}}\right)^3 = \left(\frac{t_{\max}}{t_{\text{vir}}}\right)^2 = \frac{1}{1.81^2}$$

- An alternative is to define t_{vir} as the time when $a_p=0$, so that $t_{\text{vir}}=t(\theta=2\pi)=2t_{\max}$. In this case the background density has dropped by $1/2^2=1/4$. The final overdensities are depending on the definition:

$$\Delta \sim 5.55 \times 8 \times 1.81^2 \approx 145$$

$$\Delta \sim 5.55 \times 8 \times 2^2 \approx 178$$



The Zeldovich approximation I

- In the next approximation called the Zeldovich approximation it is assumed that the perturbations were ellipsoidal with three unequal principal axes, as expected from the superposition of Gaussian fields.
- In the Zeldovich approximation, the development of the perturbations into the non-linear regime is followed in Lagrangian coordinates. If \mathbf{x} and \mathbf{r} are the proper and comoving position vectors of the particles of the fluid, the Zeldovich approximation can be written:

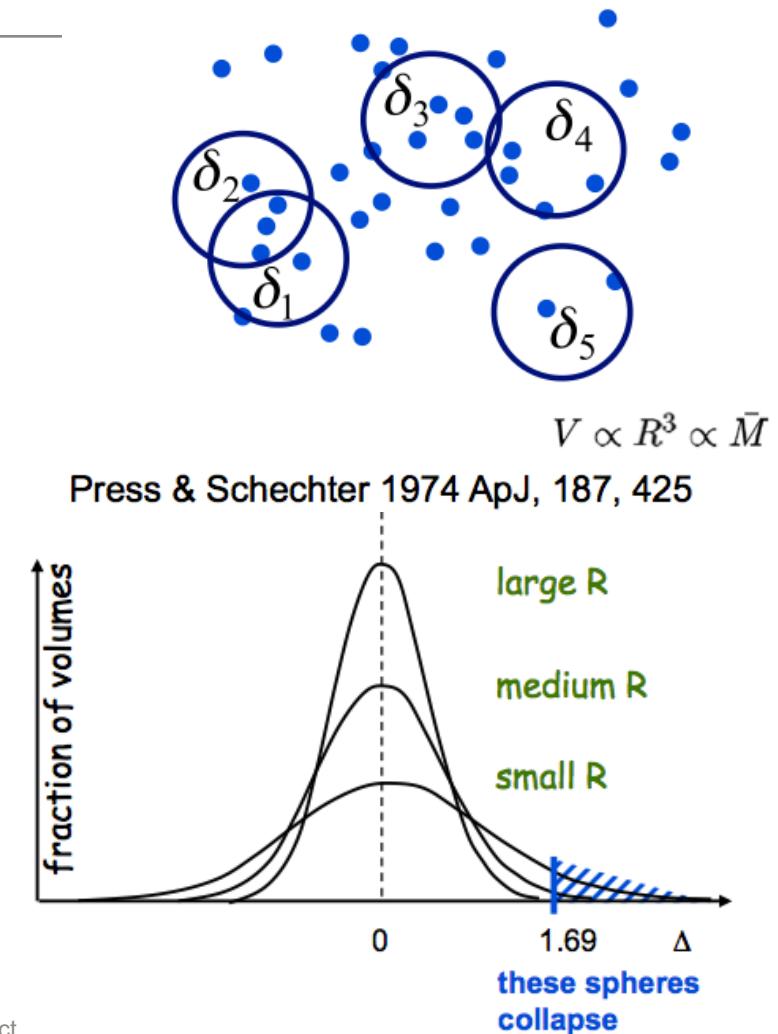
$$\vec{x} = a(t)\vec{r} + b(t)\vec{p}(\vec{r})$$

- The first term on the right-hand side describes the uniform expansion of the background model and the second term the perturbations of the particles' positions about the Lagrangian (comoving) coordinate \mathbf{r} .



Mass function of collapsed haloes

- A smoothly fluctuating density field can be described by randomly scattered spheres with each having some overdensity δ . Some of the spheres have a large enough overdensity ($\delta_c > 1.69$) that they will eventually collapse and form gravitationally bound objects.
- Now the question is: “*What is the mass function of these objects at any given cosmic epoch?*”





Press-Schechter mass function IV

- The number density of bound objects is given by:

$$N(M)dM = \frac{1}{V} = -\frac{\bar{\rho}}{M} \frac{\partial F}{\partial M} dM$$

- Combining with the expression of $F(M)$ and given that: $\frac{d\Phi}{dx} = \frac{2}{\sqrt{\pi}} e^{-x^2}$
- We finally get the expression for the mass function:

$$N(M) = \frac{1}{2\sqrt{\pi}} \left(1 + \frac{n}{3}\right) \frac{\bar{\rho}}{M^2} \left(\frac{M}{M^*}\right)^{(3+n)/6} \exp \left[-\left(\frac{M}{M^*}\right)^{(3+n)/3} \right]$$

- This expression consists of a power law times an exponential. The time dependence of $N(M)$ has been absorbed into the variation of M^* with cosmic epoch.



The NFW profile

- The NFW profile is defined by the equations:

$$\frac{\rho(r)}{\rho_{\text{crit}}} = \frac{\delta_c}{(r/r_s)(1+r/r_s)^2}$$

$$\delta_c = \frac{200}{3} \frac{c^3}{[\ln(1+c) - c/(1+c)]}$$

- $r_s = r_{\text{virial}}/c$ is the scale radius and $c \sim 5-15$ is the concentration parameter.



The NFW profile



8. Shock heating II

- In addition, if we make the assumption that the gas fall in from a very large distance where the gravitational potential $\Phi(r) \sim 0$, we get the following expression for the initial velocity:

$$v_{\text{in}} \simeq v_{\text{esc}}(r_{\text{sh}}) = \sqrt{2|\Phi(r_{\text{sh}})|}$$

- Finally v_{in} can be related to virial velocity v_{vir} by making the common assumption $r_{\text{sh}} = r_{\text{vir}}$ then:

$$v_{\text{in}}^2 = \zeta \frac{GM_{\text{vir}}}{r_{\text{vir}}} = \zeta v_{\text{vir}}^2$$

$$T_{\text{sh}} \simeq \frac{\zeta \mu m_p}{3k_B} v_{\text{vir}}^2$$

- The correction factor ζ is of the order of unity and depends on the details of the dark matter halo profile, i.e. isothermal, NFW, etc.



Virial temperature II

- If we ignore the external pressure P_{ext} we get for the virial temperature exactly the same result as for the shock-heated gas:

$$T_{\text{vir}} = \frac{\zeta \mu m_p}{3k_B} v_{\text{vir}}^2$$

- As an example we can take the truncated, singular isothermal sphere of gas (no dark matter), $\zeta=3/2$, then the virial theorem implies a virial temperature of:

$$T_{\text{vir}} = \frac{\mu m_p}{2k_B} v_{\text{vir}}^2 \simeq 3.6 \times 10^5 \text{ K} \left(\frac{v_{\text{vir}}}{100 \text{ kms}^{-1}} \right)^2$$

- Here the molecular weight $\mu=0.59$ appropriate for fully ionized primordial gas. The virial temperature is a very useful diagnostic, although realistic virialized gas in dark matter haloes is in general not isothermal.



Cooling time I

- The cooling time is defined as the time it takes the gas to radiate away its internal energy for a given metallicity and is given by:

$$t_{\text{cool}} = \frac{\rho \epsilon}{C} = \frac{\rho \epsilon}{n_H^2 \Lambda(T)}$$

- For monatomic ideal gas ($\gamma=5/3$) of primordial composition and fully ionized the cooling time can be written as:

$$t_{\text{cool}} = \frac{3nk_B T}{2n_H^2 \Lambda(T)} \simeq 3.3 \times 10^9 \text{yr} \left(\frac{T}{10^6 \text{K}} \right) \left(\frac{n}{10^{-3} \text{cm}^{-3}} \right)^{-1} \left(\frac{\Lambda(T)}{10^{-23} \text{ergs}^{-1} \text{cm}^3} \right)^{-1}$$

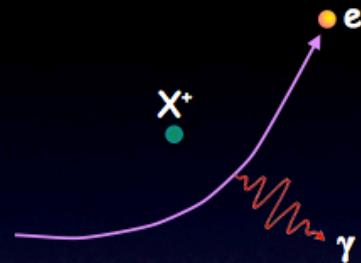
- The cooling time $t_{\text{cool}} \propto n^{-1} \propto \rho^{-1}$ -> denser gas cools faster.
- The impact of cooling can be estimated by comparing it to the Hubble time ($t_H \propto H(z)^{-1}$) and the dynamical free-fall time ($t_{\text{ff}} \propto \rho^{-1/2}$) Lect. 1.



Cooling processes I

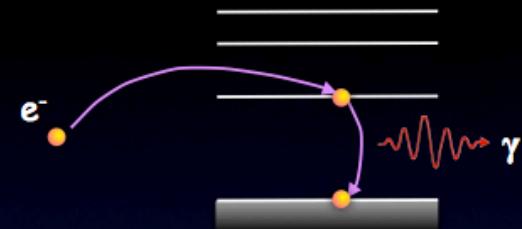
All processes require free electrons and are strongly dependent on temperature.

1) free-free (bremsstrahlung)



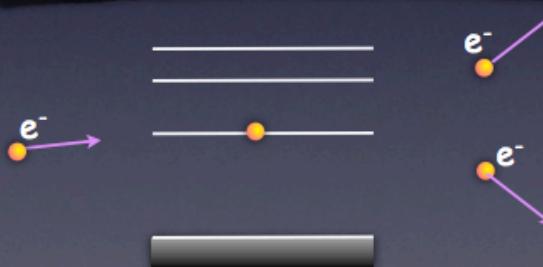
Free electron is accelerated by ion. Accelerated charges emit photons, resulting in cooling. For bremsstrahlung, $\Lambda \propto T^4$

2) free-bound (recombination)



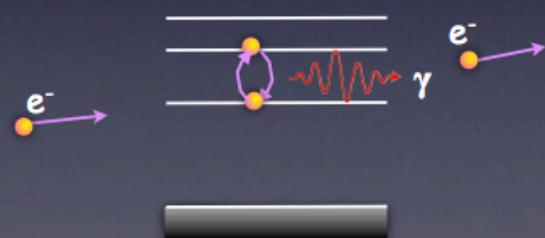
Free electron recombines with ion. Binding energy plus free electron's kinetic energy are radiated away. Only the latter counts as a loss (=cooling), though, since binding energy is already accounted for as loss in collisional ionization process.

3) bound-free (collisional ionization)

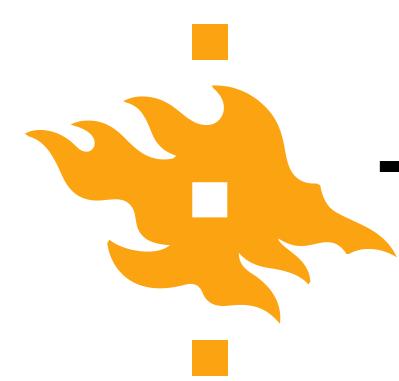


Impact of free electron ionizes a formerly bound electron, taking (kinetic) energy from the free electron

4) bound-bound (collisional excitation)

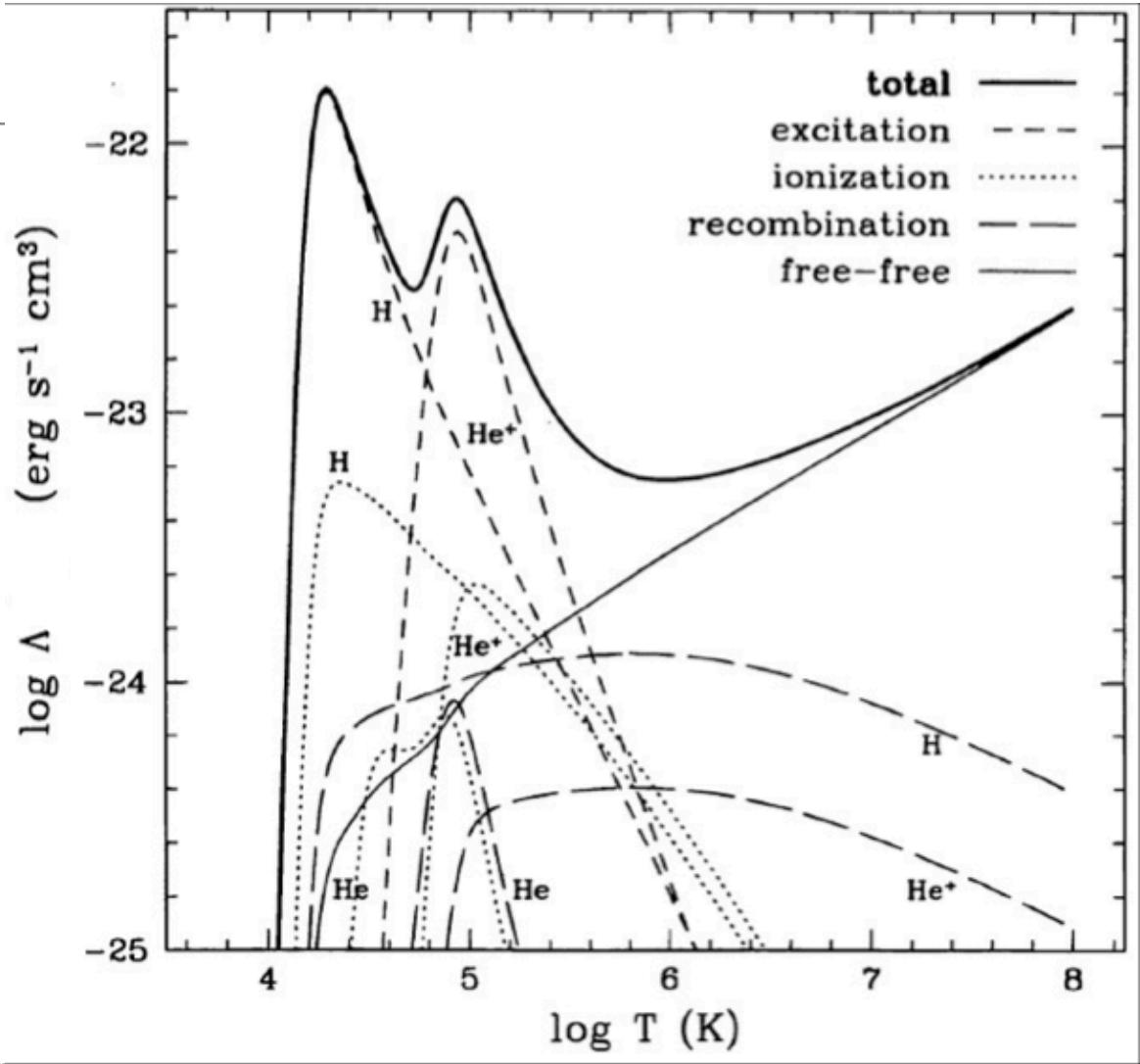


Impact of free electron knocks a bound electron to an excited state. As it decays, it emits a photon



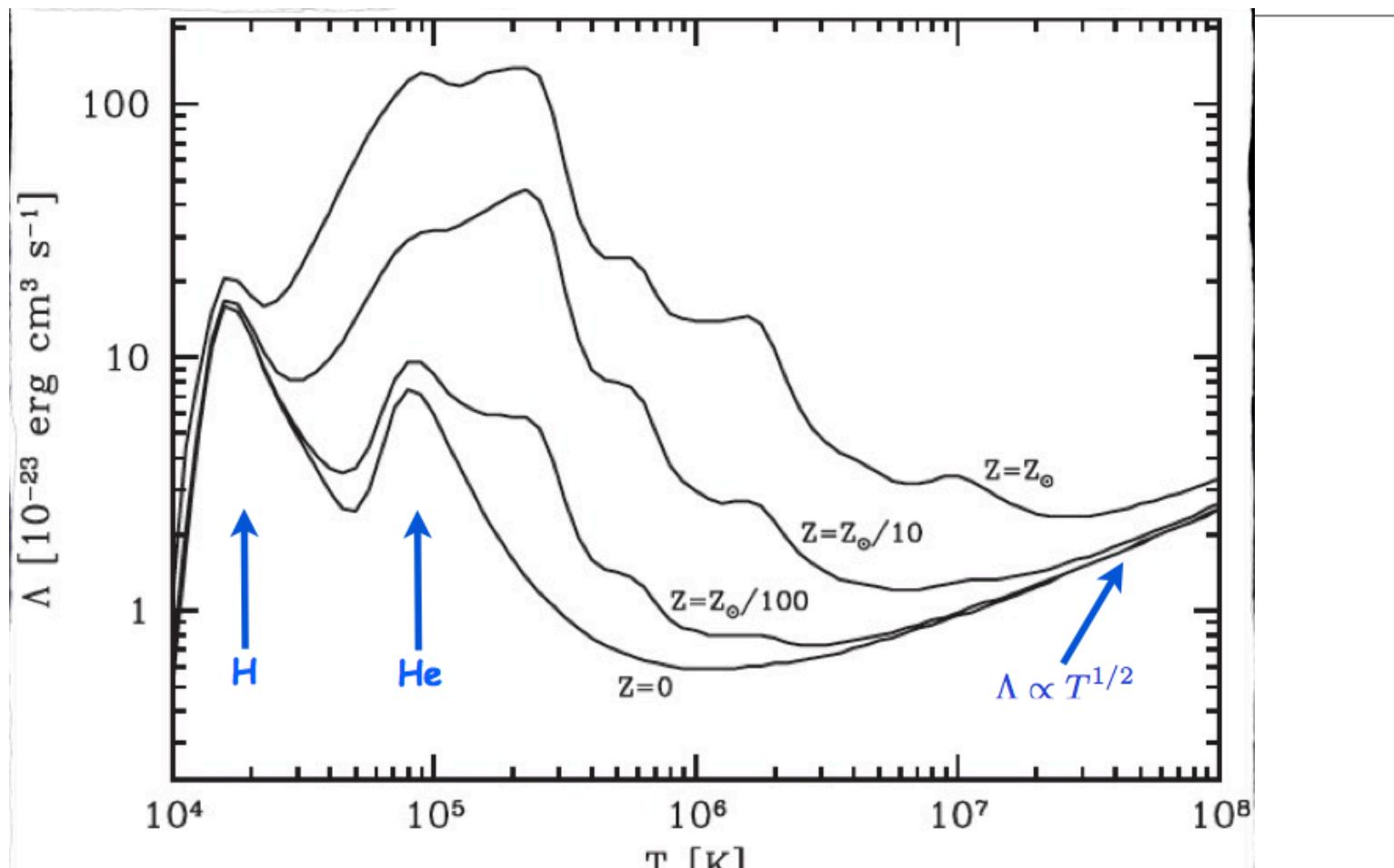
The primordial cooling function

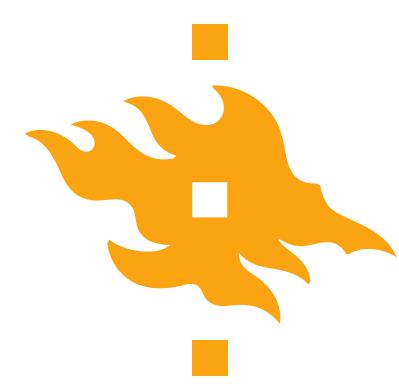
- The primordial cooling function has two distinct peaks due to hydrogen and helium.
- The excitation of H and He are the dominant sources of cooling.





The cooling function with non-zero metallicity I



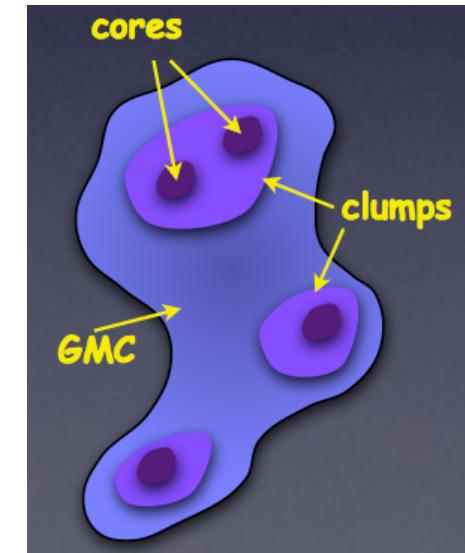


9. Giant molecular clouds

- GMCs are strongly correlated with young massive stars ($t < 10^7$ yrs).
- The free-fall time of a GMC can be estimated as:

$$t_{\text{ff}} = \left(\frac{3\pi}{32G\rho} \right)^{1/2} \simeq 3.6 \times 10^6 \text{ yrs} \left(\frac{n_{\text{H}_2}}{100 \text{ cm}^{-3}} \right)^{-1/2}$$

- However, the observed lifetimes of GMC is much longer at $t \sim 10^7$ yrs, GMC must somehow be supported against collapse.
- We can also define a star formation efficiency of a GMC as $\epsilon_{\text{SF}} = t_{\text{ff}} / t_{\text{SF}}$, where the star formation timescale $t_{\text{SF}} = M_{\text{GMC}} / \text{SFR}$. Observations indicate that $\epsilon_{\text{SF}} \sim 0.002$. Why is the star formation efficiency in GMCs so low?



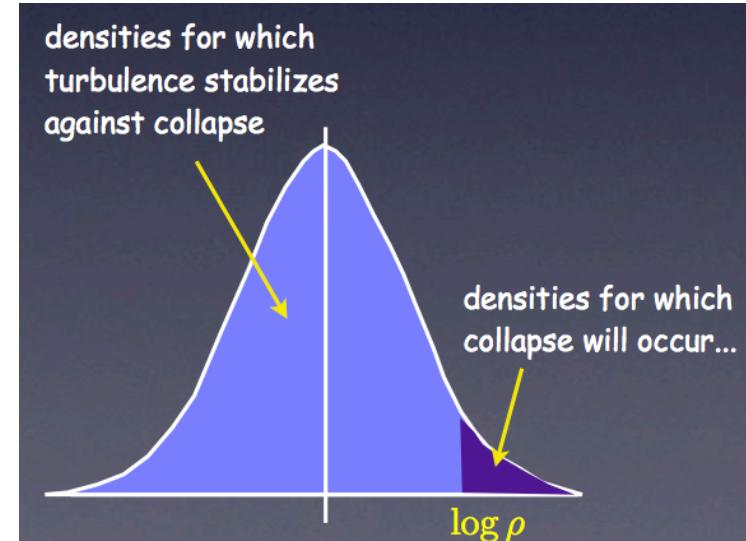


GMCs and supersonic turbulence II

- Turbulence is driven at some large scale and then decays to smaller scales until the turbulent energy is dissipated at the dissipation scale.
- Turbulence affects both the effective sound speed of the gas and its density (at areas of compression the density is boosted by the Mach number squared):

$$M_J \propto \frac{(c_s^2 + \sigma_v^2)^{3/2}}{Mach^2 \rho^{1/2}}$$

- On large scales $\sigma_v \gg c_s$ turbulent motions increase the effective pressure. On small scales $\sigma_v < c_s$ and turbulent compression boosts gas densities. -> log-normal density distribution.





Self-regulated star formation

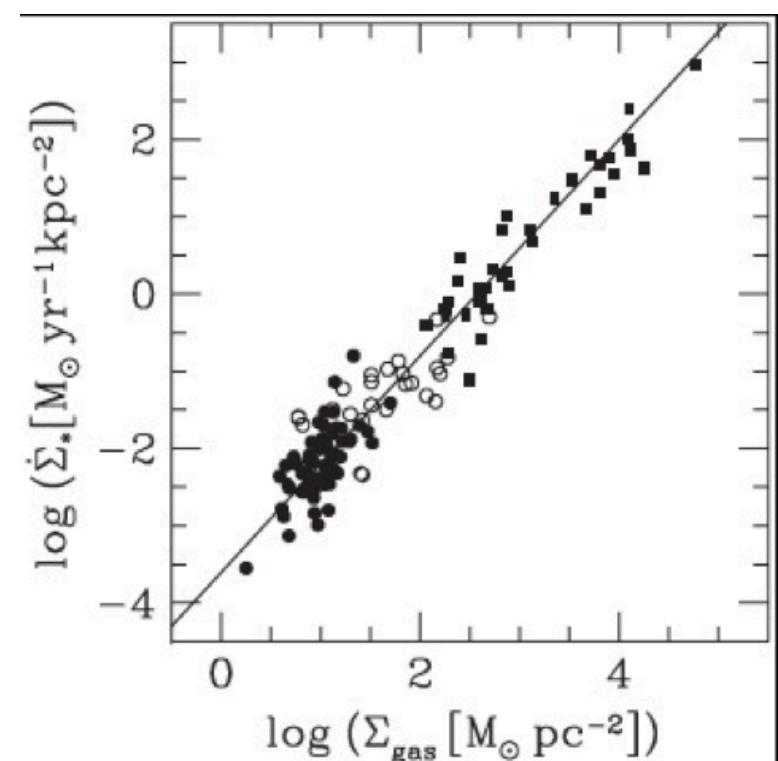
- In addition to turbulence, the overall star formation efficiency (SFE) of GMCs can also be influenced by star formation itself -> **self-regulation**.
- For example, feedback from proto-stellar winds are believed to regulate star formation efficiency of the stellar cores.
- GMCs as a whole are believed to be ultimately destroyed by energy feedback from massive OB stars (photo-evaporation by HII regions, stellar winds and SN explosions).
- Star formation may also provoke star formation (positive feedback). Shock waves associated with supernovae, stellar winds and ionization fronts may compress neighbouring gas and therefore inducing star formation.



Empirical star formation laws: Global relations I

- In galaxy formation model it is in practise impossible to resolve the ~20 orders of magnitude relevant for star formation. Instead one typically resorts to empirical star formation laws, which are scaling relations between the SFR and global properties such as gas density, etc.
- The most well-known empirical star formation law is the Schmidt-Kennicutt law that relates the global averaged star formation density to the surface-gas density (atomic+molecular):

$$\dot{\Sigma}_* \simeq 2.5 \times 10^{-4} \left(\frac{\Sigma_{\text{gas}}}{M_\odot \text{pc}^{-2}} \right)^{1.4} M_\odot \text{yr}^{-1} \text{kpc}^{-2}$$





Supernova feedback

- One of the most important problems in galaxy formation is the overcooling problem. Preventing overcooling requires some heat input.
- According to the current paradigm the main heating mechanisms are feedback from supernovae and AGN feedback (Lecture 13).



Supernova feedback in M82



Supernova feedback in numerical simulations

1. Thermal feedback:

- A fraction $\varepsilon_{\text{SN}} \leq 1$ of the SN energy is given to neighbouring gas particles in the form of thermal energy. A problem with this approach is that the gas in star-forming regions is very dense resulting in rapid cooling and consequently most of the SN energy is rapidly radiated away. A solution is to turn off cooling for a time period Δt to allow the thermal pressure to disperse the gas.

2. Kinetic feedback:

- A fraction $\varepsilon_{\text{SN}} \leq 1$ of the SN energy is given to neighbouring gas particles in form of kinetic energy, the wind velocity has to be put in by hand. A problem is that the star-forming gas is dense preventing gas from escaping to large distances. A solution would be to turn off the hydrodynamics for wind particles for a time period Δt to allow the kinematic motion to disperse the gas.



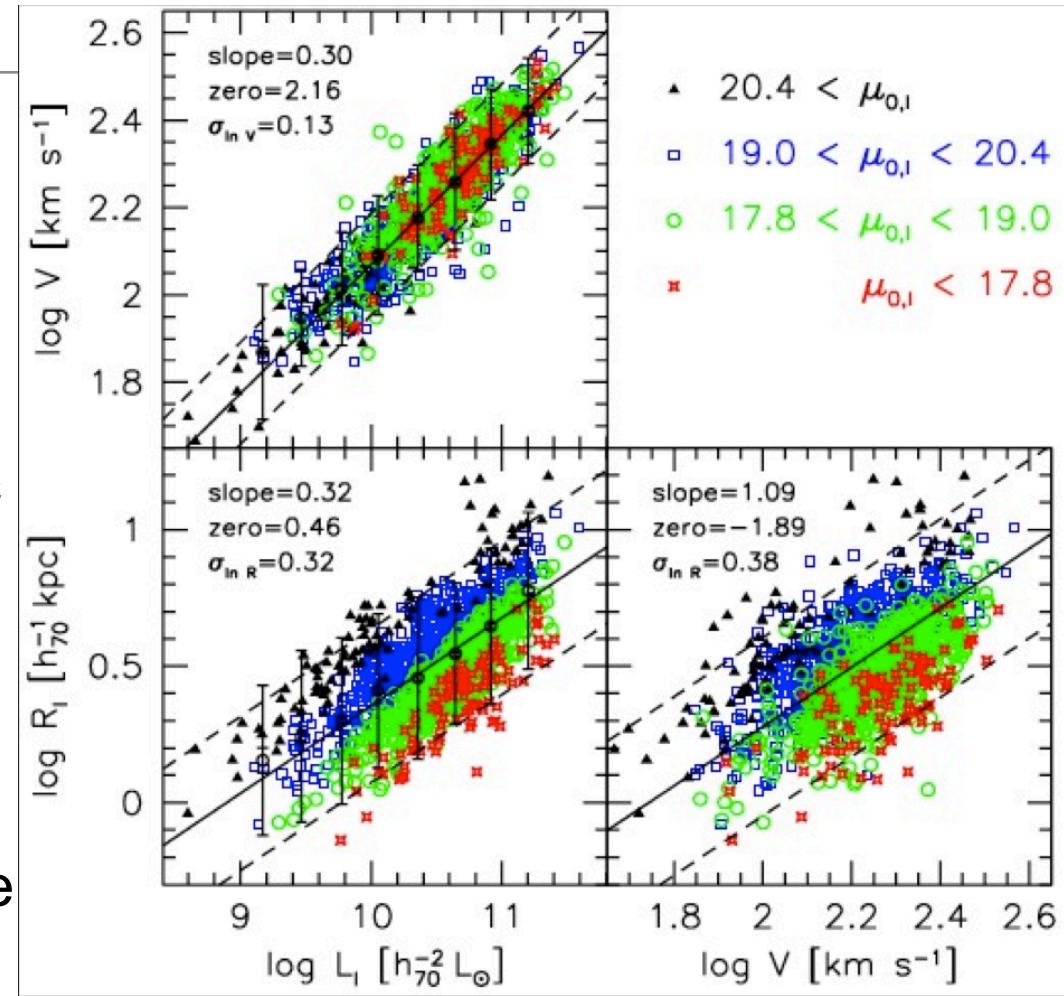
10. Disk galaxies: Observational results

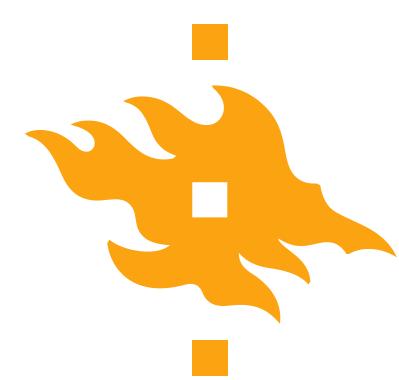
- A disk galaxy consists of a disk component made of stars and atomic and molecular gas, often with spiral arms and in about half of the cases a central bar component.
- A successful model of disk galaxy formation should be able to explain the observed structure of disk galaxies and the following observational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



The Tully-Fisher relation

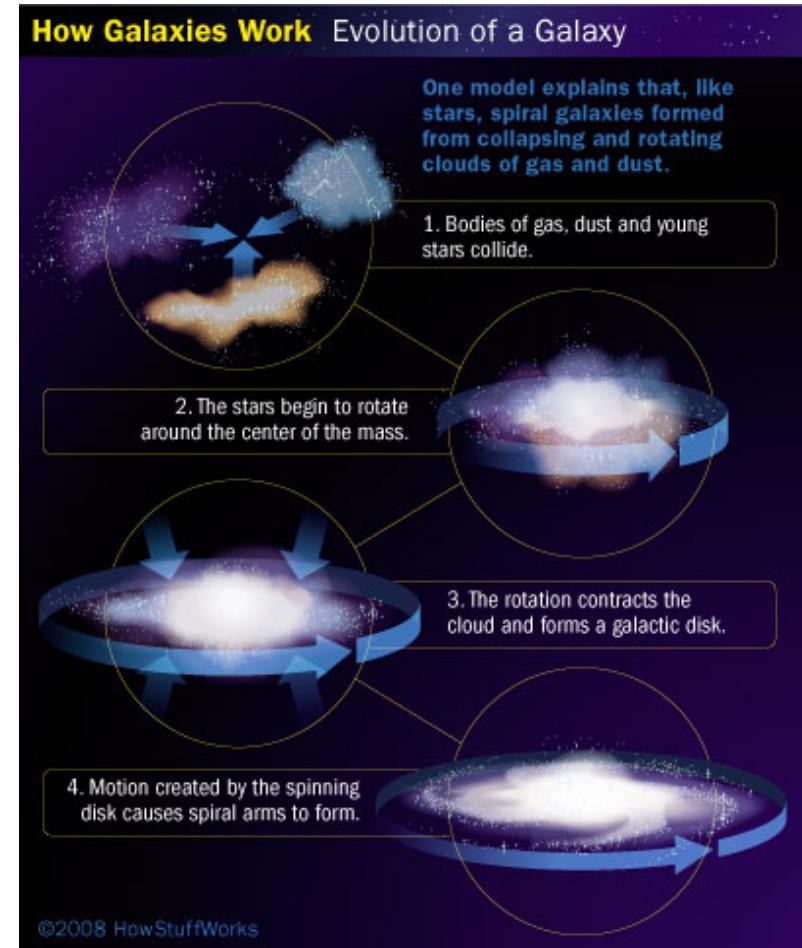
- Brighter disks also rotate faster as manifested in the Tully-Fisher relation:
$$V \propto L_I^{0.30}$$
- The slope of the TF-relation depends on the photometric band. The scatter in the TF-relation is not correlated with the surface brightness.
- Theoretically and numerically it is very challenging to reproduce the TF zero point.

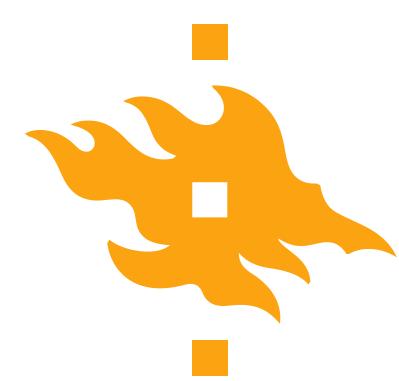




Formation of disk galaxies: Basic processes I

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 Σ for star formation is reached and a disk galaxy is born.





Disk angular momentum II

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

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

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

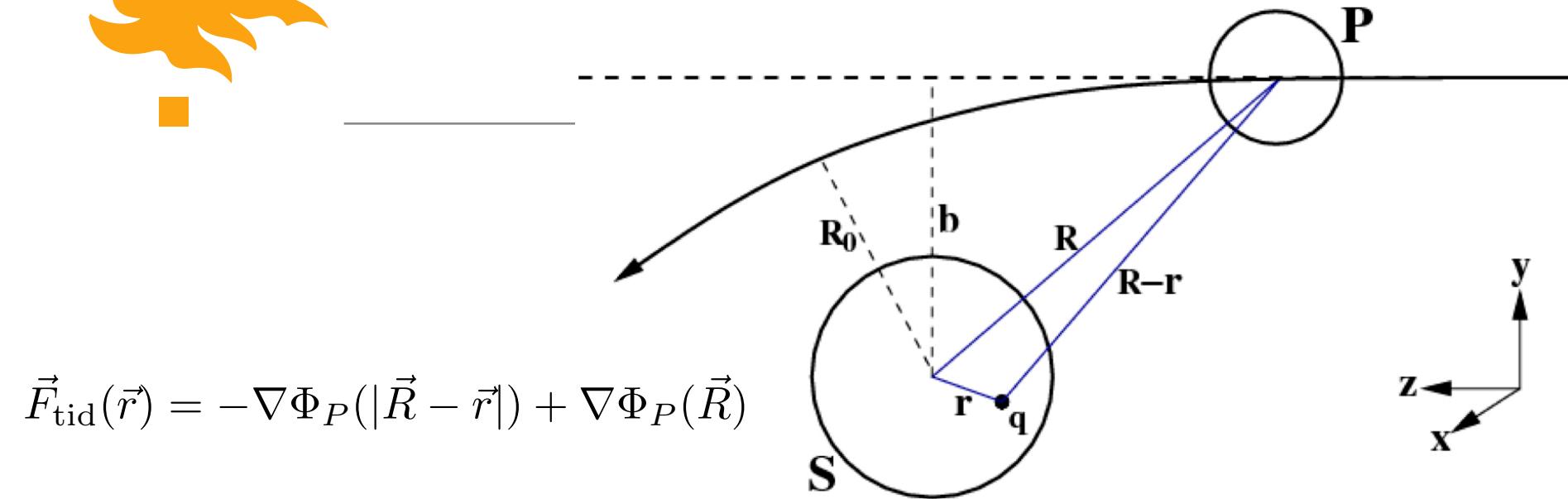
- Combining with the cosmological expression for R_{vir} :

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

- Using the values for the Milky way $V_{\text{vir}}=220 \text{ km/s}$ and $j_d=m_d$, $M_d=5 \times 10^{10} M_\odot$ and $R_d=3.5 \text{ kpc}$ results in $m_d \sim 0.01$ and $\lambda \sim 0.011$. $f_{\text{bar}} \sim 0.17$ only 6% of the baryons in the disk and the low spin of the MW halo makes it relatively rare (mean $\lambda \sim 0.04$).



11. Galaxy interactions



- Consider a body S which has an encounter with a perturber P with impact parameter b and initial velocity v_∞ .
- Let q be a particle (star) in S, at a distance $r(t)$ from the centre of S and let $\mathbf{R}(t)$ be the position vector of P from S.
- Since the gravitational force due to P is not uniform over the body S, the particle q experiences a tidal force per unit mass.



Impulse approximation I

- Taking the gradient of the potential and dropping the second term (constant acceleration) gives the tidal force per unit mass:

$$\vec{F}_{\text{tid}}(\vec{r}) = \nabla \Phi_P = d\vec{v}/dt$$

- Integrating the force over time gives the cumulative change in the velocity (see MBW page 546 for details):

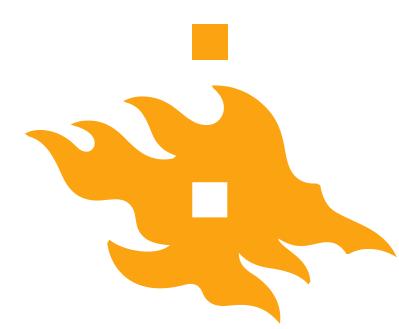
$$\Delta \vec{v} = \frac{2GM_P}{v_P b^2}(x, 0, -z)$$

- Substituting this expression for the total change in energy of S:

$$\Delta E_S = \frac{1}{2} \int |\Delta \vec{v}(\vec{r})|^2 \rho(r) d^3 \vec{r} = \frac{2G^2 M_P^2}{v_P^2 b^4} \int \rho(r) (x^2 + z^2) d^3 \vec{r} = \frac{2G^2 M_P^2}{v_P^2 b^4} M_S \langle x^2 + z^2 \rangle$$

- Assuming spherical symmetry: $\langle x^2 + z^2 \rangle = 2/3 \langle r^2 \rangle$

$$\Delta E_S = \frac{4}{3} G^2 M_S \left(\frac{M_P}{v_P} \right)^2 \frac{\langle r^2 \rangle}{b^4}$$



Impulsive heating

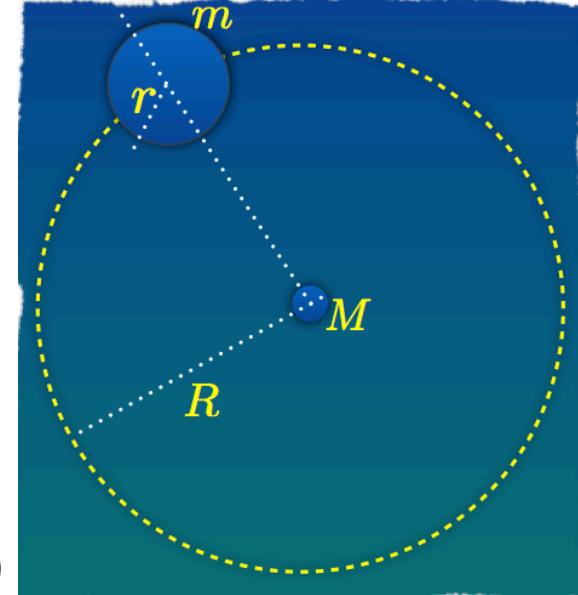
- In the impulse approximation the encounter only changes the kinetic energy of S and leaves the potential energy constant.
- After the encounter S is no longer in virial equilibrium and consequently S must undergo relaxation to re-establish virial equilibrium. Let K_S be the original pre-encounter kinetic energy, then:
 1. Virial equilibrium: $E_S = -K_S$.
 2. After encounter: $E_S \rightarrow E_S + \Delta E_S$.
 3. All new energy is kinetic: $K_S \rightarrow K_S + \Delta E_S$.
 4. After relaxation: $K_S = -(E_S + \Delta E_S) = -E_S - \Delta E_S$
- Relaxation decreases the kinetic energy by $2\Delta E_S$ and this energy is transferred to the potential energy, which becomes less negative. Hence, tidal shocks cause the system to expand and make it less bound.



Tidal stripping I

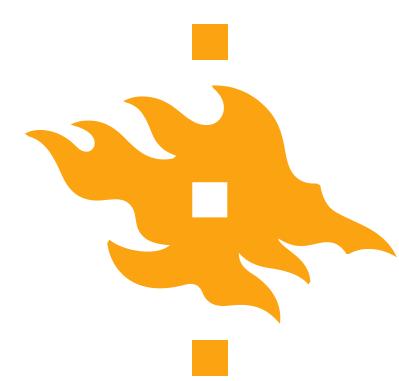
- Even in the general non-impulsive case, tidal forces can strip matter (=tidal stripping).
- Let us consider a mass m , with radius r , orbiting a point mass M on a circular orbit of radius R .
- Calculating the gravitational acceleration at the edge of m closest to the central mass M :

$$\vec{g}_{\text{tid}}(r) = \frac{GM}{R^2} - \frac{GM}{(R-r)^2} \simeq \frac{2GMr}{R^3} \quad (r \ll R)$$



- If the tidal acceleration exceeds the binding force per unit mass (Gm/r^2), the material at distance r from the centre of m will be stripped. This defines the tidal radius:

$$r_t = \left(\frac{m}{2M} \right)^{1/3} R$$



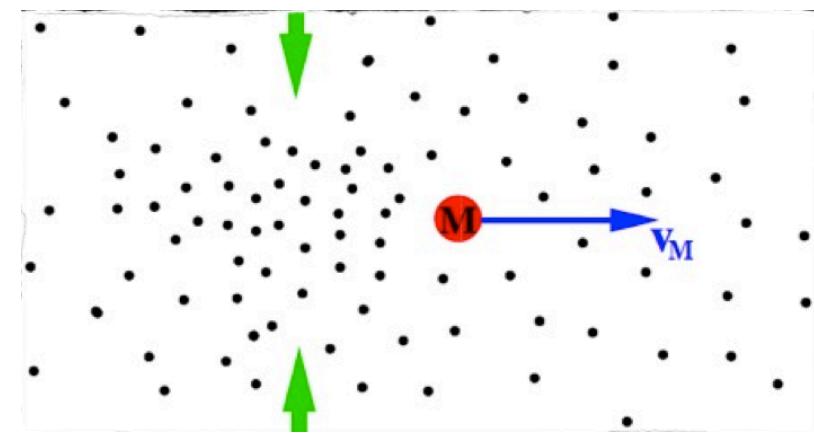
Dynamical friction: Intuitive pictures

1. Two-body interactions move the system towards equipartition.

$$m_1 \langle v_1^2 \rangle = m_2 \langle v_2^2 \rangle = m_3 \langle v_3^2 \rangle = \text{etc}$$

Since initially $v_s \sim \langle v_{\text{field}}^2 \rangle^{1/2}$ and $M_s \gg m$, the subject mass will loose energy and slow down.

2. Gravitational wake: The moving subject mass perturbs the distribution of the field particles creating a trailing density enhancement (“wake”). The gravitational force of this wake on M_s slows it down.



Gravitational wake formed by the action of dynamical friction.

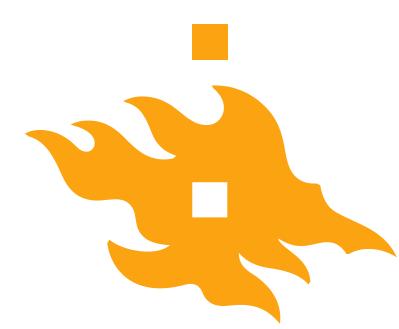


Orbital decay III

- Finally, using the following approximations $r_H/V_C \sim 1/[10H(z)] = 0.1t_H$ and that $\ln \Lambda \sim \ln(M_h/M_s)$ results in:

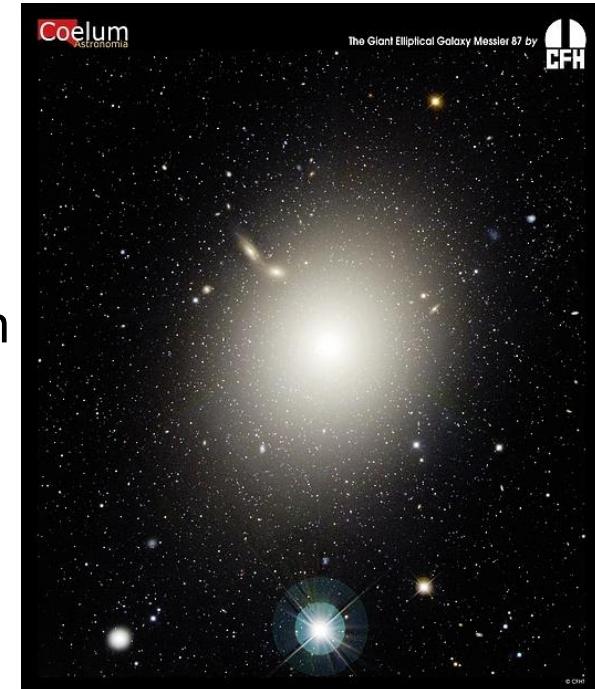
$$t_{df} \simeq 0.117 \frac{(M_h/M_s)}{\ln(M_h/M_s)} t_H$$

- This means that only systems with $M_s/M_h > 0.03$ experience significant mass segregation during the age of the Universe.
- The above formulas are based on questionable assumptions, which in general are not true. Haloes are not singular, isothermal spheres, orbits are not circular and tidal stripping implies mass loss.
- Also when the orbits are eccentric, dynamical friction may cause the orbit's eccentricity to evolve. The net result is that more eccentric orbits decay faster, see MBW pages 557-558 for details.



12. Ellipticals: Basic properties

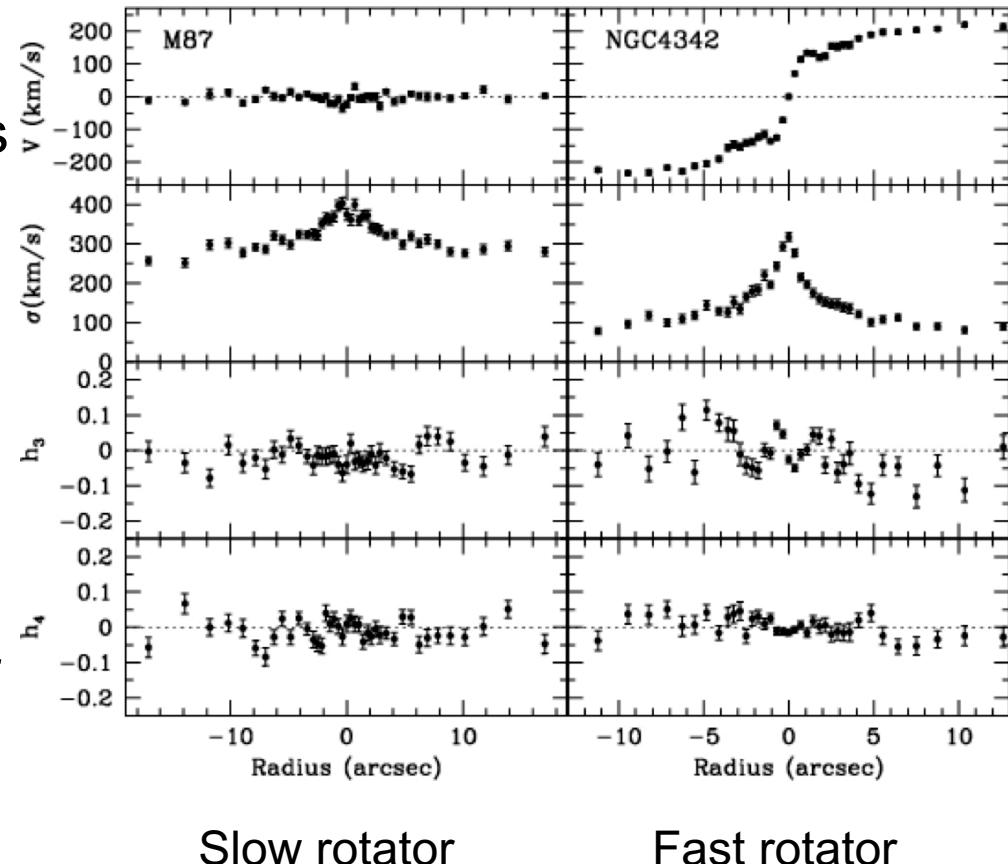
- In the 1970s it was assumed that ellipticals are relaxed systems, flattened by rotation and with isotropic velocity distributions. However, now we know that ellipticals are far more complex systems.
- The ellipticals can be divided into 3 classes:
 1. Bright Es ($M_B \leq -20.5$) are dominated by systems with little rotation, typically boxy isophotes and relatively shallow central surface brightness profiles.
 2. Intermediate Es ($-20.5 \leq M_B \leq -18$) seem to be supported by rotation, have disk-like isophotes and steep central surface brightness profiles.
 3. Faint Es ($M_B \geq -18$) reveal no, or very little rotation, and have roughly exponential surface brightness profiles.





Kinematic properties I

- Bright, boxy Es are slow rotators, supported by anisotropic velocity dispersions and are triaxial.
- Fainter, disk-like Es typically have much higher rotation velocities and are often consistent with them being purely rotationally flattened.
- LOSVDs are closed to Gaussian and the skewness of the profiles can be quantified using the Gauss-Hermite moments h_3 and h_4 .



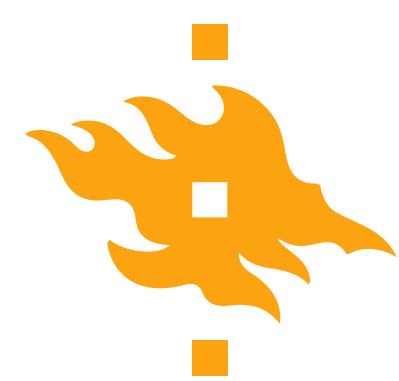


Dynamical modelling

- Modelling the dynamics of elliptical galaxies have two goals: 1) Constrain the mass-to-light ratio and 2) Constrain the orbital structure, or equivalently, the phase-space distribution functions of the elliptical.
1. Jeans Models: One way to constrain the mass distribution from the observed kinematics is to solve the Jeans equations relating the gravitational potential to the various intrinsic velocity moments. The largest uncertainty lies in the velocity anisotropy $\beta(r)$:

$$\frac{1}{\rho} \frac{d(\rho \langle v_r^2 \rangle)}{dr} + 2\beta \frac{\langle v_r^2 \rangle}{r} = - \frac{d\Phi}{dr}, \quad \beta(r) = 1 - \langle v_\vartheta^2 \rangle / \langle v_r^2 \rangle$$

2. Schwarzschild models: Assuming a stellar-mass-to-light ratio the deprojected light distribution is transformed into a stellar mass distribution. Then a larger number of orbits are calculated by numerical integration in order to find the model that best fits the derived gravitational potential.



Monolithic collapse

- In this scenario ellipticals form on a short time scale through collapse and virialization from idealized uncollapsed initial conditions.
- If the star-formation time scale is short compared to the free-fall time scale the collapse is effectively dissipationless.
- The main characteristic of this scenario is that the stars form simultaneously with the assembly of the final galaxy.
- This scenario was inspired by the fact that elliptical galaxies appear to be a remarkably homogeneous class of objects with uniformly old stellar populations.

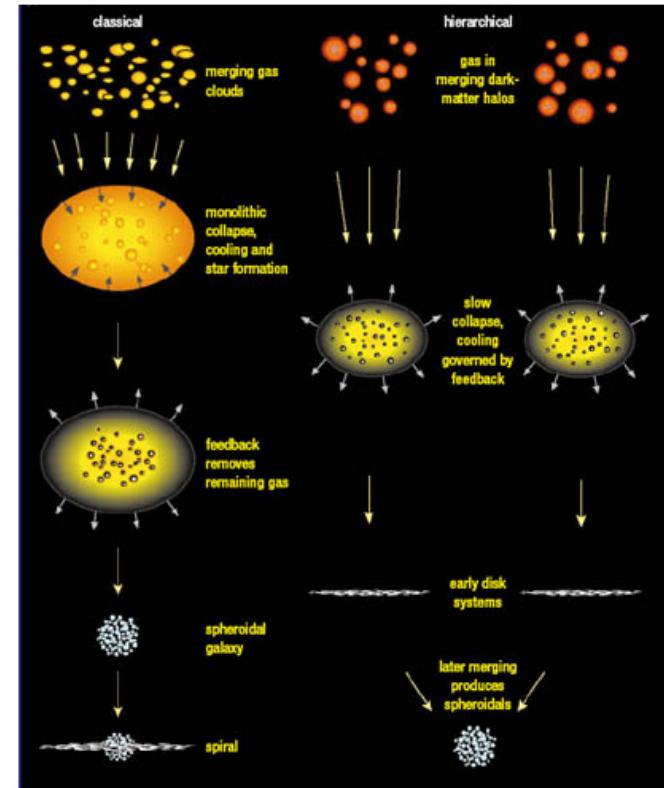
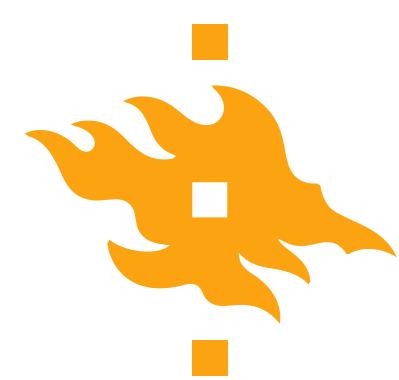
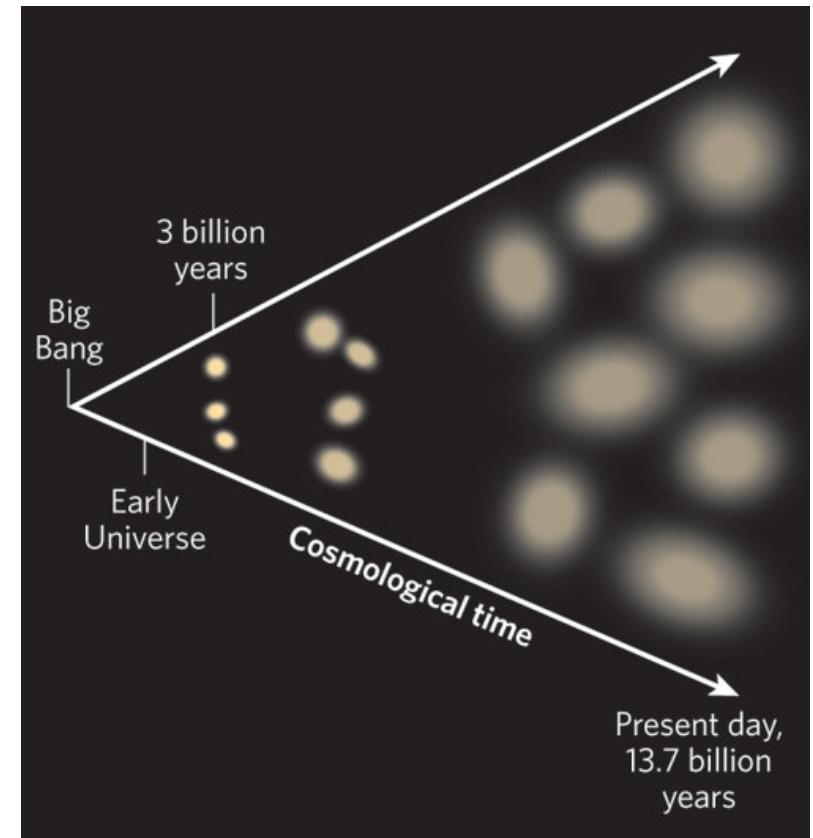


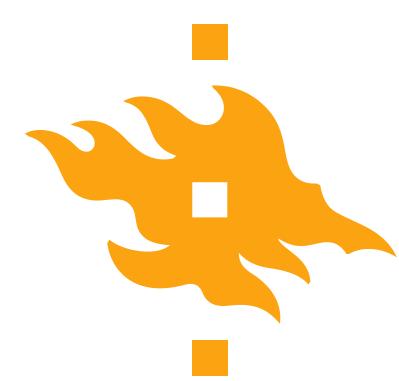
Figure 1: The two competing models of galaxy formation and evolution that could produce the galaxies we observe today. The classical 'top-down' or monolithic model is shown on the left. This involves the collapse of a large cloud over time. The hierarchical or 'bottom-up' model is shown on the right and involves successive mergers of small bodies.⁴



The merger scenario I

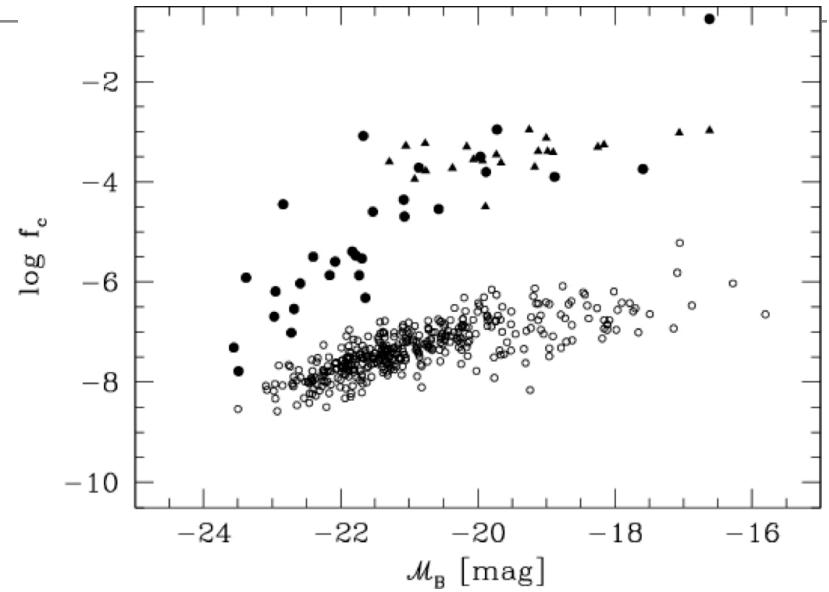
- In this scenario, an elliptical forms when two or more pre-existing and fully formed galaxies merge together.
- The main difference with respect to the monolithic collapse scenario is that the formation of the stars occurs before, and effectively independently of, the assembly of the final galaxy.
- The outstanding issues are 1) whether mergers of observed galaxies (not necessarily $z=0$ galaxies) can produce ellipticals and 2) whether the integrated merger rate as a function of progenitor properties and environment can reproduce the $z=0$ population.



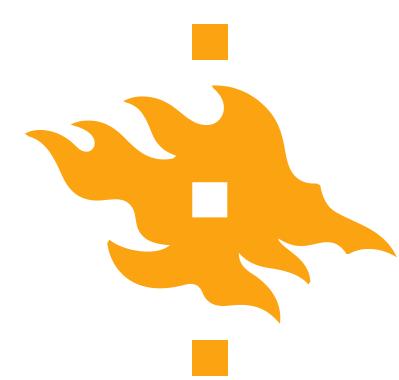


Observational constraints: The phase-space density

- The maximum value of the coarse-grained phase-space density of a collisionless system cannot increase during its evolution.
- The phase-space density in the central region of a galaxy can be estimated by dividing the central mass density by the volume occupied by an ellipsoid with its axes equal to the three velocity dispersions.
- The central phase-space densities of $M_B \geq -22$ Es are more than three orders of magnitude higher than those of observed disks.

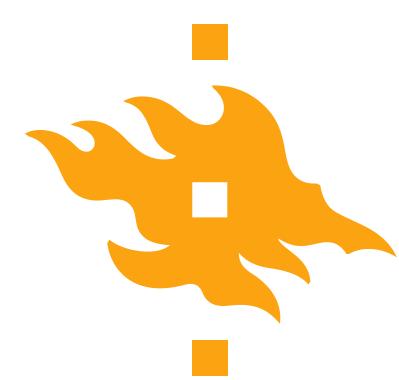


- Higher redshift disks were more compact (bulges). But most importantly mergers with **significant amounts of gas** can easily enhance the central phase-space densities to the observed values.



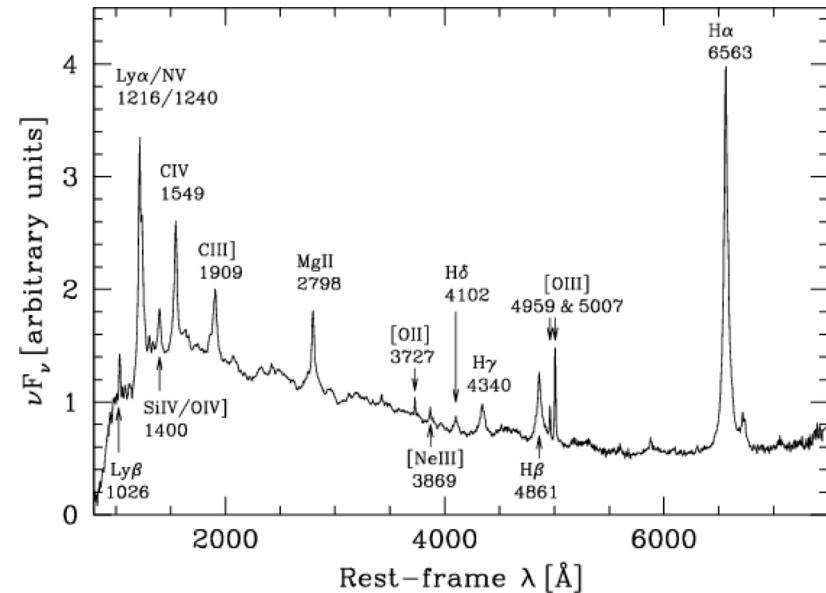
13. The population of Active galactic nuclei (AGNs)

- The temperatures of stars cover a relatively narrow range of $3,000 \text{ K} \leq T \leq 40,000 \text{ K}$. The spectra of normal galaxies is roughly the sum of the Planck spectra corresponding to these temperatures and thus the spectra are largely confined in the range between $\sim 4,000\text{-}20,000 \text{ \AA}$.
- A small (but important) fraction of galaxies have a spectral energy distribution that is much broader than expected from a collection of stars, gas and dust. These galaxies typically emit non-thermal radiation over the full wavelength regime from radio to X-rays.
- The non-thermal emission emanates from a very small central region, often less than a few parsecs across, which is called the AGN.
- Local number densities in units of Mpc^{-3} : Field galaxies: 10^{-1} , Luminous spirals: 10^{-2} , Seyfert galaxies: 10^{-4} , Radio galaxies: 10^{-6} , Quasars: 10^{-7} , Radio-loud Quasars: 10^{-9} .



Classification of AGNs

- Roughly speaking, an object is defined to be an AGN if one or more of the following properties are observed:
 1. A compact nuclear region much brighter than a region of the same size in a normal galaxy.
 2. Non-stellar (non-thermal) continuum emission.
 3. Strong emission lines.
 4. Variability in the continuum emission and/or in emission lines on relatively short time scales.



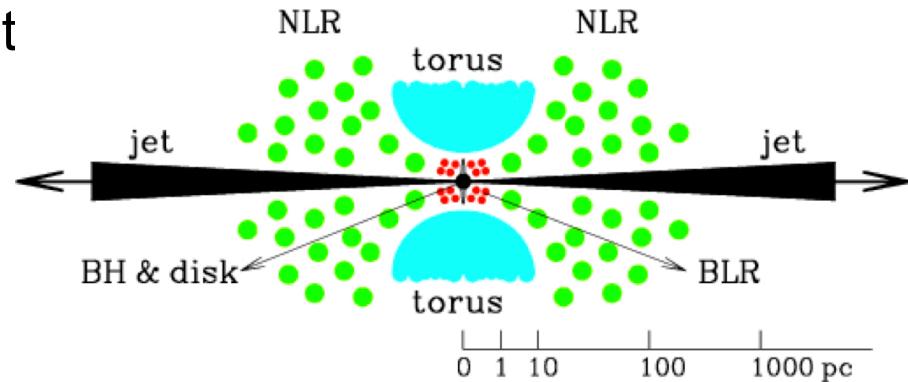
A composite spectrum of QSOs revealing the typical non-thermal continuum and permitted and forbidden [in brackets] lines.



Physics of the AGNs: The central engine I

- The small size of the emission region and the large amount of energy output suggest that this central engine is a supermassive black hole (SMBH).
- In addition to the SMBH, the standard model also assumes the existence of the broad-line region (BLR), narrow-line region (NLR), an accretion disk, torus and jets.
- The energy can be estimated by equating the radiation pressure with the gravity on the gas:

$$L_{\text{edd}} = \frac{4\pi G c m_p}{\sigma_T} M_{\text{BH}} \approx 1.28 \times 10^{46} \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \text{ ergs}^{-1}$$



$$P_{\text{rad}} = \frac{L}{4\pi r^2 c} \Rightarrow \mathbf{F} = \sigma_T P_{\text{rad}}(r) n_e(r) \mathbf{r}$$

$$\mathbf{F}_{\text{rad}} \leq F_{\text{grav}} = \frac{GM_{\text{BH}}\rho(r)}{r^2}$$



Emission-line regions and torus

- **Broad-line region:** The broad lines with velocity widths of the order of ~1000 km/s suggests that these lines are produced in a small inner region (≤ 0.5 pc) surrounding the accretion disk. The lack of forbidden lines indicates a high electron density (10^{10} cm $^{-3}$).

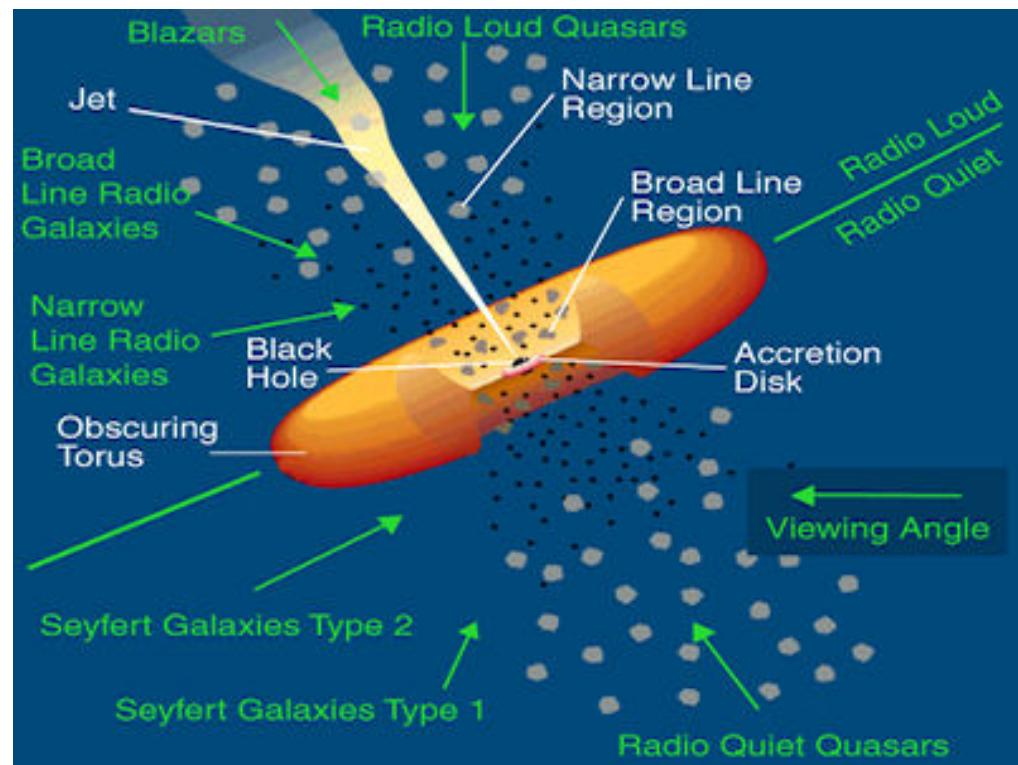
$$R \sim \frac{GM}{\sigma_v^2} \lesssim 0.5 \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \left(\frac{\sigma_v}{1000 \text{ kms}^{-1}} \right)^{-2} \text{ pc}$$

- **Narrow-line region:** Many AGNs reveal strong narrow lines with velocity widths of ~100 km/s. These lines are produced in a region of size ~50 pc around the central engine. To observe narrow lines the electron density must be $\leq 10^6$ cm $^{-3}$.
- **Obscuring Torus:** The innermost region is covered in an obscuring torus, which blocks our line-of-sight in type II AGNs. The torus must have a large column density of gas in order to block the BLR and strong X-ray and UV radiation. The torus might be clumpy.



AGN unification

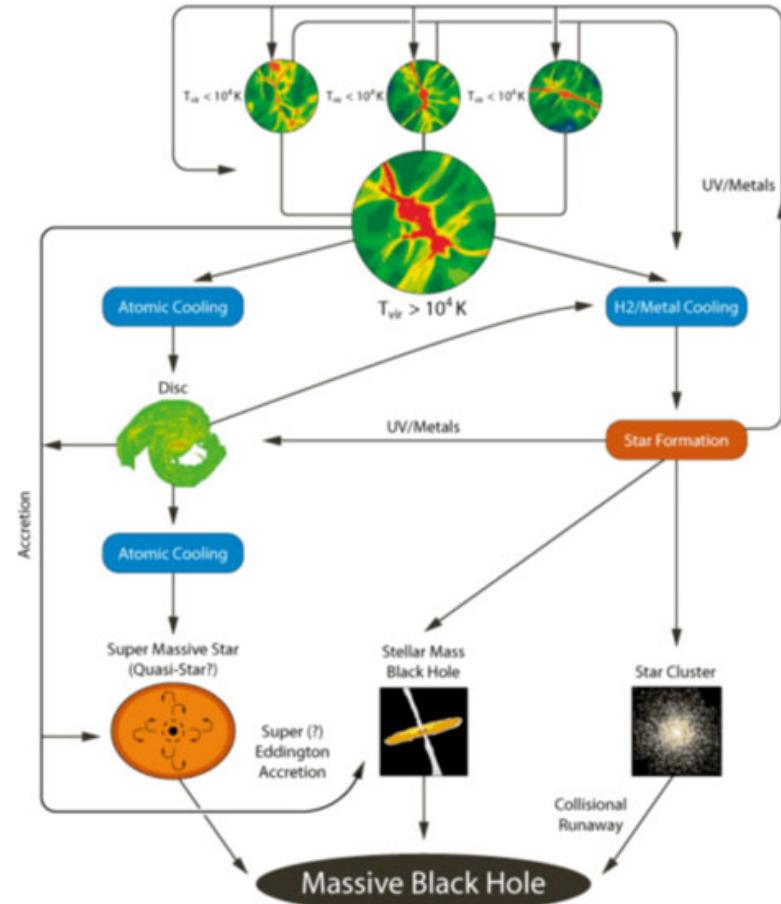
- In the AGN unification model all different types of AGNs are basically similar types of objects just viewed from different lines of sight.
- If we view the axisymmetric AGN from the side towards the obscuring torus we will observe a type II AGN, whereas we can directly view the inner BLR we classify the AGN as type I.
- Radio properties also change depending on viewing angle.





The formation and growth of SMBHs

- The formation mechanism of SMBHs at high redshift is uncertain. There are two leading models:
 1. **Pop III remnants:** The first stars in the Universe were presumably very massive ($\geq 50\text{-}100 M_{\odot}$). Black holes formed from the death of these stars should have masses of the order $\sim 10 M_{\odot}$.
 2. **Direct collapse:** There were no metals in the early Universe and it is possible that the inefficient cooling resulted in near isothermal collapse of $10^4\text{-}10^5 M_{\odot}$ gas clouds resulting in more massive SMBHs.





Formation of AGN I

- One important fact that any theory of AGN formation must take into account is the fact that quasars with $M_{\text{BH}} \sim 10^9 M_{\odot}$ at $z \sim 7$. The cosmic time at such a redshift is around $t \sim 0.5$ Gyr.
- Since black holes form in collapsed objects, the free-fall time scale of a virialized halo $t_{\text{ff}} \sim r_{\text{vir}} / V_c \sim 1/[10H(z)]$ must also be considered. At $z \sim 7$ t_{ff} is about 50 million years.
- If the growth of the SMBH is through radiative accretion the mass accretion can be written as:

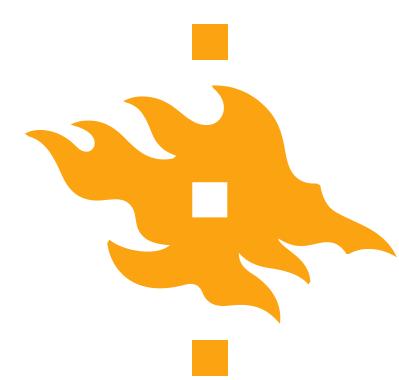
$$\dot{M}_{\text{BH}} = \frac{L}{\epsilon_r c^2} = \left(\frac{L}{L_{\text{edd}}} \right) \frac{M_{\text{BH}}}{\epsilon_r t_{\text{edd}}} \quad t_{\text{Edd}} = \frac{\sigma_T c}{4\pi G m_p} \approx 4.4 \times 10^8 \text{ yr}$$

- If L/L_{edd} and ϵ_r are independent of time: $M_{\text{BH}}(t) = M_{\text{BH},0} e^{t/t_{\text{BH}}}$
 $t_{\text{BH}} = (L/L_{\text{edd}})^{-1} \epsilon_r t_{\text{edd}} \approx 4.4 \times 10^7 (\epsilon_r/0.1) (L/L_{\text{edd}})^{-1} \text{ yr}$



AGNs and galaxy formation

- AGNs can release a huge amount of energy during their lifetimes,
$$\frac{dE}{dt} = \epsilon \dot{M}_{\text{BH}} c^2$$
 where $\epsilon = \epsilon_r + \epsilon_m$ is an efficiency factor the combines the radiative and mechanical feedback efficiencies.
- We can compare this energy output to the binding energy of the galaxy, which according to the virial theorem is $W \sim M_{\text{gal}} \sigma^2$:
$$\frac{E}{|W|} \sim \frac{\epsilon M_{\text{BH}}}{M_{\text{gal}}} \left(\frac{c}{\sigma} \right)^2$$
- $M_{\text{BH}}/M_{\text{gal}} \sim 10^{-3}$ and for $\sigma \sim 300 \text{ kms}^{-1}$, the ratio $E/|W| = 10^3 \epsilon$ indicating that the AGN energy can easily surpass the total binding energy of the galaxy.
- There are two outstanding questions: 1) What is the value of the feedback efficiency, ϵ and 2) how effective is the coupling of the feedback energy to the surrounding gas.

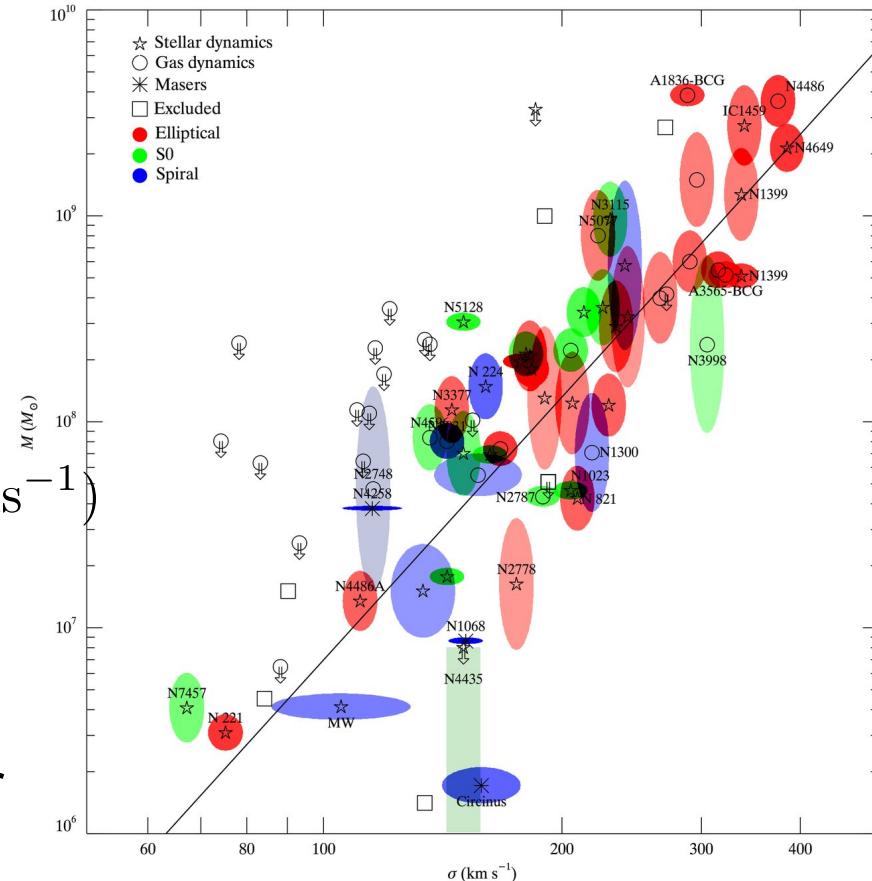


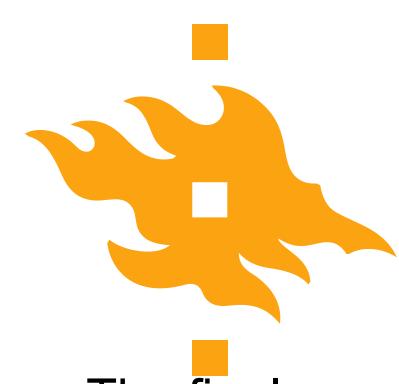
Co-evolution of SMBH and galaxies

- An important observed result of the co-evolution of SMBH and their host galaxies is the relatively tight $M_{\text{BH}}-\sigma$, which relates the mass of the supermassive black hole and the velocity dispersion (\propto mass) of the stellar bulge.

$$\log(M_{\text{BH}}/M_{\odot}) = 8.12 + 4.24 \log(\sigma/200 \text{ km s}^{-1})$$

- This means that the growth of the SMBHs and the bulge component of galaxies is linked, suggesting that they both grew together in mergers or in secular disk instabilities.





The final exam

- The final exam will be a home exam held on Friday 18.12 at 10.00-14.30. Note that exam starts at 10.00 sharp, when the exam sheet will be emailed to you. The answers are also returned by email to: Peter.Johansson@helsinki.fi
- The exam is a open book exam, thus you can use all of the course material in answering the questions. Especially the Sheet of Equations found on the course homepage will be useful during the exam.
- Have pens and a calculator at hand.
- The exam will consist of 5 questions, all of which you need to answer. The questions will be mostly mathematical problems, but you might need to answer some questions also in writing.
- The questions will be quite similar in spirit to the problem set questions, so go through them carefully.
- The maximum points from the exam is $5 \times 6 = 30$ points, and the problem sets provide a maximum of 6 extra points. To pass the course (1/5) you need a minimum of 13 points and for the highest grade (5/5) you need >26 points.