



a hands-on introduction

Copyright © 2017 Andrey Kravtsov

KICP PRESS

<HTTP://A-KRAVTSOV.GITHUB.IO/A304S17/>

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at <http://creativecommons.org/licenses/by-nc/3.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

First printing, April 2016

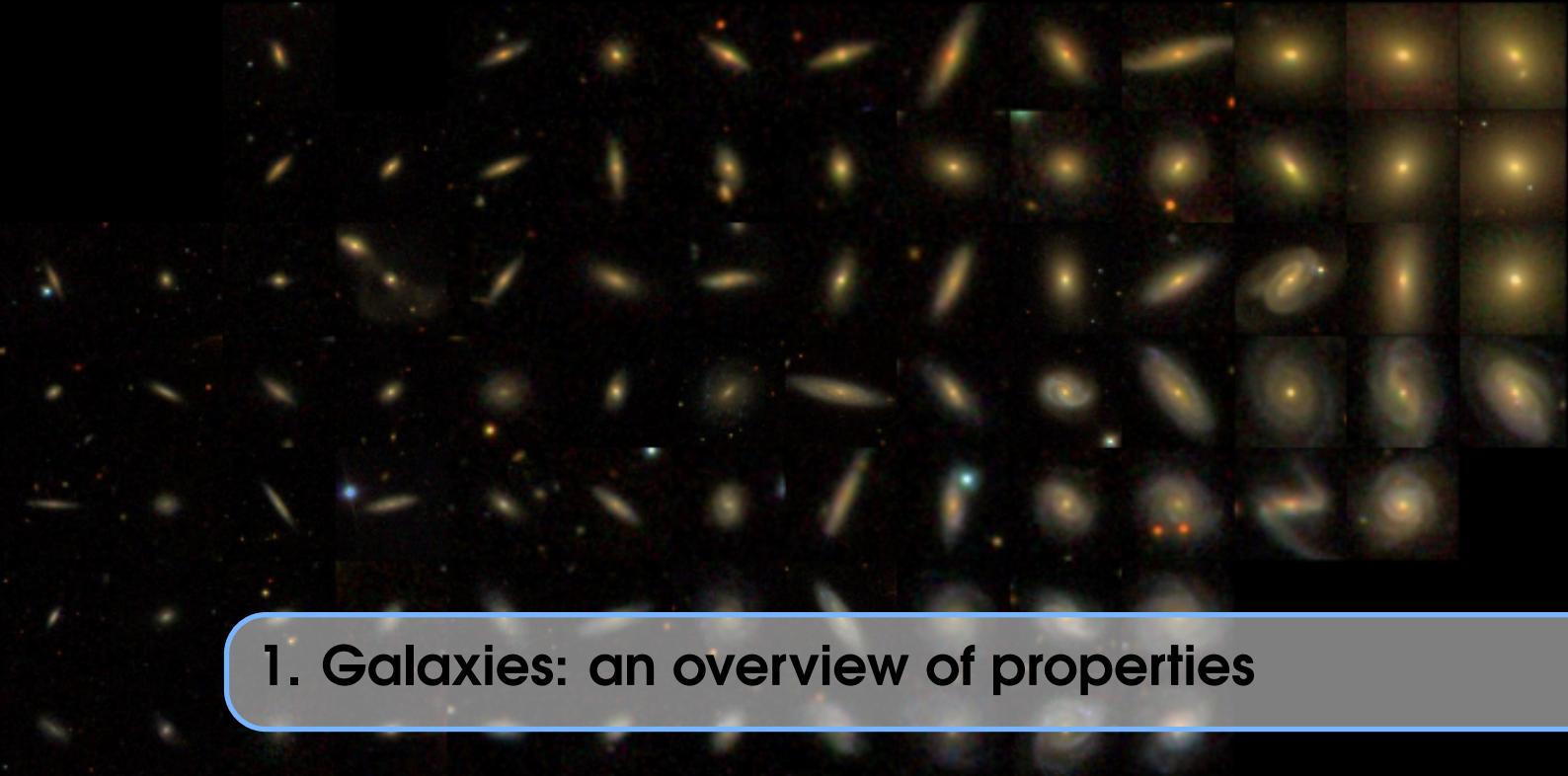


Contents

1	Galaxies: an overview of properties	5
1.1	Morphology	5
1.2	Galaxy magnitudes and luminosities	7
1.3	Surface brightness profiles	8
1.3.1	Exponential, de Vaucouleurs, and Sérsic profiles	9
1.3.2	Surface brightness profiles of spheroidal galaxies	11
1.3.3	Surface brightness profiles of late type disk galaxies	11
1.3.4	Multi-component descriptions of the surface brightness profiles	12
1.4	Galaxy magnitudes and luminosities	13
1.4.1	Aperture, Holmberg, and Petrosian magnitudes	13
1.4.2	Model magnitudes	14
1.5	Luminosity distribution of galaxies	14
1.5.1	Malmquist bias	16
1.5.2	Galaxy luminosity function	17
1.6	Galaxy spectra and colors	19
1.6.1	k -correction	19
1.6.2	From luminosity to stellar mass: stellar population synthesis	22
1.7	Structural properties of galaxies and their relation to morphology	24
1.7.1	Galaxy surface brightness	24
1.7.2	Light concentration	25
1.7.3	Vertical structure of disks	28
1.8	Galaxy clustering and large-scale structure.	28
1.9	Evidence for dark matter in galaxies and galaxy clusters	31
1.9.1	Disk of the Milky Way.	31
1.9.2	Evidence for dark matter in clusters of galaxies	32
1.9.3	Relative motion of M31 and Milky Way	33

1.9.4	Rotation curves of galaxies	34
1.9.5	Evidence from disk stability	36
1.9.6	Evidence for dark matter from the cosmic microwave background	36
1.10	Additional reading	36
1.11	Ideas for exercises and explorations	37
2	Galaxy formation	39
2.1	Inflationary origin of the seed density perturbations	40
2.2	Evolution of density perturbations in the linear regime	41
2.2.1	Linear evolution of perturbations in the radiation domination regime	42
2.2.2	Evolution of perturbations smaller than horizon	42
2.2.3	Summary perturbation evolution in the linear regime	43
2.3	Evolution of perturbations in the non-linear regime	44
2.3.1	The spherical tophat collapse model	44
2.3.2	Nonlinear mass and peak height	48
2.3.3	The ellipsoidal collapse model	50
2.3.4	Mass function of halos	50
2.3.5	Collapse of peaks of non-uniform density	52
2.3.6	Mass accretion histories of halos	54
2.3.7	Characteristic “formation” time of halos	56
2.3.8	Secondary infall model	57
2.3.9	“Virial” mass definition	58
2.3.10	Collapse of real dark matter peaks	59
2.3.11	Density profiles of CDM halos	61
2.3.12	Physical model for CDM halo profiles and concentrations	63
2.3.13	Halo angular momentum	64
2.3.14	Scaling relations of halos	65
2.4	Baryonic processes affecting galaxy formation	66
2.4.1	The filtering scale in baryon perturbations after recombination	66
2.4.2	Thermodynamics of baryon plasma	66
2.4.3	Baryon cooling and condensation	68
2.4.4	Radius of angular momentum support	68
2.5	Additional reading	69
A	Cosmological distances	71
A.1	<i>h</i>-scalings	72
A.2	Redshift space distortions	72
A.3	Cosmological distances	73
A.3.1	Comoving line-of-sight distance	73
A.3.2	Angular diameter distance	74
A.3.3	Luminosity distance	75
A.4	Apparent and absolute magnitudes and <i>k</i>-correction	75
A.5	Surface brightness	76
A.6	Comoving volume	77
A.7	Lookback time	77

B	Supplementary material	79
B.1	Equilibrium equations	79
B.2	Power spectrum definition	80
B.3	Gaussian peaks	81
B.4	Linear perturbation evolution equation	81
C	python info	85
C.1	Ipython notebooks	85
C.2	colossus	85
C.3	AstroML	85
D	Data access info	87
D.1	SDSS DR8 main galaxy sample	87
D.1.1	Petrosian magnitudes and sizes	87
D.1.2	cmodel magnitudes	88



1. Galaxies: an overview of properties

The term galaxy is derived from the Greek word “galaxias” ($\gamma\alpha\lambda\alpha\xi i\alpha\zeta$) which means “milky” and which was used to describe the milky band on the sky formed by stars within our Galaxy — the Milky Way.¹ Our Galaxy is only one out of two galaxies visible in the northern hemisphere and only one out of four visible to naked eye over the entire sky.

Although many galaxies have been cataloged after the advent of observations with telescopes as “nebulae,” the name stemming from their nebulous appearance, their extragalactic nature was not understood until the 1920s. In fact, the name “nebulae” has persisted as the dominant names in astronomical literature through the mid-1930s. For example, the title of the seminal Hubble (1929) paper presenting what we now call the Hubble law was “A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae,” which reflects the (then) new knowledge that galaxies are extra-galactic, but still uses the word “nebulae.”²

1.1 Morphology

“From a study of all available photographs with large reflectors, it is found that the characteristic features of non-galactic nebulae are rotational symmetry about dominating nuclei. About 2.5% are irregular, lacking both these features, and form a homogeneous class of which the Magellanic Clouds are the most conspicuous members. The regular nebulae fall into a progressive sequence ranging from globular masses of unresolved nebulosity through lenticular forms to the open spirals with arms swarming with stars. This purely observational sequence conforms very closely to Jeans’ theory of the origin and evolution of spirals.” – Edwin Hubble (1926, PASP)

Even before the extra-galactic nature of galaxies was proven beyond any doubt, people classified them based on their appearance or *morphology* using hand-sketches of how they appeared through a telescope.³. William Herschel introduced the notion of “spiral nebulae” in 1755 based on some of his sketches. In the same year, the German philosopher Immanuel Kant, building upon earlier idea of Thomas Wright about the

¹The connotation of the apparent Galaxy on the sky with milk is surprisingly universal and is present in most cultures. The two notable exceptions are the Ukrainian legend about a salt trader’s way and the Cherokee legend about scattered cornbread crumbs.

²Hubble’s 1934 book is still titled “The realm of the nebulae” (Hubble, 1934). The term “galaxies” only truly took off and displaced “nebulae” after 1940. See, e.g., the use of word “galaxies” in books as a function of time through Google’s ngram-viewer.

³For example, see sketches by William and John Herschel, Lord Rosse in the 19th century

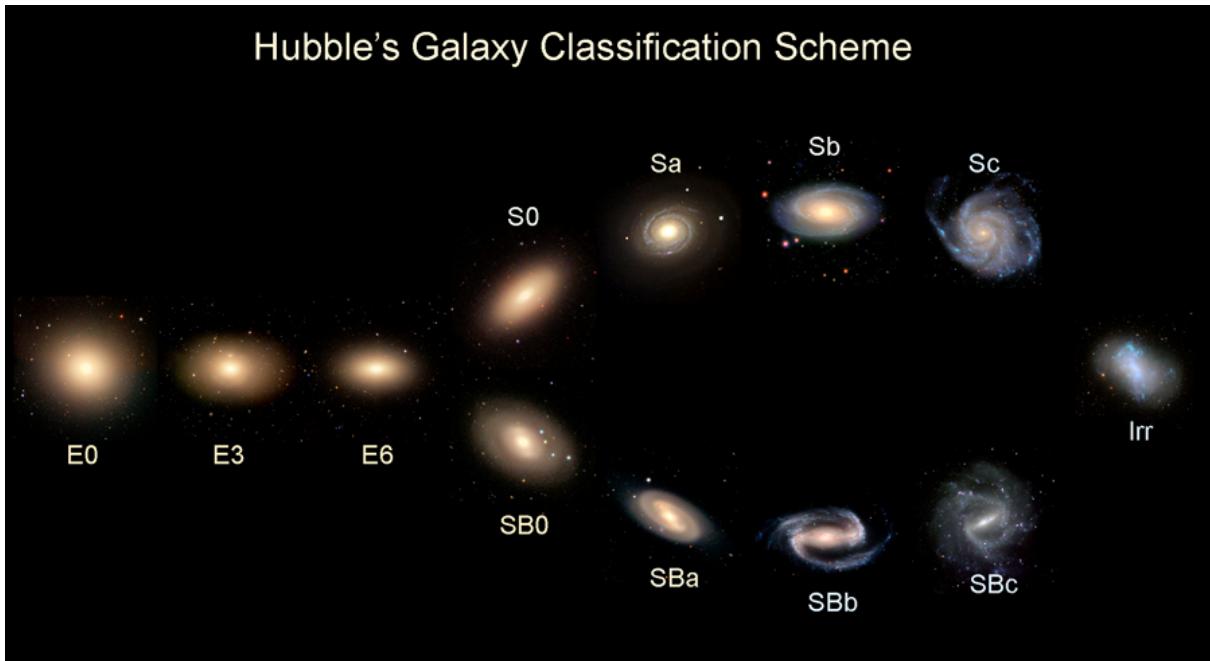


Figure 1.1: Hubble’s famous tuning fork (Hubble, 1926b). This diagram was constructed by the Galaxy Zoo team using SDSS images.

Milky Way as a disk of stars, put forth the idea that spiral nebulae were “island universes” of stars similar to our own Milky Way. The popularity of morphological studies and classification schemes is probably due both to the human tendency to classify as well as the relative success of the morphological classification approach in both the biological sciences and in the classification of the photometric properties of stars. The popularity of Galaxy Zoo can attest that morphological classification comes naturally to the human brain, which evolved to see patterns even where there are none and suffers from curious idiosyncrasies (Land et al., 2008) when it comes to classifying patterns.

The most persistent morphological classification scheme was developed by Hubble (Hubble, 1926b) and is usually represented via the “tuning fork diagram” (see Fig. 1.1). Hubble developed this scheme from the early 1920s to 1936 and it was definitely influenced by earlier ideas of Reynolds, who developed a classification based on the changing bulge-to-disk ratio of spirals. Later, more elaborate schemes have also been developed by de Vaucouleurs, van den Bergh, and others (see Sandage, 2005, for a comprehensive review), but these schemes are less widely used.

Hubble’s classification, in particular, was motivated by the evolutionary scenario advocated by James Jeans (Jeans, 1919), in which spheroidal galaxies evolved into spirals by undergoing a sequence of transformations through intermediate morphologies. The bulge-less late-type spirals in this scheme were envisioned to be on the verge of break up into pieces that would then fly apart. The relative stages of this evolutionary sequence still reflected in the widely used terms “early” and “late” galaxies.

There is definitely value to classification (see, e.g., Sandage, 2005, for discussion) and sometimes classification does map onto key physical properties of the classified systems (e.g., the original spectral classification of stars mapped onto the sequence of effective temperatures of stars). When it comes to galaxies, none of the classification schemes have mapped onto any particular evolutionary scenario or physical processes, but they did provide a language for astronomers to talk about when they discuss galaxies. You should know this language because it will not go away any time soon.

The minutiae of morphological classification, such as the details of spiral structure, are probably not that important. First, they depend on the wavelength. Second, the morphology of galaxies becomes increasingly irregular and unclassifiable at higher redshifts (see Fig. 1.2) and some of the morphological features are definitely transient features that can be triggered through tidal interactions (e.g., M51). Third, robust

classification is difficult and requires sophisticated methods, such as neural networks, or human classifiers (the problem that was solved to some extent by the "citizen scientist" movement). This becomes increasingly difficult in the era of very large galaxy surveys. In the SDSS, for example, appearance of a galaxy becomes increasingly "fuzzy" beyond $\sim 200 - 300$ Mpc and details of morphology are difficult or impossible to discern. However, some key properties of these galaxies, such as color or concentration, can still be measured reliably. This is the reason why such quantities are increasingly used in lieu of morphology.

Nevertheless, the overall morphology of an individual galaxy and presence of structures, such as bulges, bars, or prominent spiral arms *do* tell us useful information about the formation history or current dynamical state of the system. Thus, studies of galaxy morphology and how these features relate to systematic variations of galaxy properties continue unabated (e.g., Huertas-Company et al. 2015). The challenge, however, is that classification of large numbers of galaxies (by either human or automated computer algorithms) into fine morphological sub-classes is difficult for shallow surveys and/or distant galaxies. Thus, both automated and "citizen science" mass classifications use relatively coarse-grained, broad classes of elliptical, lenticular, and spiral classes (e.g., Huertas-Company et al. 2011, Kuminski and Shamir 2016).

It makes sense then to also examine other "broad brush" properties of galaxies, such as luminosity, light profiles, surface brightness, concentration, colors, etc. We will examine these properties and their mutual correlations as well as their correlations with morphology next.

We will start with an overview and exploration of the basic galaxy properties using the Sloan Digital Sky Survey (SDSS) and other data sets. One way to get a feel for these properties is to try to construct your own version of the Hubble's tuning fork diagram by finding random examples of galaxies of each morphological class and examining their key properties: spectra, color, luminosity, size, and bulge-to-disk ratio, as well as useful quantities derived from these, such as surface brightness. For example, even a quick examination of the tuning fork diagram in Figure 1.1 shows that the sequence does correspond to a regular trend in color of galaxies. The trend in color means a trend in spectral shapes, which, in turn, implies a trend in properties of stellar populations (age, metallicity, dust, etc.).

1.2 Galaxy magnitudes and luminosities

Astronomers use a logarithmic magnitude scale to quantify the apparent and intrinsic brightness of sources, which is akin to the historical visual magnitude system based on the roughly logarithmic sensitivity of our eyes.⁴ A number of different magnitude systems exist in the literature, especially in older papers. Modern magnitude systems, e.g. in SDSS, are based on the AB-magnitude system (Oke and Gunn, 1983), which is described in detail in Fukugita et al. (1996).⁵

In practice magnitudes are defined using light from a limited range of wavelengths, defined by a filter. In the AB system this is given by (Fukugita et al., 1996):

$$m_f = -2.5 \log_{10} \frac{\int f_v S(v) d(\log_{10} v)}{\int S(v) d(\log_{10} v)} - 48.6, \quad (1.1)$$

where f_v is the specific flux per unit frequency.

The SDSS filter transmission curves, S_λ , can be found [here](#). SDSS database reports spectra as the flux density per unit wavelength at a given λ , f_λ , instead of f_v . Converting from f_v to f_λ ($v = c/\lambda$, $dv = -cd\lambda/\lambda^2$, so $f_v = \lambda^2 f_\lambda / c$), the magnitude in a given filter can be expressed as (see, e.g., eqs 2 and 3 in Smolčić et al., 2006):

$$m_f = -2.5 \log_{10} \left[\frac{10^{19.44}}{c} \frac{\int f_\lambda S(\lambda) \lambda d\lambda}{\int S(\lambda) \lambda^{-1} d\lambda} \right]. \quad (1.2)$$

The flux density, f_λ , has a simple meaning, but its measurement is often non-trivial in practice because it requires integration of photons over some area on the sky. For point sources like stars and quasars the total

⁴The actual sensitivity of the human eye is complicated and is currently a subject of continued study. It may be closer to a square root response than to a logarithmic response.

⁵See also online info [here](#).

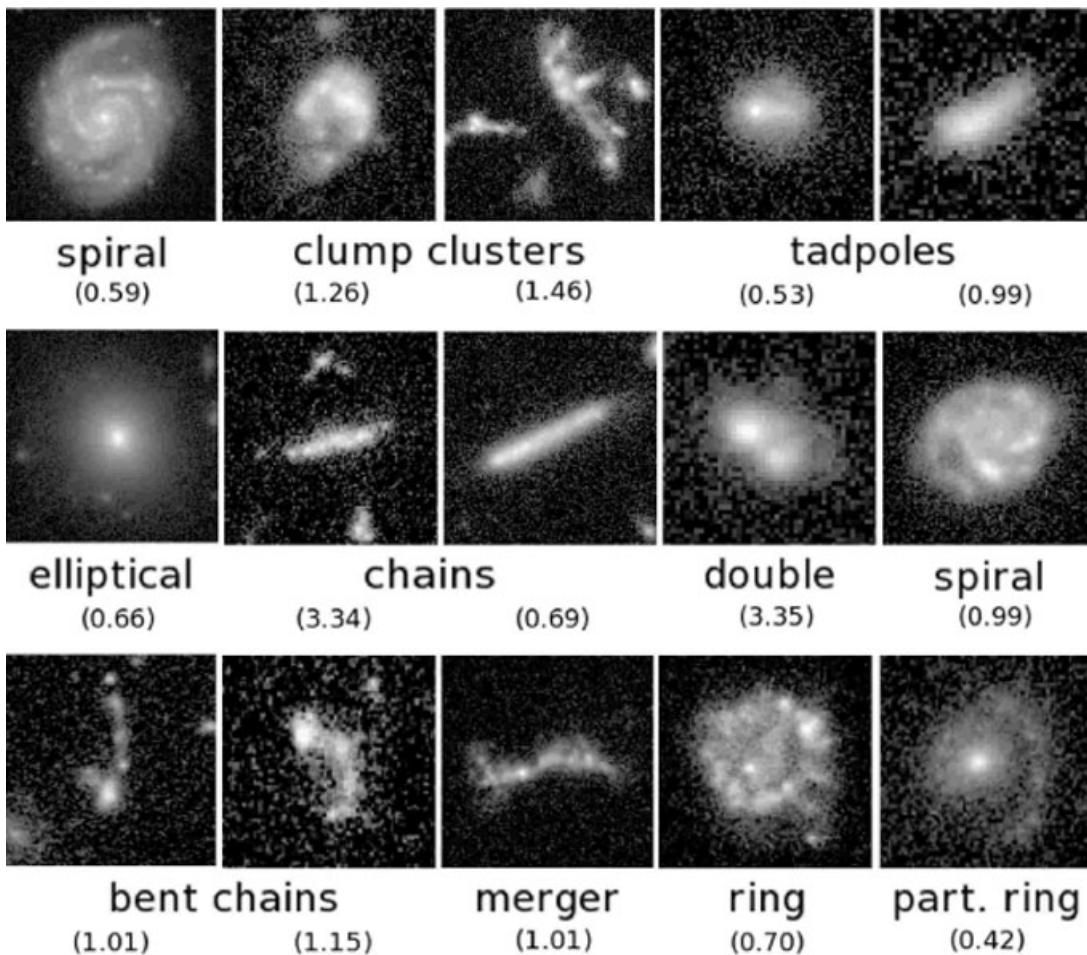


Figure 1.2: Example of morphological classification at higher redshifts: z is indicated in brackets under morphophological classification under each galaxy. At higher redshift we mostly observe light from shorter wavelength and galaxies are in earlier stages of their evolution. The fraction of "irregular" systems that do not fit into the main branches of the tuning fork grows steadily with increasing redshifts (although nice disks are found at least to $z \sim 3$). Here is an image with some new morphology "classes" at intermediate and higher z from (Buta, 2011, see here for the online version).

flux can simply be collected from an area that encloses the telescope point spread function, although for stars in crowded regions even this task is non-trivial. Galaxies, however, are intrinsically extended and diffuse (as discussed earlier, this is the reason they were dubbed “nebulae”) and collecting flux associated with galaxies requires either defining a galaxy boundary or devising some way to calculate the total flux.

As we will see below, different ways of doing this may result in significantly different estimates of galaxy intrinsic luminosities. Thus, although this subject may seem technical it is actually quite important to understand how total flux is estimated for galaxies. To understand how this is done or how it should be done, we need to discuss the radial distribution of light around galaxies.

1.3 Surface brightness profiles

Galaxy surface brightness is defined as the amount of light that we receive from a given area of the galaxy. This quantity makes sense for extended objects where we can compare amount of light we get from an area from the total amount of light we can get from a point source (see, e.g., S 5.1.2 in the SG book). Even if we don't know the distance to a galaxy we can compute the surface brightness in units of magnitude per square

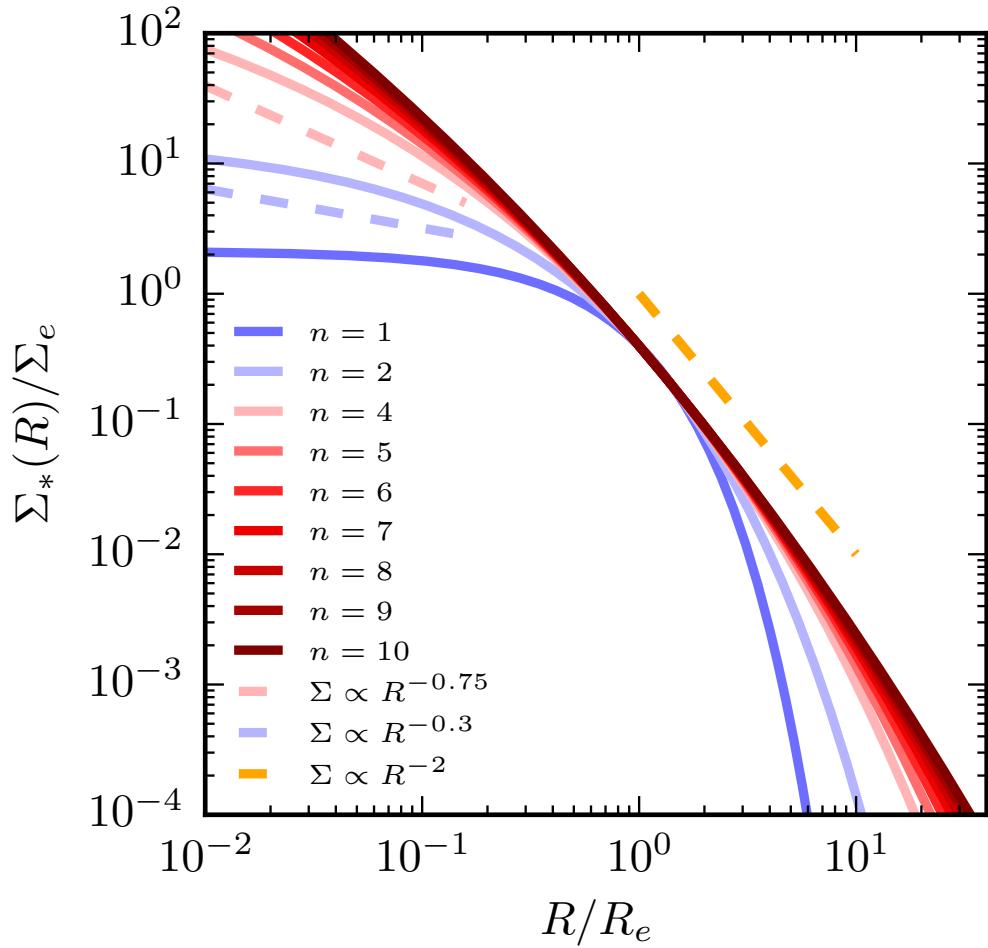


Figure 1.3: The Sérsic surface density profiles for $n \in [1, 10]$ along with three power laws for illustration.

arcsec - i.e., the flux in the r -band magnitudes we receive from a solid angle $d\Omega$ in arcsec²: ⁶

$$\mu_f = -2.5 \log_{10}(f_f/d\Omega) + \text{const} = m_f + 2.5 \log_{10} d\Omega, \quad (1.3)$$

where f_f is the flux collected from the solid angle on the sky $d\Omega$ (usually, in square arcseconds). The choice of $d\Omega$ depends on one's goals. For example, we can define $d\Omega$ to be an annulus of radius R and thickness ΔR , as is done in the estimates of the Petrosian magnitudes. Or we can estimate surface brightness in equal size patches of a given $d\Omega$.

Note that we can always express the flux f_f in equation 1.3 in terms of intrinsic luminosity in the same filter: $f_f \propto L_f/d^2$, where d is galaxy distance. Thus, μ_f and $\Sigma_f = L_f/\mathcal{A}$, where \mathcal{A} is a physical area (e.g., in kpc²), are tightly related (see eq. A.25 in §A.5 for the exact relation) and we can talk about μ_f and Σ_f interchangeably. Often, L_f is also converted into an estimate of stellar mass of the underlying stellar population stellar using some type of population modelling, as discussed below. In these cases we talk about stellar surface density, Σ_* .

1.3.1 Exponential, de Vaucouleurs, and Sérsic profiles

The measured surface brightness profiles of galaxies are often parameterized using simple functional forms. The most frequently used forms are the exponential profile which tends to provide a better fit to the surface

⁶Why is it meaningful to compute and consider surface brightness in such units?

brightnesses of disk galaxies and the de Vaucouleurs profile (de Vaucouleurs, 1953b) that tends to provide a better fit to profiles of spheroidal systems.⁷

The exponential and de Vaucouleurs profiles are specific cases of a more general Sersic (1968) profile:

$$\Sigma(R) = \Sigma_e \exp \left\{ -b_n \left[\left(\frac{R}{R_e} \right)^{1/n} - 1 \right] \right\}, \quad (1.4)$$

where Σ_e is the surface brightness at the effective radius R_e that encloses half of the total light from the model. The constant b_n is related to n – the Sérsic index – that controls the overall shape of the profile. Its detailed properties are well described by Graham and Driver (2005).⁸ In particular, $b_n = 1.9992n - 0.3271$. This is an approximation, but a very accurate one. The profile above is sometimes expressed as

$$\Sigma(R) = \Sigma_0 \exp \left[- \left(\frac{R}{R_d} \right)^{1/n} \right], \quad (1.5)$$

where $\Sigma_0 = \Sigma_e \exp(b_n)$ and $R_d \equiv R_e/b_n^n$ is the scale length.

If profiles are constructed using magnitudes rather than surface brightness or intensity, the corresponding expression is:

$$\mu(R) = \mu_e + \frac{2.5b_n}{\ln(10)} \left[\left(\frac{R}{R_e} \right)^{1/n} - 1 \right], \quad (1.6)$$

where R_e is the half-light radius and n is the Sérsic index.

The exponential profile is given by $n = 1$ case:

$$\mu(R) = \mu_e + \frac{2.5b_1}{\ln(10)} \left[\frac{R}{R_e} - 1 \right], \quad (1.7)$$

where $b_1 = 1.678$. R_e defined here encloses half of the total light of the profile. Note that this linear increase in magnitude per area corresponds to an exponential decrease in luminosity per area, thus the name. The de Vaucouleurs profile is given by $n = 4$:

$$\mu(R) = \mu_e + \frac{2.5b_4}{\ln(10)} \left[\left(\frac{R}{R_e} \right)^{1/4} - 1 \right], \quad (1.8)$$

where $b_4 = 7.669$. Here too R_e encloses half of the light is enclosed within R_e . Note that in the figures below, surface brightness is plotted in terms of surface mass density; these are related to μ via eq. A.25 in §A.5.

Although pure exponential and de Vaucouleurs profiles *do* provide a good description of the surface brightness profiles of some galaxies, examination of large samples shows that they are over simplifications of the typical profiles (see Figure 1.4 below).

The approximate 3D deprojected profile of the projected Sérsic profile is discussed in Lima Neto, Gerbal, and Márquez (1999). A good approximation for the 3D profile that gives the Sérsic profile in projection is given by the Einasto (1965) profile that is the second most commonly used profile to describe the radial density profiles of dark matter halos in simulations of structure formation (see Merritt et al., 2006, for detailed information about this profile). Analytical functions of the 3D profiles corresponding to exponential and de Vaucouleurs profiles are also discussed by Pitts and Tayler (1997) and Hernquist (1990), respectively. Note though that this deprojection of the exponential profile makes sense for a spherical distribution, not for disk distribution of stars.

⁷Hubble (1930) proposed another form to describe profiles of spheroidal galaxies, given by $\mu \propto 5 \log(1 + R/R_e)$. This form does fit many *bright* spheroidal galaxies quite well, but is too shallow in the central regions for many other early type systems (see Figure 1.3).

⁸See here for version online.

Figure 1.3 shows Sérsic profiles for $n \in [1, 10]$. Note the main difference between $n = 1$ and $n = 4$ profiles is in the central regions, where the exponential profile is shallow, while the de Vaucouleurs profile is steep. At the largest radii the exponential profile becomes steeper due to the higher power of R/R_e in the exponent. As we will see below, profiles of most galaxies are not traced to sufficiently large radii for this difference to be apparent. Note, also, that there is a fairly wide range of intermediate radii (nearly a decade), where the exponential and de Vaucouleurs profiles are close to each other. For $n > 6$, the profile shape changes only mildly with increasing n . This property sometimes leads to fit degeneracies, where best fit values of n are driven to large values. At large R , the profiles are close to $\Sigma \propto R^{-2}$ independent of n . This asymptotic behavior can be understood analytically in the limit of large R and n - see the last paragraph in S 2.1 of Graham and Driver (2005).

We will often use the radii containing half, R_e , of total galaxy light. For the Sérsic profile:

$$L(< R_{1/2}) = L_{\text{tot}}/2 = 2\pi n \Sigma_e R_e^2 \frac{e^{b_n}}{(b_n)^{2n}} \gamma(2n, x),$$

where $x = b_n(R/R_e)^{1/n}$, $\gamma(2n, x)$ is the incomplete gamma function:

$$\gamma(2n, x) = \int_0^x e^{-t} t^{2n-1} dt$$

and L_{tot} is the total light given by integration of the profile over all radii - it can be obtained by replacing $\gamma(2n, x)$ with the $\Gamma(2n)$ in the above equation, where Γ is the complete gamma function.

1.3.2 Surface brightness profiles of spheroidal galaxies

The left panel of Figure 1.4 shows the Sérsic profiles of a couple hundred nearby ($z < 0.1$) spheroidal galaxies with stellar masses in the range $\log_{10}(M_*/M_\odot) \in [10.7, 11.6]$ from Szomoru et al. (2013). The profiles are colored by the stellar mass of the galaxies with darker red colors representing more massive, i.e. more luminous, galaxies.

The solid orange line shows the de Vaucouleurs profile. It is definitely a good description for some of the galaxies, but it is clear that profiles of spheroidal galaxies have a range of n values.

Also, the dashed orange line shows the profile Hubble (1930) proposed for elliptical galaxies. As you can see, it is not a bad description of the modern measurements, although individual galaxies do exhibit a diversity of profile shapes. Note that some galaxies have rather cored profiles, as implied by the Hubble's form, while others have cuspy profile (surface density continuing to increase towards smaller radii). There are arguments that there is a real dichotomy in the profiles of spheroidal galaxies (e.g., Faber et al., 1997). Brightest galaxies tend to have cored profiles. Not surprisingly, Hubble made his measurements for the brightest spheroidals and this is probably why he came up with the profile that is quite flat in the center.

1.3.3 Surface brightness profiles of late type disk galaxies

The right panel of Figure 1.4 shows the stellar surface density profiles of late-type disk galaxies in the combined THINGS and LITTLE THINGS sample which spans a wide range of stellar masses $\log_{10}(M_*/M_\odot) \in [7, 11]$. The green curve shows an exponential profile with normalization corresponding to the parameters of the Milky Way. To take away wide variation of galaxy sizes, the radii are normalized by the half-mass radius of each galaxy, while surface densities are normalized by $\Sigma_{1/2} = M_*/(2\pi R_{1/2}^2)$. The profiles are colored by the logarithm of stellar mass with lighter blue corresponding to lower stellar masses. The figure shows that $\Sigma_*/\Sigma_{1/2}$ is larger for dwarf galaxies in these rescaled units, which reflects their systematically lower surface densities $\Sigma_{1/2}$.

The figure shows that the exponential profile is a reasonable description of the surface density profiles. Nevertheless, many profiles are somewhat shallower than exponential in the outer regions and steeper than exponential in the center. The latter is often due to the presence of central mass concentration in the form of a bar or a bulge. Such concentration are not well fit by the exponential profile, which motivates fits of profiles with varying Sérsic index n or multi-component fits to $\Sigma(R)$.

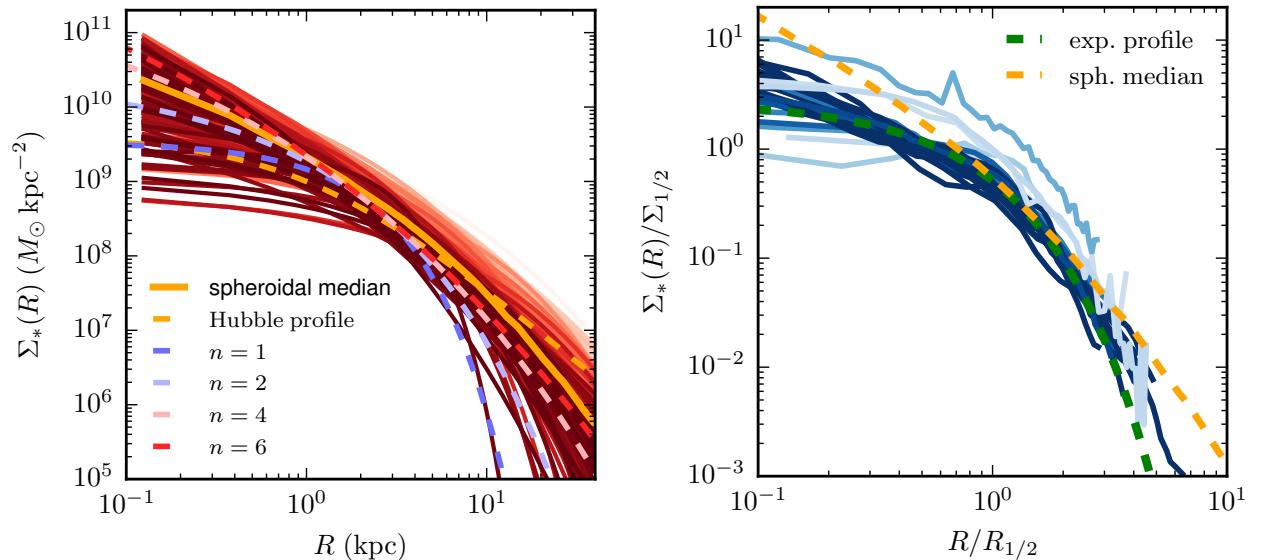


Figure 1.4: Left panel: the Sérsic profiles of a couple hundred nearby ($z < 0.1$) spheroidal galaxies with stellar masses in the range $\log_{10}(M_*/M_\odot) \in [10.7, 11.6]$ from Szomoru et al. (2013). The profiles are colored by the stellar mass of the galaxies with darker red colors representing more massive, i.e. more luminous, galaxies. Right panel: surface density profiles of the THINGS galaxies spanning a wide range of stellar masses $\log_{10}(M_*/M_\odot) \in [7, 11]$. The green curve shows the exponential profile with normalization corresponding to the parameters of the Milky Way. To take away wide variation of galaxy sizes, the radii are normalized by the half-mass radius of each galaxy, while surface densities are normalized by $\Sigma_{1/2} = M_*/(2\pi R_{1/2}^2)$. The profiles are colored with logarithm of stellar mass with lighter blue corresponding to lower stellar masses. The figure shows that $\Sigma_*/\Sigma_{1/2}$ is larger for dwarf galaxies in these rescaled units, which reflects their systematically lower surface densities $\Sigma_{1/2}$. The dashed orange line is the normalized median profile of spheroidal galaxies.

Figure 1.4 also show the median surface density profile of spheroidal galaxies. The radii in this case are normalized to the median half-mass radius, while vertical normalization is $\xi \bar{M}_*/(2\pi \bar{R}_{1/2}^2)$, where \bar{M}_* and $\bar{R}_{1/2}$ are the median stellar and half-mass radii and ξ is a factor adjusted so that the normalized profile roughly matches that of the exponential profile at $R \approx R_{1/2}$. Figure shows that the median profile of spheroidal galaxies is actually quite close to the profiles of late type galaxies at $R \gtrsim R_{1/2}$, and the main difference is at smaller radii: the central surface densities of spheroidal galaxies are larger than those of spiral galaxies (see Kravtsov 2013 for additional discussion on this). One implication of this fact is that one cannot turn a spiral galaxy into spheroidal by simply stopping star formation and letting stellar distribution fade and become red. Redistribution of stellar mass must occur in the process (Fang et al. 2013).

1.3.4 Multi-component descriptions of the surface brightness profiles

Galaxies with both a spheroidal component (e.g., bulges and bars) and a disk component are commonly modelled as a sum of the exponential disk and de Vaucouleurs bulge. In the main SDSS sample we've been examining, the `fracdeV` parameter is an indicator what fraction of light is well fit by the de Vaucouleurs profile. This is a rough indication of the fraction of light in spheroidal component. Nowadays, surface density profiles of late type disk galaxies are often fit with two Sérsic components with independent n values.

Purely spheroidal galaxies were thought to be much simpler and well-described by a single Sérsic component. However, a number of recent studies have shown that if one analyzes surface density profiles of ellipticals in detail, they actually require 2 or 3 Sérsic components to fit accurately. This was first found for the Brightest Cluster Galaxies (BCGs), in which the inner light distribution is described by the usual high- n de Vaucouleurs like profile, but the outer component is often described by a much flatter profile with $n \sim 1$

(Seigar, Graham, and Jerjen 2007, Kravtsov, Vikhlinin, and Meshcheryakov 2014). Huang et al. (2013) showed that two or three Sérsic components are required to accurately describe surface brightness profiles of all elliptical galaxies, with the outer components tending to have low values of n . In the single Sérsic component fits to such galaxies n is driven to large values (Graham et al., 1996).

The reason multiple components have not been detected earlier is because they are relatively subtle and require accurate measurements of profile at low surface brightnesses (which is difficult due to sky background). The existence of these distinct components indicates different formation mechanisms likely corresponding to different stages of evolution of early-type galaxies (inner regions are built up during early stages, while the outer profiles are built during late stages by accretion of other galaxies rather than by *in situ* star formation).

The outer regions of disks exhibit variation from a sharp cutoff to a more steady decline (e.g., Pohlen and Trujillo 2006 and §3.8 of Blanton and Moustakas 2009).

1.4 Galaxy magnitudes and luminosities

Now that we are familiar with the surface brightness profiles of galaxies, we can discuss various methods for estimating galaxy magnitudes (and thus luminosities, if the distance to the galaxy is known).

One thing that is clear from Figure 1.4 and the discussion in §1.3.4 is that the light distribution in galaxies extends to large radii. In fact, our ability to trace it is typically limited by the sensitivity of observations. This would not be an issue if the fraction of light at large radii was small. This is indeed the case for steeply falling profiles, such as the exponential profile. However, Figure 1.4 shows that the profiles of many spheroidal galaxies have outer slopes comparable to that predicted by the Hubble profile: $\Sigma \propto R^{-2}$, for which the total luminosity diverges logarithmically: $L(< R) \propto \int_0^R \Sigma(R) R dR \propto \ln R$. Thus, in galaxies with such profiles a large fraction of the total light may be contributed by the low surface brightness outer regions that may actually be below the sensitivity of observations, defined by the observation-specific level of background.

In practice, this is addressed in two ways. First, one can simply define the total magnitude as an *aperture magnitude* within a certain well-defined radius, ignoring light beyond it. Second, one could try to construct the best possible model for the distribution of light using the portion of the galaxy that is reliably measured in observations and compute the total light by extrapolating the model profile. In the SDSS magnitudes obtained in this way are called *model magnitudes*.

Once the apparent magnitude for a galaxy is estimated using one of these approaches, the absolute magnitude and luminosity can be computed if the galaxy distance or redshift are known using the equations A.18 and A.22 in §A.4.

1.4.1 Aperture, Holmberg, and Petrosian magnitudes

One of magnitudes provided in the SDSS database, called m_{fiber} , is based on light integrated within a fixed aperture of radius $1.5''$. However, it is a bad idea to use such fixed aperture magnitudes for physical studies of galaxies, because this would correspond to different physical scales for galaxies of different intrinsic size and galaxies of the same type but at different distances.

An example of a radius definition that does not suffer from these problems is the *Holmberg radius*, defined as the major axis of the observed galaxy's ellipsoidal light distribution, which itself is defined as the contour at which surface brightness of a galaxy in the B band is equal to 26.5 magnitudes per square arcsecond (Holmberg, 1958). Similar definitions could of course be used for other bands. This definition works as long as observations allow for measures of surface brightness to $26.5 \text{ mag arcsec}^{-2}$ reliably. What if this is not the case, like when a galaxy is far away or when observations are shallow?

Petrosian (1976) proposed to measure radii based on the integrated galaxy flux to make such estimates more robust. The idea is that as one traces the surface brightness profile $\Sigma(R)$, one can compare the light within some shell of thickness $\Delta R/R$, ΔI , to the total light within R , $I(< R)$ and define the galaxy radius as the radius corresponding to $\Delta I = \eta I(< R)$. One could then tune η in such a way that the measurement

recovers all or most of the flux for a particular assumed form of the profile $\Sigma(R)$ extending to infinity. The specific implementation of the Petrosian magnitude and sizes in the SDSS pipeline is described in §D.1.1.

One can show that theoretically the Petrosian magnitudes defined as in the SDSS should recover almost all of the flux of an exponential galaxy profile and about 80% of the flux for a de Vaucouleurs profile. As we discussed above, however, galaxy surface brightness profiles are often described by different functions (corresponding to different components) at different radii. Thus, in practice the Petrosian magnitude may underestimate total galaxy light by $\approx 0.2 - 0.5$ mag. The magnitude of the underestimate depends on the fraction of light in the outer component, which increases with increasing galaxy luminosity.

Thus, this underestimate will affect the brightest galaxies the most (see Bernardi et al., 2013). This issue will come up in the practical calculation of the galaxy luminosity function we will discuss below.

1.4.2 Model magnitudes

An alternative way is to estimate total magnitude by using an extrapolation of the model for surface brightness profiles obtained via a fit to the regions of the galaxy where the light profile is measured reliably. One could, for example, perform a fit of one of the surface brightness profile models discussed above or their combination (e.g., see Meert, Vikram, and Bernardi, 2015, for a recent example of such fits for SDSS galaxies). The total model magnitude is then obtained by integrating the best fit model profile to infinity.

The SDSS pipeline measures two kinds of model magnitudes based on fits of the exponential and de Vaucouleurs profiles to each galaxy in the r -band (the most sensitive band). The goodness of fit is evaluated for each of these models and the profile providing a better fit is chosen as the surface brightness profile model for computing SDSS model magnitudes. The model magnitudes are then computed by integrating this model to infinity.

On the other hand, *cmodel magnitudes* use surface brightness profile model constructed as a linear combination of the exponential and de Vaucouleurs models used to define model magnitudes in the r band: $F_{\text{composite}} = \text{fracDeV} \cdot F_{\text{dev}} + (1 - \text{fracDeV})F_{\text{exp}}$, where F_{dev} and F_{exp} are the de Vaucouleurs and exponential surface brightness profiles computed using the corresponding independent model fits in the r band. The cmodel magnitude is computed using *fracDeV* value that fits the surface brightness profile of a particular galaxy best. These magnitudes are recommended as the best choice for estimating galaxy luminosities and other physical properties. As we will see below, however, for bright galaxies they underestimate the total light.

We will also use magnitudes obtained for SDSS galaxies using the Sérsic profiles or multi-component fits by Meert, Vikram, and Bernardi (2015). In general, such magnitudes will give a result different from that given by the Petrosian magnitude definition or by the model and cmodel magnitude algorithms described above. For example, Graham et al. (2005) provide fitting formulae for correcting the Petrosian magnitudes to account for the total light in the Sérsic profile of a given n .

1.5 Luminosity distribution of galaxies

Figure 1.5 shows the distribution of galaxies in the SDSS main galaxy sample in the plane of redshift and absolute r -band Petrosian magnitude. Galaxies at a given z are approximately at the same distance from us⁹ and thus the distribution of galaxies in a vertical band of a given dz shows us the distribution of galaxies as a function of their intrinsic luminosity in that redshift interval — or the *galaxy luminosity function*. We can see that galaxies have a wide range of intrinsic luminosities.

This may seem trivial to you, but it took more than two decades for people in 1920s-1940s to realize this basic fact. When Hubble was considering galaxies in the samples he used for his work on expansion of the universe, he found that fainter galaxies had to have larger distances and he has postulated that galaxies have a narrow range of intrinsic luminosities with a Gaussian distribution. If you take a look at his seminal paper on the cosmic expansion, you will see that Hubble actually used apparent magnitudes to estimate distances to some of the galaxies – i.e., he treated galaxies as standard candles.

⁹Can you explain why this is only approximately true?

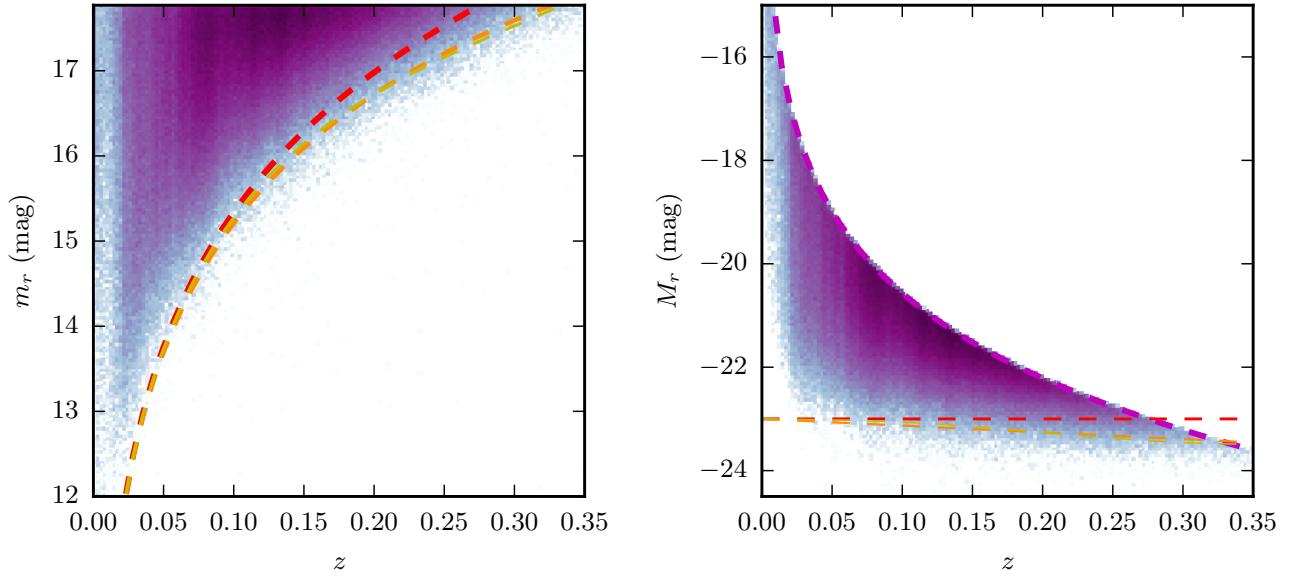


Figure 1.5: Left panel: apparent Petrosian r -band magnitude versus redshift z . The red dashed line shows expected $m(z)$ for an object of constant r -band absolute magnitude of $M_r = -23$. The orange and yellow lines correspond to an object of $M_r = -23$ at $z = 0$, but which is assumed to get brighter with redshift as $1.3z$ (orange) and which is k -corrected (yellow). Right panel: distribution of galaxies from the SDSS main sample in the redshift – r -band Petrosian absolute magnitude plane. The magenta dashed line shows the absolute luminosity limit corresponding to the apparent magnitude limit of $m_{\text{lim},r} = 17.7$ for the main SDSS spectroscopic galaxy sample. The other lines are the same as in the left panel.

It was only around 1943 that Fritz Zwicky realized that the universe is full of intrinsically faint galaxies from his studies of the Local Group neighborhood and distant galaxy clusters (where galaxies can be assumed to be at roughly the same distance). Zwicky then argued that galaxies had a broad range of intrinsic luminosities and that their luminosity distribution at low L in fact was steeply increasing with decreasing luminosity. The number of galaxies does indeed increase with decreasing luminosity, as can be seen in the distribution of M_r at a given z in Figure 1.5.

The upper envelope shown in Figure 1.5 by the dashed red line simply corresponds to the apparent magnitude limit of the main SDSS spectroscopic galaxy sample of $m_{\text{lim},r} = 17.7$: $M_r = m_{r,\text{lim}} + 5 \log_{10}(d_L/10\text{pc})$. The actual completeness is 0.93 near the limit and is then slowly decreasing with decreasing apparent magnitude (see e.g., Fig. 2 in Montero-Dorta and Prada 2009).

The figure shows that fainter galaxies (larger M_r) dominate by number and that there is a rather sharp cutoff in the number of galaxies with luminosities larger than a given threshold of $M_r \approx -22$. This sharp threshold changes somewhat to brighter values with increasing z . This increase is due to two factors: 1) the volume covered by the SDSS is increasing with increasing z , so brightest rarest galaxies are becoming more common and 2) luminosities of galaxies actually evolve across this redshift interval so that galaxies at larger z are intrinsically brighter. Thus, we can deduce the overall shape of the distribution of galaxy luminosity and rough sense of evolution right away from this plot.

The figure also shows prominent vertical stripes in distribution of galaxies in z direction. These stripes reflect large-scale structure in the galaxy distribution, which we will discuss in more detail below. In particular, the sharp feature at $z \approx 0.02$ is due to the supercluster associated with the Coma cluster of galaxies. The prominent wide stripe at $z \approx 0.07$ is due to the largest structure in the distribution of galaxies that was identified so far — the SDSS Great Wall (Gott et al., 2005). This structure is within the bottom of the region shown in the right panel of Figure 1.6 along with other prominent structures. The figure shows the cosmic web like distribution of galaxies in both redshift intervals shown in two panels. However, it is clear that the

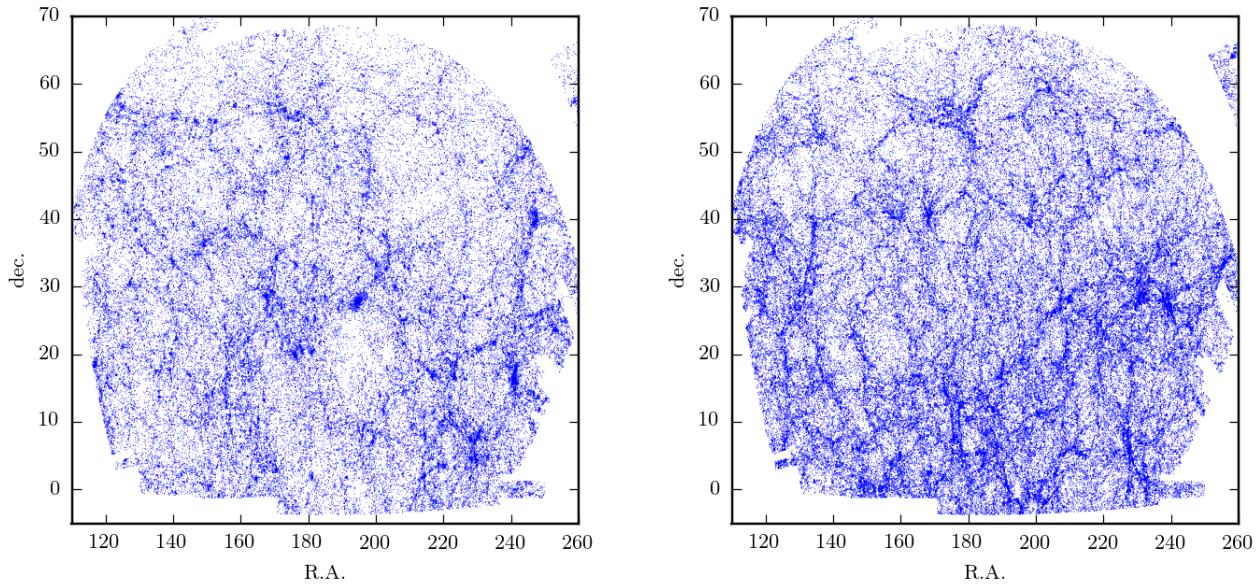


Figure 1.6: Left panel: distribution of the SDSS galaxies with $z < 0.06$ in the main northern region of the survey. Right panel: the same but for galaxies in the redshift interval $0.06 < z < 0.1$. The bottom of the region in this z interval contains many more prominent structures, including the SDSS Great Wall (Gott et al., 2005).

redshift interval $0.06 < z < 0.10$ contains much more prominent large-scale structures at the bottom of the region.

Large-scale structure as traced by galaxies of different properties is interesting in its own right and we will return to it later. For the purposes of studying the number density of galaxies, however, it causes variations. For example, it is clear from Figures 1.5 and 1.6 that if we estimated the number density of galaxies as a function of luminosity in the redshift range of $0.03 < z < 0.05$, we would get a considerably lower value than if we did this in $0.07 < z < 0.09$. Such variations are called *sample variance* (e.g., Hu and Kravtsov 2003) and are one of the sources of uncertainty in the estimation of galaxy abundance as a function of intrinsic luminosity — the luminosity function. So much so that the SDSS Great Wall region is often excluded when various galaxy statistics are computed. The other source is the usual Poisson noise due to there being a finite number of galaxies.¹⁰.

1.5.1 Malmquist bias

This apparent magnitude limit of the SDSS main galaxy sample shown by the magenta line in Figure 1.5 imposes a severe limitation in our ability to study properties of the faintest galaxies, as it limits the faint sample to only the nearest redshifts (i.e., a small volume). However, we can see from Figure 1.5 that at $z \lesssim 0.05$ we don't have a sufficiently large number of bright galaxies to probe the bright end of the distribution reliably. We can get a good number of bright galaxies if we include galaxies with redshifts $z > 0.05$, but then we are incomplete at faint magnitudes. This difficulty is a manifestation of the *Malmquist bias* (Malmquist, 1922) — the effect that only the brightest objects are visible at large distances for a flux (or magnitude) limited sample. If one is not careful, bright galaxies will dominate the volume of the sample.

This bias makes it difficult to estimate the luminosity function from a volume-limited sample of galaxies – i.e., a sample of galaxies in a given volume complete down to some limiting absolute magnitude. In practice,

¹⁰For galaxies of low luminosity, additional uncertainty arises due to the incompleteness SDSS in low surface brightness systems. Another source of uncertainty for faint galaxies, for which samples are complete only out to relatively small distances, are distance errors due to peculiar velocities that contribute significant fraction of observed redshift for nearby galaxies (Baldry et al., 2012)

to construct the luminosity function we need a way to combine information about faint galaxies from nearby distances, with information about the brightest galaxies at the largest distances.¹¹ We will discuss the ways to do this and the luminosity function of galaxies next.

1.5.2 Galaxy luminosity function

As mentioned above, the basic form of the luminosity distribution of galaxies was established in the 1940s and 1950s, which was followed by continuous refinement in measurements, but it was also followed by studies of how the LF varies with galaxy properties and environment. The LF was approximated by a variety of functional forms, including a double power law. In 1976, Paul Schechter proposed a power law + exponential form to the LF (Schechter, 1976) based on observations of galaxies in 13 clusters:

$$\phi(M) \equiv \frac{dn(M)}{dM} = 0.4 \ln(10) \phi_* 10^{-0.4(M-M_*)(\alpha+1)} \exp\left[-10^{0.4(M-M_*)}\right], \quad (1.9)$$

or in terms of luminosity:

$$\phi(L) \equiv \frac{dn(L)}{dL} = \frac{\phi_*}{L_*} \left(\frac{L}{L_*}\right)^\alpha \exp(-L/L_*), \quad (1.10)$$

where M_* or L_* is the characteristic absolute magnitude/luminosity that roughly corresponds to the sharp cutoff at bright luminosities we saw in Figure 1.5, and α is the power law slope of the function at faint luminosities. The Schechter form has become the standard way to describe the luminosity function of galaxies. In cases when the form is too simple to describe measurements well, a sum of two Schechter functions is fit.

If not for the Malmquist bias, measuring the luminosity function would be equivalent to just histogramming luminosities. The methods that are used in practice include a method for accounting for this bias, and also often methods to mitigate sample variance, distance errors, etc. The simplest method accounting for just the Malmquist bias is the $1/V_{\max}$ method.

The $1/V_{\max}$ method. Maarten Schmidt is famous for discovering an interpretation of spectra of quasars as high redshift objects. A few years after this discovery, when the sample of quasars became sufficiently large, he estimated the luminosity function of quasars (see the beginning of Section VII on p. 403 in (Schmidt, 1968)) using a simple method to correct for the Malmquist bias: instead of counting each quasar with an equal weight, thereby overweighing luminous quasars due to the larger volume to which they are detectable in a given survey, objects were counted by first calculating the maximum distance at which they could be seen given the survey limit, finding the volume of a sphere of this radius (the so-called V_{\max}) and then weighting the count by the inverse of this volume.

More precisely, continuing with the notion of the luminosity function as a histogram, in this method we construct a weighted histogram with weights $1/V_{\max}(L)$, where $V_{\max}(m_{\lim}, L)$ is the maximum *comoving* volume to which a given object of luminosity L will be visible given the apparent magnitude limit m_{\lim} of the sample. It is clear that this method assumes that galaxies are distributed uniformly in space, which, as we saw, is not strictly true due to sample variance. Thus, the $1/V_{\max}$ method estimate of the LF is approximate and is often accompanied by additional corrections (e.g., Baldry et al. 2012, Bouwens et al. 2015) that improve upon the method $1/V_{\max}$ by splining or binning variations of galaxies with redshift and then dividing the LF by a correction factor based on these variations. Felten (1977) presents detailed analysis of the $1/V_{\max}$ estimator of the luminosity function in practice and some of the associated issues.

Note that the $1/V_{\max}$ can be used not just to estimate the galaxy luminosity function, but more generally as a way to represent galaxies of different luminosities with correct relative weights in a histogram constructed from a flux-limited sample. We will be using it for such purposes below.

Figure 1.7 shows the luminosity function of galaxies in the SDSS main spectroscopic sample for different ways of measuring apparent magnitudes of galaxies. The solid red and magenta lines are luminosity functions,

¹¹An additional complication is that if the redshift interval we include is too large, galaxy luminosities and number density may evolve across the redshift interval and this evolution should also be accounted for.

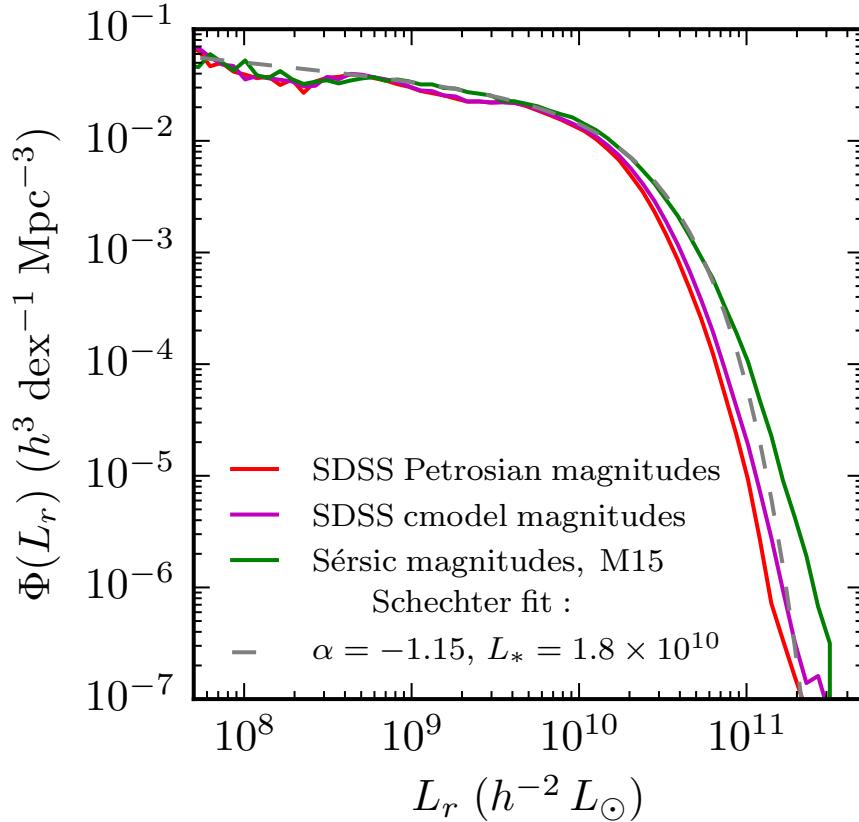


Figure 1.7: Galaxy luminosity functions for different magnitude definitions in the SDSS main galaxy sample. The solid red and magenta lines are luminosity functions, in which luminosities were computed using the Petrosian and cmodel magnitudes, respectively, without K -correction. The solid green line is the luminosity function, in which magnitudes were computed using improved photometry based on the Sérsic profile fits by Meert, Vikram, and Bernardi (2015) where galaxy luminosities were K -corrected. The gray dashed line shows the best fit Schechter function for the latter LF with the best fit faint- L slope and characteristic luminosity (in $h^{-2} L_\odot$) given in the legend.

in which luminosities were computed using the Petrosian and cmodel magnitudes estimated by the SDSS pipeline as described above (see also §D.1.1 and D.1.2), respectively, without K -correction. The solid green line is the luminosity function, in which luminosity functions were computed using improved photometry based on the Sérsic profile fits by Meert, Vikram, and Bernardi (2015) where galaxy luminosities were K -corrected. It is clear that the bright end of the luminosity function is quite sensitive to how magnitudes are estimated (see, Bernardi et al. 2013 for a more detailed discussion). One should be aware of this dependence when comparing theoretical models of galaxy evolution to observational estimates of the luminosity function or the stellar mass function derived from it.

The gray dashed line in Figure 1.7 shows the best fit Schechter function for the LF estimated from Sérsic fit photometry. Overall, it provides a reasonable fit for $L_r \lesssim 10^{11} h^{-2} L_\odot$. The best fit has a faint-end slope of $\alpha \approx -1.2$ and a characteristic luminosity (at which number density of galaxies starts to fall off exponentially) of $L_{*,r} \approx 2 \times 10^{10} h^{-2} L_\odot$, roughly consistent with other recent measurements. There is an ongoing debate about the correct slope of the luminosity function at low luminosities. This is because galaxies at these L are dominated by low-surface brightness systems, as the mean surface brightness decreases with decreasing luminosity (see below). A given galaxy survey may thus be increasingly incomplete with decreasing L thereby underestimating α . A recent measurement of the luminosity function using the GAMA survey – a survey that sampled a smaller volume survey, but going deeper to fainter luminosities and lower surface brightnesses

than the SDSS main galaxy sample – indicates a steeper faint end slope of $\alpha \approx -1.5$ (Baldry et al., 2012).

At large luminosities a Schechter fit underestimates the measured LF. For greater flexibility, the sum of two or three Schechter functions is often used to model the LF (e.g., Li and White 2009). Sometimes the Schechter function is modified to the following more flexible form (e.g., Eq. 9 in Bernardi et al. 2010):

$$\phi(L) \equiv \frac{dn(L)}{dL} = \frac{\phi_*}{L} \frac{\beta}{\Gamma(\alpha/\beta)} \left(\frac{L}{L_*} \right)^\alpha \exp(-[L/L_*]^\beta), \quad (1.11)$$

where $\Gamma(x)$ is the gamma function and β is additional parameter that provides more flexibility at high luminosities. Some studies use sums of such modified Schechter functions to model the luminosity function. Another model often encountered is modified form above but with a broken power law at faint luminosities (Blanton et al. 2005).

1.6 Galaxy spectra and colors

The color of an astronomical object is defined as the difference in apparent magnitudes between two spectral bands defined by two different filters. It is thus a quantity that depends on the specific choice of filters. The color also depends on how magnitudes are defined and are thus specific to a given magnitude system and method used to measure magnitudes. For example, Figure 1.8 shows several representative examples of SDSS galaxies of different colors ranging from blue ($g - r \approx 0.3 - 0.5$) to green ($g - r \approx 0.55 - 0.7$) and red ($g - r > 0.7$). The right panels show the corresponding spectra across the SDSS filter range. The legend in the upper right corner shows the $(g - r)$ color of each galaxy computed using the Petrosian magnitudes of the SDSS catalog in the corresponding bands. However, if we attempt to compute this color directly from the spectrum using equation 1.2 above, we find a color that is redder by $\approx 0.1 - 0.2$ magnitudes, especially for the late type galaxies. This is because the spectra are measured from the region of $1.5''$ radius around the center (the radius of the SDSS fiber), while the magnitudes are measured by integrating most of the light in each galaxy. The difference in color between the nuclei and outskirts of galaxies is visually apparent in the images of spiral galaxies in the left column of Figure 1.8.

Figure 1.8 also shows that the spectra clearly change qualitatively between blue and red galaxies. Blue and green galaxies have rather flat continua, prominent narrow emission lines, and only weak absorption lines. Red galaxies have spectra with suppressed emission in the u and g bands, do not have strong emission lines, and strong absorption lines and bands.

A rough characterization of these spectral shapes can be captured by colors defined using different bands. Figure 1.9 shows the distribution of galaxies of r -band absolute magnitude $M_r < -19$ in the color-color plane of $g - r$ and $u - g$ colors. The galaxies were selected within $200h^{-1}$ Mpc and represent a volume-limited sample. The figure clearly shows that the distribution of galaxies is not uniform, but is characterized by blue and red clumps with a “green valley” between them at $u - g \approx 1.5$ and $g - r \approx 0.7$.

Figure 1.10 shows the distribution of galaxy $g - r$ color as a function of galaxy r -band absolute magnitude, M_r , along with a grid of postage stamp images of randomly selected galaxies in $M_r - (g - r)$ bins. The blue clump and elongated “red sequence” of galaxies are clearly visible. Galaxies with luminosities $M_r > -20.5$ span a wide range of colors, although the faintest galaxies have mostly blue colors. The brightest galaxies are predominantly red and have mostly early morphological types.

1.6.1 k -correction

When we study the properties of galaxies from a wide range of redshifts in a given band, we are actually studying different, coarsely binned, parts of their spectra. For example, if the spectrum of a galaxy is steeply falling towards bluer wavelengths in roughly the same range as a given filter, a larger redshift will cause a galaxy of constant luminosity to appear dimmer just due to the filter response sliding down the spectrum. We are usually interested in the actual physical evolution of galaxy luminosities, not in the artificial effects of a limited filter range. To correct for this filter-induced evolution, one applies the so-called k -correction to the

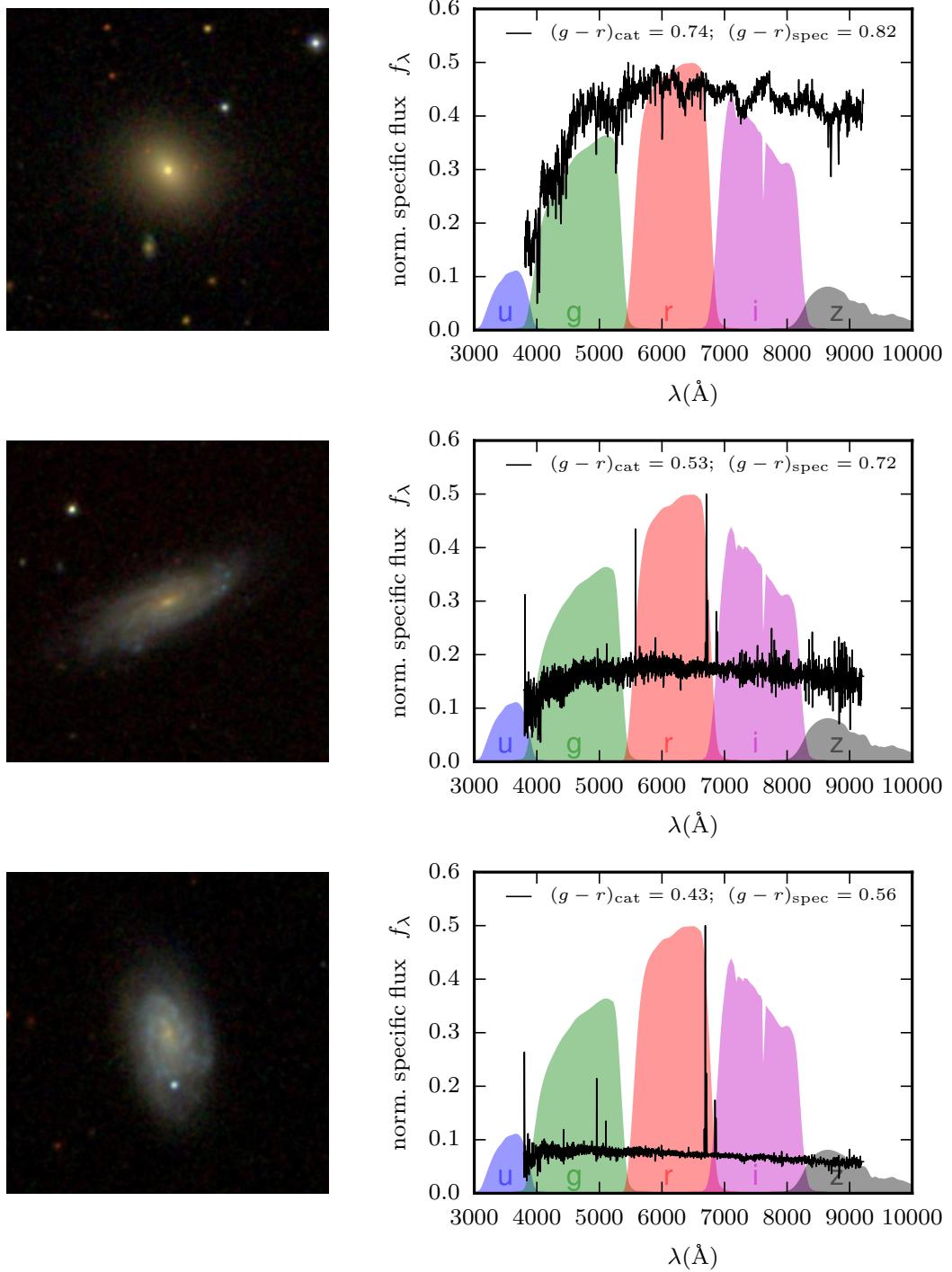


Figure 1.8: Examples of nearby SDSS galaxies of different $(g - r)$ color. Left column shows images of galaxies, while the right column shows spectra of these galaxies. The shaded regions are the sensitivity curves of the SDSS filters.

luminosity measured in a given band (see Oke and Sandage 1968 and Hogg et al. 2002) using equation A.19 in §A.4.

Given that galaxy spectra differ for different types of galaxies, k -correction is spectrum dependent. The correction is applied broadly to spectra with similar shapes, as quantified by broadband colors. The principle is thus simple, although in practice there are many subtleties (see, e.g., Blanton and Roweis 2007). For

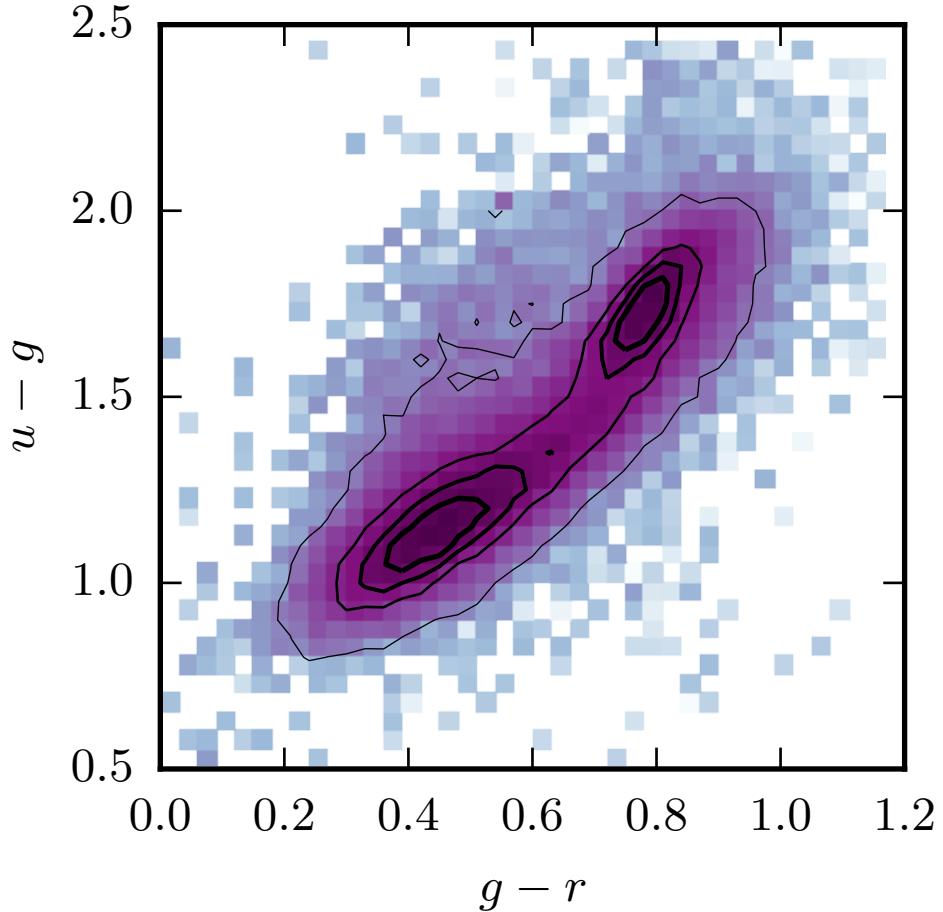


Figure 1.9: The distribution of galaxies with r -band absolute magnitude $M_r < -19$ in the color-color plane of $g - r$ and $u - g$ colors. The galaxies were selected within $200h^{-1}$ Mpc and represent a volume-limited sample. The distribution of galaxies is not uniform, but is characterized by blue and red clumps with a “green valley” between them.

example, we cannot just use the SDSS spectra to measure the k -correction for a given galaxy because spectra are taken through fibers of limited aperture and reflect the color of the region within that aperture, which can be significantly different from the spectrum of the entire galaxy. Also, the spectra of particular galaxies may be subject to observational limitations, such as limited wavelength range.

Thus, in practice sophisticated modelling of k -correction employing well-determined characteristic spectral templates is used. A widely used example is the the k -correction model of Blanton and Roweis (2007) implemented in the `kcorrect` code. At the same time, simple approximations exist for k -correction (Chilingarian, Melchior, and Zolotukhin 2010, Chilingarian and Zolotukhin 2012) with an online calculator and code snippets available [here](#). The effect of k -correction on luminosity of bright red galaxies with a typical red color ($g - r = 0.8$) can be gleaned from the difference between dark orange and yellow lines in Figure 1.5. The effect is quite small at $z < 0.3$ compared to the effects of photometry method or galaxy evolution over this redshift interval. At higher redshifts, however, effects of k -correction grow larger and it can be a source of significant systematic error in measurements of galaxy luminosities and luminosity functions (e.g., Lake and Wright 2016).

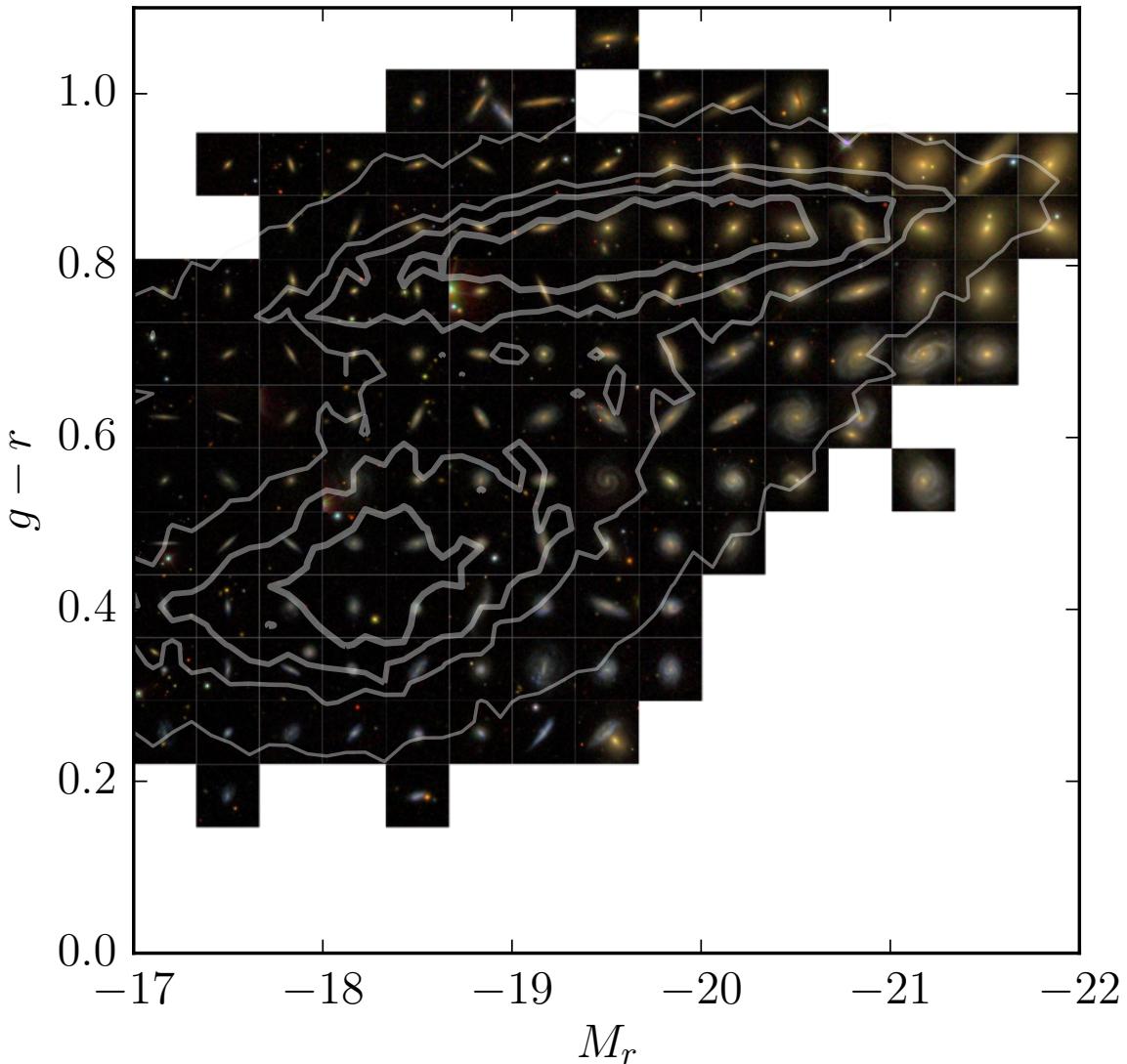


Figure 1.10: The contours show the distribution of SDSS galaxies within $150h^{-1}$ Mpc in the M_r and $g - r$ plane. In the background is a collage of images of galaxies randomly selected from within the corresponding magnitude and color bins. Each image shows a region of $25h^{-1}$ kpc. The four contours enclose 95%, 70%, 45% and 20% of all galaxies. The distribution of galaxies is not uniform, but is characterized by a blue clump, a red sequence, and a “green valley” between them. Note that within the selection volume the sample is not complete for galaxies with $M_r > -18.5$. Thus, the actual fraction of galaxies in the blue clump is larger than indicated by this plot.

1.6.2 From luminosity to stellar mass: stellar population synthesis

While there is some qualitative similarity between galaxy spectra and the spectra of stars of particular spectral classes, these spectra are also noticeably different, especially for bluer galaxies. This is because the spectra of galaxies are actually a combination of spectra of stars of different mass, metallicity, and evolutionary stage. In galaxies with gas and dust, gas emission contributes to the overall spectrum and dust affects the spectrum by preferentially scattering light at bluer wavelengths.

It is all but impossible to decompose a spectrum into properties of individual stars due to our limited

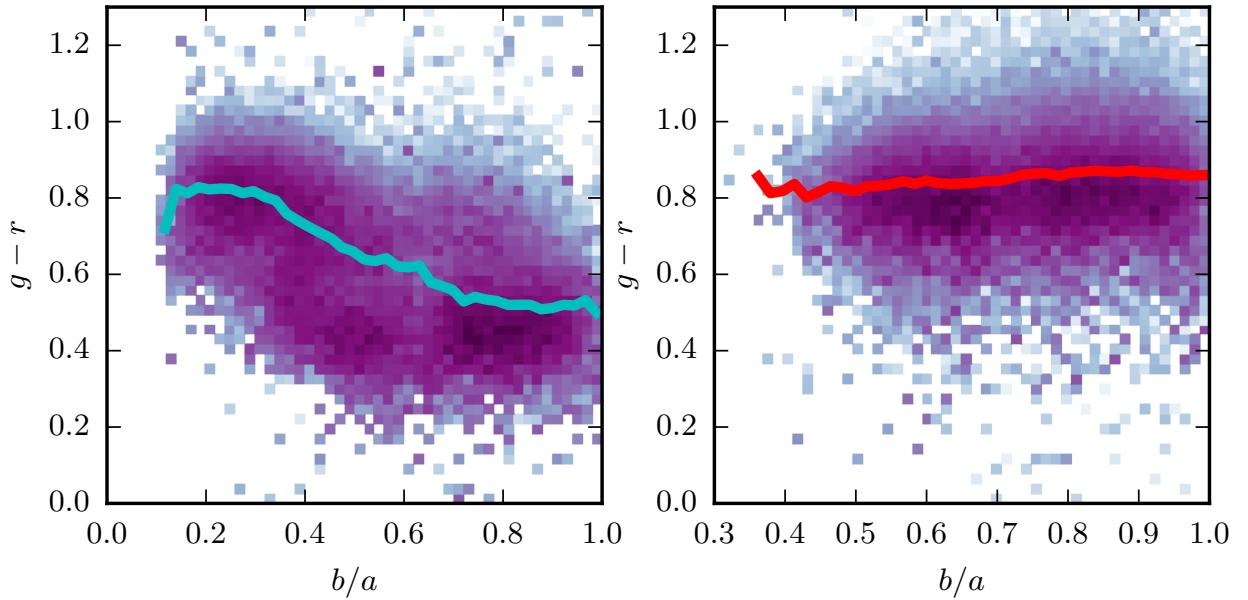


Figure 1.11: The $g - r$ color of late (left) and early (right) type galaxies as a function of the axis ratio of and ellipsoidal fit to the galaxy's light. Smaller axis ratios correspond to more flattened systems viewed closer to the edge-on view. The color maps show the distribution of galaxies of $M_r < -19$ in the Meert, Vikram, and Bernardi (2015) catalog. The distribution in the left panel includes only the galaxies with probability to have late type in the classification of Huertas-Company et al. (2011) of > 0.7 , while in the right plot distribution includes galaxies with the probability to be elliptical or S0 of > 0.7 . The lines in both panels show median $g - r$ at a given b/a . The plot shows that highly inclined, nearly edge-on disk galaxies have $g - r$ colors that are ≈ 0.3 redder on average than disks viewed face on. The effect on blue-band colors, such as $u - g$ or $u - z$ will be even larger because dust affects bluer wavelengths more strongly. The right panel shows that there is no trend of color with axis ratio for the early type galaxies. Note that the range of axis ratios shown in the right panel is smaller because distribution of b/a of these galaxies does not extend to < 0.4 . Note also the paucity of disks with $a/b < 0.1$ in the left panel.

knowledge of the star formation history, metallicity distribution, initial mass function, etc., of a given galaxy. So spectra are typically forward modelled by making simplifying assumptions about these properties in *the stellar population synthesis* (SPS) models (see Conroy, 2013, for a recent in depth review). The key assumptions that need to be made and which are the main source of uncertainty are the initial stellar mass function (IMF) of stars and star formation history (SFH) of a galaxy. There is currently no general, accepted theory that predicts IMF for given galaxy properties and evolution stage. SFH needs to be modelled within the context of a full galaxy formation theory, which is still a work in progress. While there is data on both the IMF and SFHs, the empirical information and guidance in modelling that it can provide are limited. Additional uncertainties are introduced by models of stellar evolution, dust, etc.

Nevertheless, SPS modelling is a useful tool and is commonly used in galaxy studies. Perhaps the most common use is to convert luminosities into stellar masses using galaxy spectra or colors using SPS estimates of the stellar mass to light ratio, M_*/L : $M_* = (M_*/L) \times L$. An estimate of M_*/L can be obtained with an SPS model given assumptions described above for a galaxy with a given spectrum, because such model can be used derive the best fit estimate of the stellar population properties, including stellar mass and metallicity that produces a galaxy with a given luminosity in different bands.

These models predict that mass-to-light ratios *on average* are particularly simple linear functions of galaxy colors (e.g., Bell et al. 2003, Zibetti, Charlot, and Rix 2009) – the fact that is used extensively for crude conversions of luminosities into stellar mass. For example, calculations of Bell et al. (2003)

predict that for a galaxy with a given $g - r$ color, the mass-to-light ratio in the r -band is: $\log_{10}(M_*/L_r) = -0.305 + 1.097 \times (g - r)$, where M_* is in M_\odot and L_r is in L_\odot . For similar calibrations for other bands and colors see their Appendix A.2 and Table 7.

Note, however, that in reality there is substantial scatter around these relations and their exact form depends on assumptions of the SPS modelling (see, e.g., Roediger and Courteau 2015 for a detailed discussion of uncertainties in the SPS-derived M_*/L ratios). Thus, the results of the SPS modelling should be treated as fairly crude approximations, not precise predictions (despite the large number of significant digits provided in the fit parameters). For comparisons of models to data, it is much more preferable to forward-model luminosities in a galaxy formation model, because star formation history and metallicity evolution are available in the context of such model. IMF still would remain a significant and important assumption, because modern galaxy formation models do not include a model for IMF.

1.7 Structural properties of galaxies and their relation to morphology

We have already considered surface brightness profiles and saw that they can be described by the Sérsic profile with a range of indices n . We also discussed that galaxies often exhibit indications of multiple components in the distribution of light. Here we will discuss simple quantities parametrizing basic properties of light distribution and their relation to morphology.

1.7.1 Galaxy surface brightness

Let us return to Figure 1.10 showing images of representative galaxies in the $M_r - (g - r)$ plane. Recall that each postage stamp image in the grid shows a region of the same physical size of $25h^{-1}$ kpc. Thus, the relative sizes of galaxies are represented correctly. One can see from the figure that the physical size of galaxies correlates with its luminosity. We will examine this correlation in more detail later, but for now note that for faint galaxies correlation is actually quite weak (if present at all).

As we discussed above, surface brightness is defined as the amount of light that we receive from a given area of the galaxy. This quantity makes sense for an extended object. It can be defined for example using the total magnitude or luminosity computed within a given filter using one of the methods discussed above and the radius enclosing half of this total light. So, if the total luminosity is L and half-light radius is $R_{1/2}$, we can define galaxy surface brightness as $\Sigma = L/(2\pi R_{1/2}^2)$. One can also use stellar population modelling of galaxy colors or spectra to derive underlying stellar mass under the model assumptions. Thus, sometimes we use total stellar mass M_* in lieu of L , in which case we talk about stellar *surface density*.

The weak correlation of size with luminosity means that for galaxies of increasing luminosity, size increases slowly and so *surface brightness* increases with increasing galaxy luminosity. Conversely, fainter galaxies tend to have lower surface brightness. This is illustrated in Figure ??, which shows distribution of the r -band surface brightness as a function of M_r for SDSS galaxies. The figure shows a strikingly large variation of surface brightnesses of galaxies of a given absolute magnitude.

Historically, it was believed that early type galaxies tend to have a narrow range of high values of the *central* surface brightnesses, which is sometimes called the "Kormendy law" (Kormendy, 1977), while the tendency of for late type galaxies to concentrate near the peak of the low-concentration clump is called the "Freeman law" (Freeman, 1970). Specifically, Freeman (1970) found that the central surface brightnesses of spiral galaxies in the B -band occupied a narrow range: $\mu_B \approx 21.6 \pm 0.3 \text{ mag arcsec}^{-2}$). As can be seen from Figure ??, large galaxy surveys showed that these "laws" simply reflect the tip of the iceberg of the surface brightness distributions of late and early type galaxies. The overall population exhibits a much wider range of surface brightnesses.

The actual distribution may be even wider than Figure ?? indicates, especially for low luminosity galaxies. This is because a population of *low-surface brightness* (LSB) galaxies is missing in them. Such galaxies form a tail of the distribution of late type galaxies towards surface brightnesses dimmer than the "blue island" of late type galaxies and are particularly abundant at faint magnitudes (Bothun, Impey, and McGaugh 1997, Driver 2004). You can see the distribution of surface brightnesses as a function of galaxy luminosity

(expressed as absolute magnitude) [here](#).

The LSB galaxies have disk or irregular morphologies. However, they are difficult to detect against the brightness of the night sky and are thus largely missing in galaxy surveys unless a special care is taken to include them as is done in some special deep surveys (e.g., Driver et al. 2005).

One striking manifestation of this difficulty is reflected by the fact that we are still discovering copious numbers of new satellite galaxies of the Milky Way (Koposov et al. 2015, Drlica-Wagner et al. 2015). Very diffuse, low-surface brightness galaxies are also being discovered in significant numbers at larger distances (van Dokkum et al. 2015).

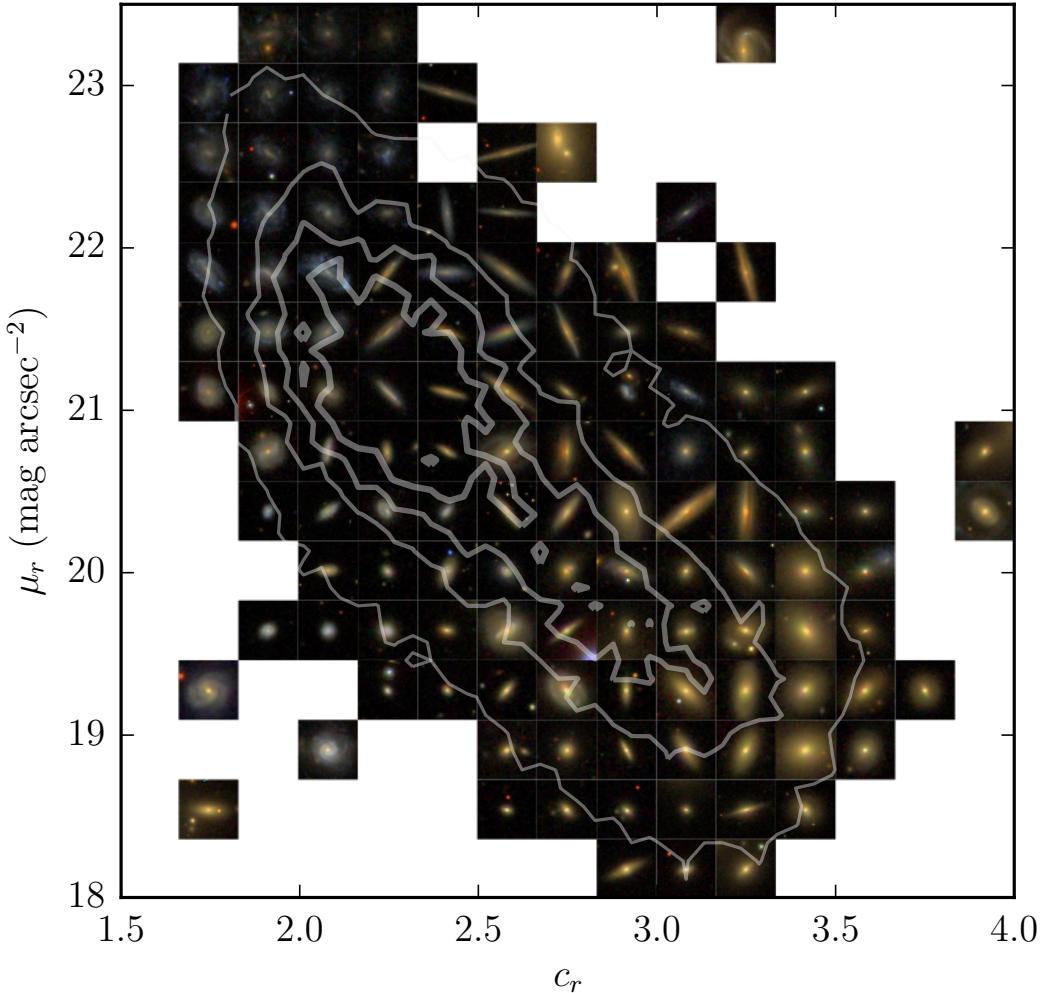


Figure 1.12: Contours show the distribution of SDSS galaxies in the plane of light concentration vs the r -band surface brightness. In the background a grid of postage stamp images of randomly selected galaxies in bins of c_r and μ_r are shown for illustration. Each image shows a region of $25h^{-1}$ kpc. The four contours enclose 95%, 70%, 45% and 20% of the galaxy distribution. The distribution of galaxies is not uniform, but is characterized by two clumps of low and high concentration and surface brightness. These clumps reflect bi-modality in the *structural* properties of galaxies and are thus qualitatively different from the bimodality and color in Figure 1.10.

1.7.2 Light concentration

Another simple quantity that can be defined without knowledge of galaxy distance, is the *concentration of its light distribution*. SDSS database provides estimates of the radii (in arcsec) enclosing 50% and 90% of the

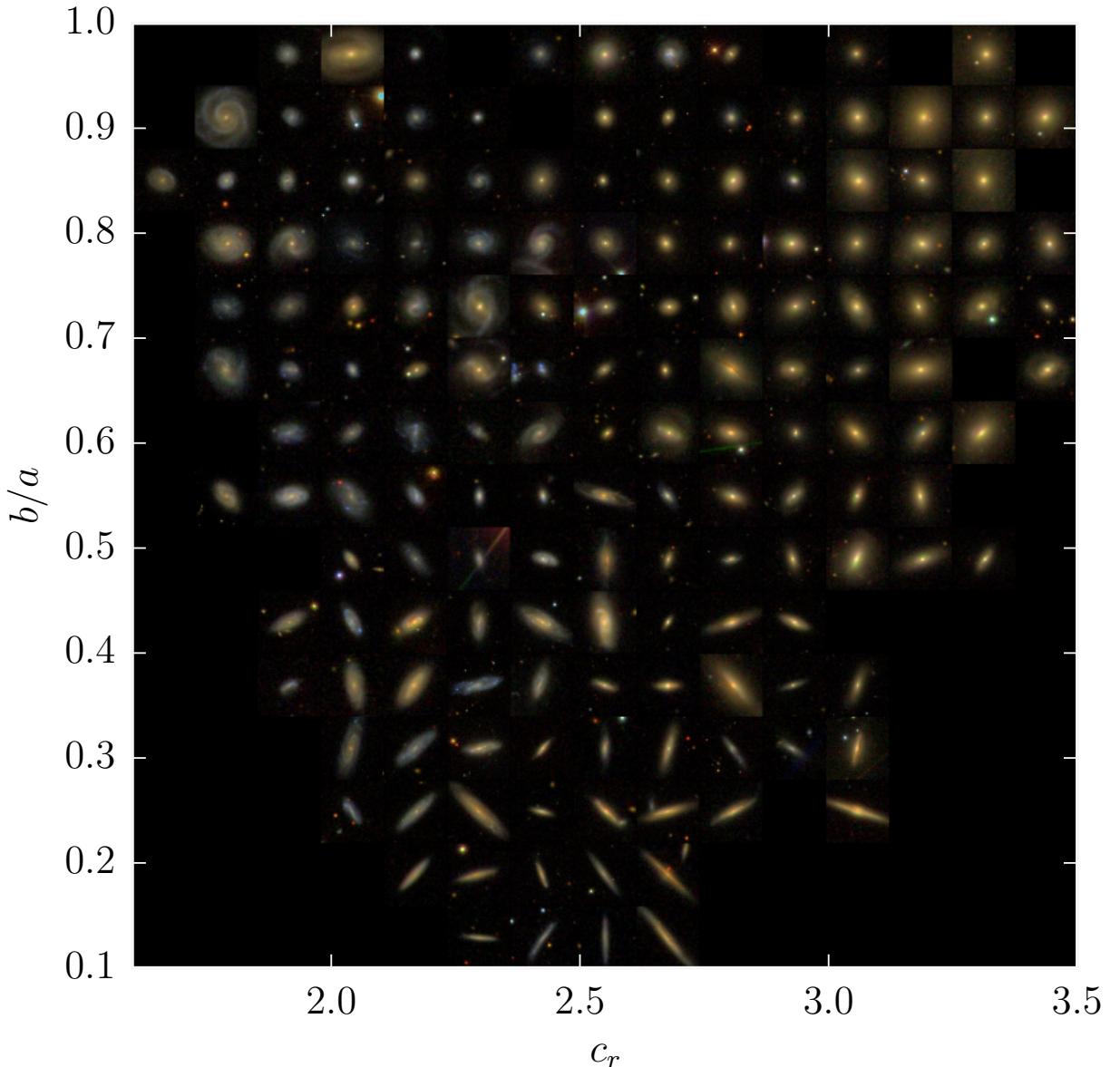


Figure 1.13: Collage of images of the SDSS galaxies randomly drawn from the sample of galaxies with $M_r < -19$ at distances $d_L < 200$ Mpc in the concentration–image axes ratio plane $c_r - b/a$. The figure shows that different morphological types separate well in this plane: late type close to face-on galaxies occupy upper left corner of the distribution, elliptical galaxies are in the upper right corner, edge-on disks are in the bottom–center, and lenticular galaxies occupy the middle–right part of the diagram.

total light estimated using the Petrosian (1976) method (see Appendix D.1.1). Their *ratio* provides us with a distance-independent measure of concentration of galaxy light.¹²

$$c \equiv R_{90}/R_{50}. \quad (1.12)$$

For example, if galaxy light is redistributed in such a way that more of it is concentrated towards the center, R_{90} should be largely unaffected, while R_{50} should decrease and thus c would increase. Strateva et al. (2001) showed that this concentration tightly correlates with the traditional morphological types of galaxies obtained via visual classification.

¹²For distant galaxies of $R_{50} \lesssim 3''$ concentration is affected by seeing to some extent - see Blanton et al. (2003). This is not a significant worry for our broad brush estimates, although if working with sufficiently large sample cuts on size to $> 3''$ are prudent.

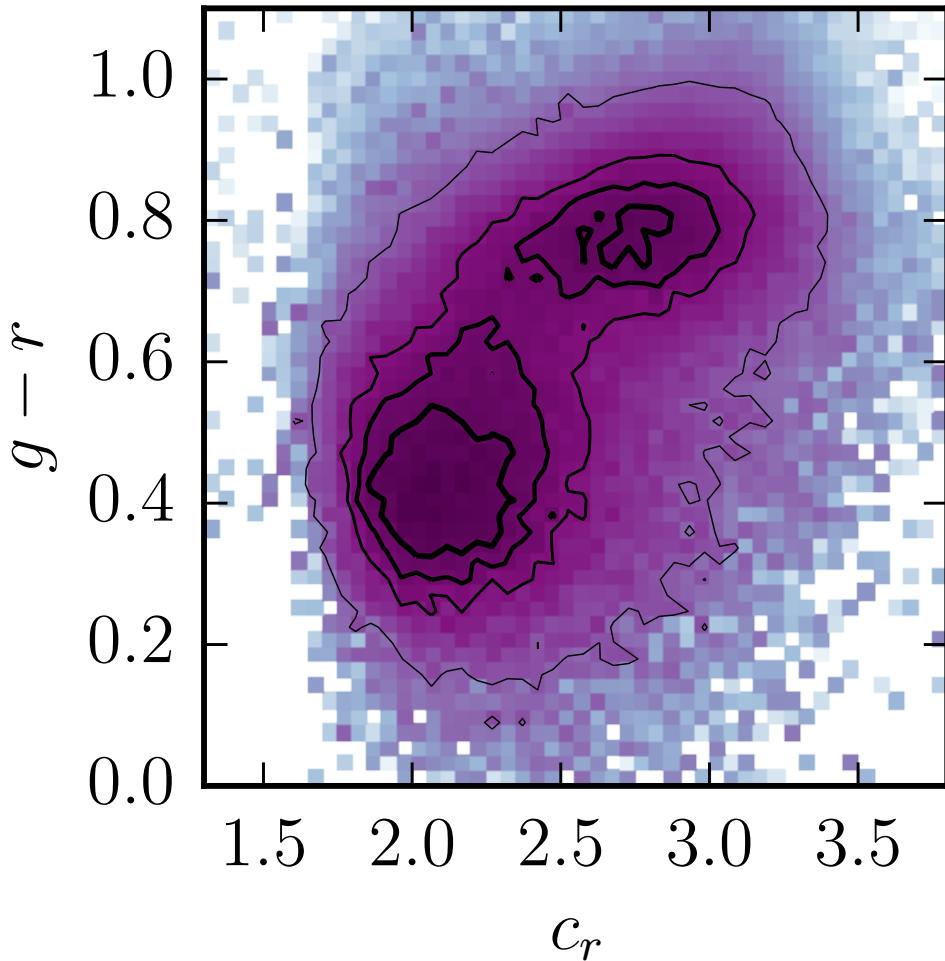


Figure 1.14: Distribution of galaxies in the r -band concentration, c_r , $g - r$ color plane. Galaxies are weighted by their $1/V_{\max}$ to remove the Malmquist bias. The four contours enclose 95%, 70%, 45% and 20% of the galaxy distribution. The distribution of galaxies is not uniform, but is characterized by two clumps. We have previously seen bimodality of color (the quantity controlled by age of stellar population, its metallicity, and affected by dust reddening), this plot indicates that there is also a weaker bimodality in the concentration of light distribution – a property describing structure of galaxies.

It may seem like concentration is similar to surface brightness, but they are actually distinct properties: concentration can be thought of as parametrizing shape of the luminosity profile, while surface brightness quantifying its normalization. In practice, however, the two quantities are correlated. Figure 1.12 shows the distribution of galaxies in the c_r - μ_r plane along with randomly selected postage stamp images of galaxies in bins. The distribution is broad but exhibits a clear correlation between the two quantities. We can see that early type galaxies mostly populate the high concentration, high surface brightness end of the distribution, while the late type spirals occupy the low-concentration, low surface brightness end.

If we examine galaxies in the horizontal rows of a given μ_r in Figure 1.12, we see fairly uniform morphologies at low surface brightness, but a wide range at intermediate and high μ_r values. The surface brightness is thus not a good indicator of morphology on its own.

If we consider vertical columns of galaxies at a given c_r , the morphological mix tends to be uniform, except at high concentrations, where elliptical and S0 galaxies are inter-mixed with disk systems. Note, however, that the disk systems at high c_r are all highly inclined. This is because the apparent concentration of an exponential disk depends on the inclination under which the disk is viewed. Nearly edge-on disks have

high concentrations and are thus located in the high- c tail. We can see that concentration is a potentially good indicator of galaxy morphology, if we could correct for the effects of inclination on disk galaxy concentrations. This can be done by dividing concentrations by the function that describes the change of exponential disk concentration with inclination, as proposed by Bailin and Harris (2008).

The good separation of galaxies in the axes ratio – concentration plane is illustrated in the collage of randomly drawn images of SDSS galaxies from the sample with $M_r < -19$ and distances of $d_L < 200$ Mpc in Figure 1.13. The figure shows that different morphological types separate well in this plane: late type close to face-on galaxies occupy upper left corner of the distribution, elliptical galaxies are in the upper right corner, edge-on disks are in the bottom–center, and lenticular galaxies occupy the middle–right part of the diagram.

This nice separation of types motivates the use of b/a and c_r with a machine learning classification method to classify galaxy morphologies. These methods work by first definition regions occupied by galaxies of different morphological classes given the training sample with known classes, and then classifying other galaxies of unknown class based on where they fall within the set of parameters on which training was performed. In this case, classification can be done using both b/a and c_r , as well as other parameters correlating with morphology, such as color.

Figure 1.14, for example, shows distribution of $g - r$ color vs concentration, which features two distinct clumps of galaxies. We observed color bimodality before. However, here a bimodality in concentration is also present. This bimodality is in the *structural* property of galaxies and is thus qualitatively different from the bimodality of color. For example, if a galaxy was building up its stellar mass at a steady rate, but color was changing nonlinearly with time due to peculiarities of shape evolution of stellar spectra with time, the bimodality of color could arise just from this nonlinearity. However, the fact that we see bimodality in the structural parameter in light distribution means that it reflects a real physical bifurcation of galaxy properties. In particular, this means that *galaxies become red when their light becomes concentrated due to some physical process*.

1.7.3 Vertical structure of disks

Perpendicular to the plane disks are generally well described by an exponential atmosphere with a roughly isothermal profile at a given R , although the scale height varies with R and disks often "flare" - increase scale height significantly in the outer regions. Figure 1.11 discussed above shows colors of SDSS disk galaxies as a function of their apparent ellipticity characterized by the axis ratio of light distribution. Smaller axis ratios correspond to more inclined disks. You can see that axis ratios are rarely less than 0.1. This reflects the finite thickness of the disks in the direction perpendicular to the disk plane. Indeed, a detailed study by Yoachim and Dalcanton (2006) shows that the ratio of the disk scale-length to the scale-height ranges in $\sim 8 - 12$ with a weak trend with galaxy mass and rotation velocity.

1.8 Galaxy clustering and large-scale structure.

Awareness that there is large-scale structure in galaxy distribution stemmed from identification of groups and clusters of nebulae in the General Catalogue of Nebulae and Clusters of Stars compiled first by William Herschel and then his son John Herschel from the late 18th through mid-19th century. For example, Proctor (1869) shows the maps of all sky distribution of nebulae in the Herschel's catalog and comments that it is highly non-uniform with nebulae in "streams" and "groups". The most prominent concentration of galaxies in Proctor's maps was what we now call the Local Supercluster. This structure was later re-discovered by R.H Reynolds in 1923.

In the early 1950s de Vaucouleurs (1953a) has studied the band in detail and argued that it represented a "super-galaxy". The galaxies in a band were thought to be akin to the Milky Way stars that form a band on the sky. He has later changed the name of this structure to the *Local Supercluster*. However, the original name survives in the name of celestial coordinate systems aligned with the galaxy band, called *supergalactic coordinates*. In fact, distribution of galaxies within 10 Mpc of the Milky Way does resemble a disk, as shown in Figure 1.15. Two 3D projections of galaxy distribution within 15 Mpc of the Milky Way. The

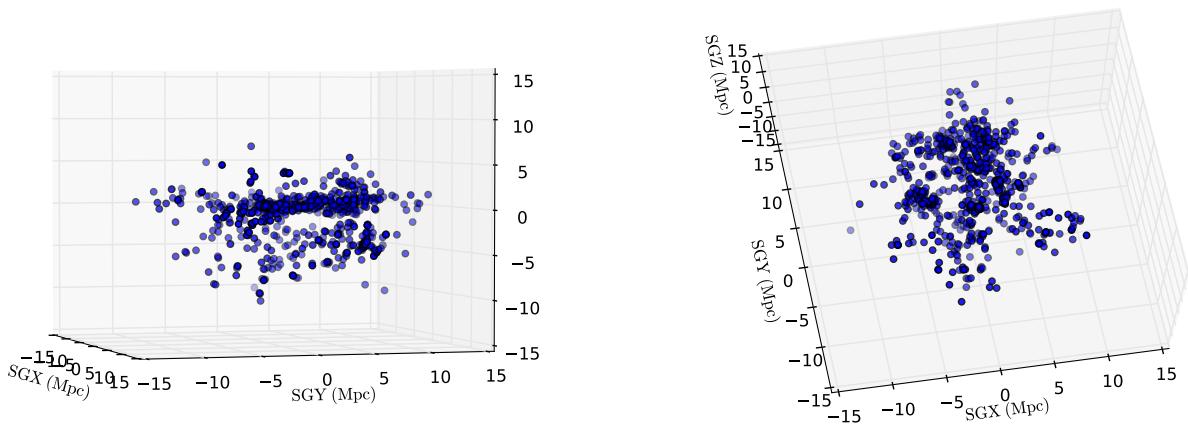


Figure 1.15: Two 3D projections of galaxy distribution within 15 Mpc of the Milky Way. The galaxies are from the catalog of the Local Group galaxies compiled by McConnachie (2012) and nearby galaxy catalog of Karachentsev, Makarov, and Kaisina (2013). The figure shows that distribution of nearby galaxies is flattened around the $SGZ = 0$ plane. This is by definition, because it is this flattened structure that corresponds to the “super-galaxy” of des Vaucouleurs. The system of supergalactic coordinates, in which the distribution is plotted is defined relative to this structure.

galaxies are from the catalog of the Local Group galaxies compiled by McConnachie (2012) and nearby galaxy catalog of Karachentsev, Makarov, and Kaisina (2013). The figure shows that distribution of nearby galaxies is flattened around the $SGZ = 0$ plane. This is by definition, because it is this flattened structure that corresponds to the “super-galaxy” of des Vaucouleurs. The system of supergalactic coordinates, in which the distribution is plotted is defined relative to this structure. The distribution of galaxies within 5 Mpc of the Milky Way is particularly flattened. The best fit ellipsoid to this distribution is oblate with axes ratios of $\approx 1 : 0.84 : 0.29$. The distribution of galaxies within the ellipsoid is not uniform, however, but is concentrated in several groups. The flattened structure – a *sheet* – is characteristic structure forming at the interface of two under-dense regions (voids). Indeed, the Local Group is surrounded by two prominent voids on both sides of the sheet (e.g., Karachentsev, Makarov, and Kaisina 2013). Interestingly, the distribution of satellites around Milky Way, M31, and other nearby massive galaxies is flattened, so one often hears discussion of “*planes*” of satellites. Orientations of these “*planes*” are aligned with the large-scale distribution of galaxies (Libeskind et al. 2015). This illustrates that properties of galaxies can correlate or depend on their large-scale environment.

Another interesting feature is that the flattened structure extends well beyond the 30 Mpc region shown here, so our local “sheet” is part of a larger coherent structure – the Virgo supercluster. What about even larger scales? Existence of even larger structures beyond the local supercluster seen by R. Proctor and others in mid-19th century was revealed in systematic catalog of galaxies by Harlow Shapley and collaborators undertaken in 1925–1939. The most prominent structure identified by Shapley – a concentration of galaxies in the Centaurus constellation had an elongated “oval” shape (Shapley 1930). This was one of the first examples of what we now call “superclusters”. The Shapley’s structure now carries his name – the Shapley supercluster (Raychaudhury 1989).

Studies of large-scale distribution of galaxies were difficult because they required painstaking compilation of imaging data over large areas of the sky and identifying and cataloging galaxies in the images. Such surveys have been constructed starting with Harvard Observatory one in 1920s–1930s, Zwicky’s survey in the 1950s–1960s, Lick survey of galaxies by Shane and Wirtanen, etc. These surveys revealed that structures like clusters and superclusters were common in galaxy distribution. At the same time, the surveys indicated

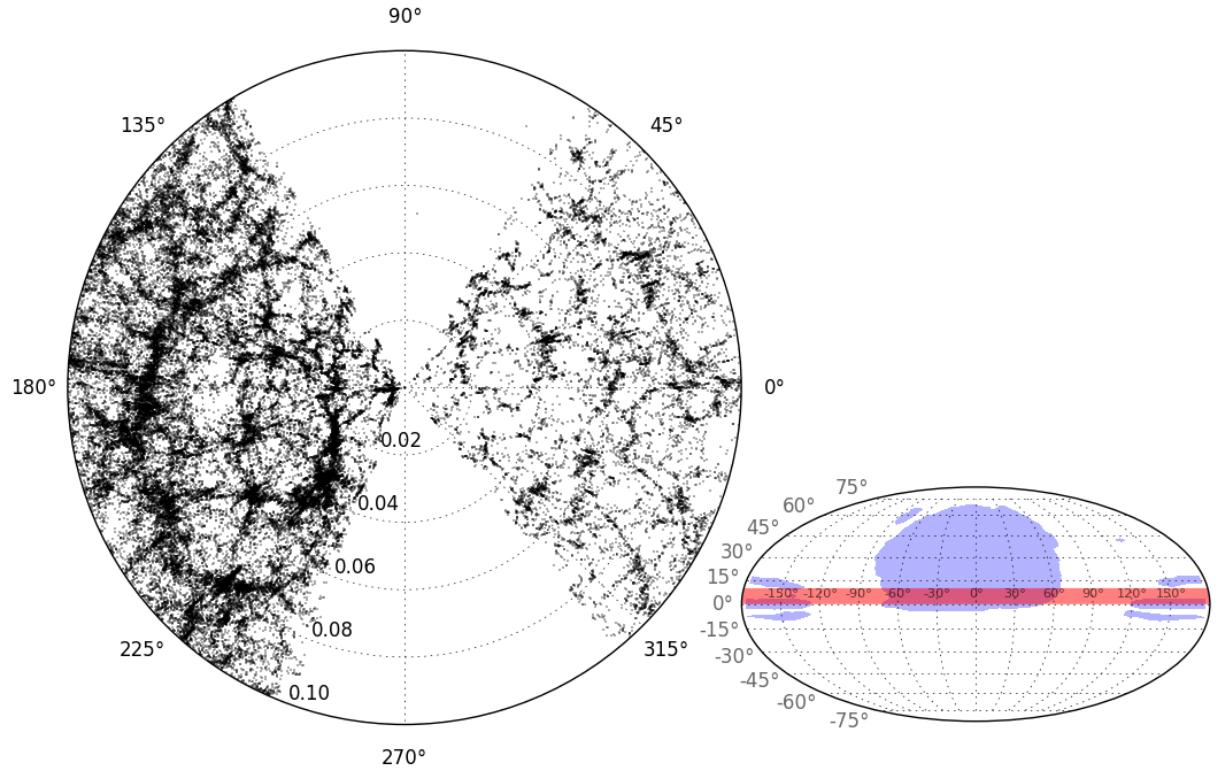


Figure 1.16: Large-scale structure in the distribution of SDSS galaxies at redshifts $0 < z < 0.1$ and declination slice of $0 < \delta < 10^\circ$; the declination range of the wedge is shown as a red band in the sub-plot showing Mollweide projection of the sky with the SDSS area shown in lavender color. Large ribbon-like structure at $z \approx 0.07 - 0.08$ in the left wedge of the diagram is the SDSS Great Wall. The figure was made using the [LSSGalPy](#) code.

presence of ever larger structures comparable to the survey extent. This prompted de Vaucouleurs (1970) to argue that distribution of galaxies is highly non-uniform on large-scales and is characterized by *hierarchical structure* at different scales: i.e. that galaxies exhibited an hierarchy of structures at different scales. Peebles (1974) and collaborators in a series of papers (e.g., Peebles and Hauser 1974, Groth and Peebles 1977) showed that the 2-point correlation function of galaxies, defined as excess probability to find a pair of galaxies at a given separation compared to uniform distribution, had a power law form and was positive on the largest scales in all the surveys available at the time. Positive correlations indicated that there was structure on all probed scales ($\lesssim 30 - 40$ Mpc), and the absence of negative correlations indicated that these scales were still far from the scale of uniformity in galaxy distribution. The power law form of the 2-point correlation function also indicated *self-similarity* of galaxy structures lending support to the hierarchical clustering picture.

These developments strongly motivated exploration of galaxy distribution on ever large scales, including the third dimension along the line of sight. Mapping out galaxy positions in that dimensions was becoming easier with advances in detector and spectrograph technology. The first modern-style modern-style redshift “pie-slice” diagram were presented by Gregory and Thompson (1978) and Joeveer and Einasto (1978). These showed the main elements of the cosmic web as we now know it: clusters, filaments connecting them, and surrounding voids. The “frothy” (or “cellular” according to Joeveer and Einasto 1978) structure of the large-scale distribution of galaxies was further quantified in the surveys during the 1980s. The image of galaxy distribution that has left indelible impression is elaboration of the distribution around the Coma

cluster in the CfA survey (de Lapparent, Geller, and Huchra 1986), which looked like a “stick-man” with outstretched arms forming an elongated structure that the authors dubbed “the Great Wall.” This structure still span the entire survey. The extension of the CfA survey into the southern sky, have found a counterpart “wall” that seemed to be an extension of the “Great Wall” (da Costa et al. 1994). So again, the largest structure was spanning the entire survey. This has continued with the studies of ever increasing volumes with the Las Campanas, 2dF, and SDSS redshift surveys. The culmination was the discovery of the SDSS Great Wall in the Sloan digital sky survey (Gott et al. 2005), a filamentary structure that spans ~ 400 Mpc ($\sim 80\%$ larger than the CfA Great Wall), which can be seen in Figure 1.16.

No prominent coherent structures on even larger scales were found, but existence of structures like the SDSS Great Wall is the reason why modern galaxy surveys still do not allow to reliably measure the scale were correlations turn negative, expected to be at $\approx 130 h^{-1}$ Mpc.

1.9 Evidence for dark matter in galaxies and galaxy clusters

"It is incidentally suggested that when the theory is perfected it may be possible to determine the amount of dark matter from its gravitational effect." - Jacobus Kapteyn (1922)

Dark matter, by definition, emits little or no electromagnetic radiation. The evidence for it thus is almost always derived by its gravitational effects on the visible matter. These effects are used to estimate total gravitating mass (e.g., see Courteau et al. 2014 and Einasto 2009 for review of the methods by which this is done).

The history of discussion of evidence for the presence of dark matter in galaxies is almost as old as the study of galaxies themselves, even if we do not count Herschel’s 18th century arguments for “dark” matter in the 18th century due to strange gaps in his map of the Galaxy from star counts in different directions, which turned out to be due to dust. This section briefly reviews the evidence for dark matter in galaxies and galaxy clusters obtained with different methods. The order of the methods discussed below roughly follows the historical order in which the methods were proposed and applied. The review ends with the most compelling and precise measurement of dark matter density in the universe from the temperature fluctuations of the cosmic microwave background (CMB) and baryonic acoustic oscillations feature in the galaxy correlation function.

The history of dark matter investigations starts with studies of the kinematics of stars within the Milky Way in 1920s.

1.9.1 Disk of the Milky Way.

Kapteyn (1922) used measurements of velocities of stars around the Sun to estimate the gravitational potential of the Milky Way disk. The idea is that in equilibrium there is a relation between spatial and velocity distributions of tracers moving in a gravitational potential and the gradient of the potential expressed in the hydrostatic equilibrium equation: $\nabla\phi(\mathbf{x}) = -\nabla p(\mathbf{x})/\rho(\mathbf{x})$, where $\rho(\mathbf{x})$ is spatial density of tracers at the location \mathbf{x} . In the case of stellar “fluid”, the pressure is $p(\mathbf{x}) = \rho \sigma^2$, where σ^2 is the velocity dispersion of tracers¹³ at \mathbf{x} .

Kapteyn used distribution of stars perpendicular to the disk and their velocity dispersion in that direction (usually denoted as z -direction, with x and y oriented in the plane of the disk) to constrain $\partial\phi/\partial z$ and compare it to the potential expected from the visible stars alone. He did not find any significant evidence for the extra dark mass beyond what could plausibly be attributed to stars. He was, however, aware of the possibility judging by the words in the abstract of his paper: *"It is incidentally suggested that when the theory is perfected it may be possible to determine the amount of dark matter from its gravitational effect."* One often sees the term “dark matter” attributed to Zwicky’s papers in the 1930s (see below). However, this quote from Kapteyn shows the use of the term 10-15 years earlier.

¹³Velocity dispersion is analogous to thermodynamic temperature of gas in the expression for gas pressure: $p = nkT$. Note that constant normalization of ρ does not matter, as it cancels out in the equation.

Kapteyn's analysis relied on the assumption that distribution of velocities in vertical direction is Maxwellian and does not vary with z , which make the analysis simple but are not quite correct. Perhaps motivated by the prospect of discovering "dark matter" Jeans (1922) corrected Kapteyn, by carrying out proper analysis using what is now known as the "Jeans equations" and assuming "Jeans equilibrium". The Jeans equations are essentially the hydrostatic equilibrium equations, but take various terms arising in a collisionless stellar fluid properly into account (see §B.1).

Jeans obtained local gravitating mass that's about 3 times larger than the mass in visible stars, and thus concluded that there are "3 dark stars per 1 luminous star". This measurement is very difficult, however, due to the need to take numerical derivatives of noisy data, and the actual number obtained by Jeans is incorrect. In fact, just 10 years later Oort (1932) redid the measurement and concluded that all of the gravity governing the vertical motion of stars in the Milky Way disk can be plausibly explained by stars and putative interstellar gas.

Debate about this is still ongoing (e.g., see Moni Bidin et al. 2012 and Moni Bidin et al. 2015 vs. Bovy and Tremaine 2012). The reason is that despite the reasonable logic of the method, in practice getting a robust measurement is difficult for a variety of reasons (e.g., Hessman 2015). First of all, dark matter contributes only $\approx 10\%$ of local mass density within the disk at the solar radius according to modern models (Bovy and Tremaine 2012). In order to detect the dark matter contribution with such method, error of the measurement has to be small, which is difficult to achieve. One of the potential systematics may stem from the fact that the Milky Way disk is not in complete equilibrium but exhibits large-scale motions due to bending and "breathing" modes (Widrow et al. 2014).

1.9.2 Evidence for dark matter in clusters of galaxies

Based on his earlier ideas published in 1933, Zwicky (1937) proposed to use the virial theorem and gravitational lensing of galaxy clusters to constrain their masses. Gravitational lensing idea had to wait about five decades for observations to improve. Virial theorem Zwicky applied immediately because velocity dispersion could be estimated from galaxy spectra and a number of measurements for galaxies in the Coma cluster already existed.

The virial theorem, which posits that total kinetic and potential energies of a self-gravitating system, K and W in equilibrium are related by $2K = -W$. This relation can be obtained by taking the volume integral of the Jeans equilibrium equations over the system (see §4.8.3 of Binney and Tremaine 2008). Kinetic energy of a system of mass M is $K = M\langle v^2 \rangle$, while potential energy is $W = -\alpha GM^2/R$, where R is physical extent of the system and α is a constant that depends on the internal mass distribution. Thus, measurement of velocity dispersion $\langle v^2 \rangle$ and distance to the system, which gives R , allows one estimate total gravitating mass of the system as $M = \alpha^{-1} \langle v^2 \rangle R$.

Zwicky showed that typical velocity dispersion in the Coma cluster was of order $\langle v^2 \rangle \approx 2100 \text{ km s}^{-1}$ and that the implied dynamical mass is a factor of ~ 500 larger than the sum of all stellar masses of galaxies. This was also shown to be true for the Virgo cluster by Smith (1936), who based his study on Zwicky's ideas published earlier in 1933. Later studies showed that the discrepancy between gravitating and luminous mass was overestimated by more than an order of magnitude due to incorrect distance ladder. At that time Hubble constant was thought to be $H_0 \approx 500 \text{ km/s/Mpc}$ vs. the modern value of $H_0 \approx 67 - 72 \text{ km/s/Mpc}$. The overestimated H_0 resulted in underestimated distances and galaxy luminosities, $L \propto fd_L^2$.

The revised distances decreased the discrepancy between the dynamical mass from the virial theorem and stellar masses because dynamical mass depends linearly on distance, but stellar mass depends quadratically. Other changes reduced the mass discrepancy to a factor of $\sim 5 - 10$. At that point astronomers did not take the mass discrepancy indicated by clusters as something particularly worrying, but deemed it to be likely due to intergalactic gas (e.g., Limber 1959, van Albada 1960).¹⁴

¹⁴This interpretation was a popular explanation of mass discrepancies at the time. Little was known about intergalactic gas, but astronomers strongly believed that it must exist. In fact, Fritz Zwicky was one of the early strong proponents of its existence. Kahn and Woltjer (1959) also interpreted the discrepancy between dynamical and luminous mass of the Milky Way and Andromeda as due to intergalactic gas filling the Local Group volume.

In 1972, after early indications from rocket-based measurements (Cavaliere, Gursky, and Tucker 1971, Meekins et al. 1971), X-ray observations by the UHURU X-ray space observatory revealed presence of extended hot intracluster medium (ICM) emitting at X-ray energies by thermal bremsstrahlung that was found to be smoothly filling intergalactic space within the Coma cluster (Gursky et al. 1971, Forman et al. 1972).

Similarly to the velocity dispersion the temperature is related to the gravitating mass via (see eq. B.1):

$$kT \sim \frac{GM}{R}$$

The temperatures measured for X-ray emission were similar to those implied by galaxy velocity dispersions: i.e., $kT \sim m_p \sigma^2 \sim 1 - 10$ keV. At the same time, estimates of the gas mass associated with X-ray emission of the ICM showed that it can only explain $\lesssim 15\%$ of the gravitating mass implied by galaxy velocity dispersions and temperature. This has ended speculations that mass discrepancy could be explained by the intracluster gas.

Modern measurements of cluster masses include methods that use the Sunyaev-Zel'dovich effect (SZE) observations and weak lensing (WL) shear estimates. Modern estimates of total mass in clusters from velocity dispersions, X-rays, SZE, and WL are consistent to $\approx 10 - 20\%$ (e.g., Applegate et al. 2014, Applegate et al. 2016, de Haan et al. 2016).

1.9.3 Relative motion of M31 and Milky Way

Kahn and Woltjer (1959) presented a simple, but powerful argument for a lot of extra mass in and around galaxies. The argument uses the measured radial velocity of M31 relative to the Milky Way, which is currently measured to be $v_r \approx 109 \pm 4.4$ km/s and the distance between them, $\Delta R \approx 770 \pm 40$ kpc. Assuming that MW and M31 are on the first approach (there is no reason to think otherwise), and that there is no tangential velocity (i.e., MW-M31 pair has no angular momentum), MW and M31 will reduce their separation to zero in less than $\Delta R/v_r \approx 6.8$ Gyrs. Given that the age of the universe is $t_U \approx 13.6$ Gyrs and galaxies must have formed at $t \lesssim t_U$, half period of their orbit should be $P/2 \lesssim t_U + 6.8$ Gyrs and thus $P \lesssim 40.8$ Gyrs. Due to this part of the argument, it is often called *the timing argument*.

For a bound binary system, period of orbit P is related to orbit's semi-major axis a by:¹⁵

$$P^2 = \frac{4\pi^2 a^3}{G(M_1 + M_2)}, \rightarrow (M_1 + M_2) = \frac{4\pi^2 a^3}{GP^2}. \quad (1.13)$$

The semi-major axis of the orbit must be larger than current distance $a > \Delta R$, while $P \lesssim P_{\max} \approx 40.8$ Gyrs. Thus, we can get a lower limit on the sum of masses of MW and M31 as:

$$M_{\text{MW}} + M_{\text{M31}} \gtrsim \frac{4\pi^2 \Delta R^3}{GP_{\max}^2} \approx 2.4 \times 10^{12} M_\odot. \quad (1.14)$$

In contrast, the sum of baryon masses in the MW and M31's disks is $\approx 1.6 \times 10^{11} M_\odot$ – a factor of > 15 discrepancy.

This estimate is approximate and provides only a lower limit. One can refine the estimate by actually solving the parametrized equations of motion of the Milky Way and M31:

$$r = a(1 - \cos \eta); \quad t = \sqrt{\frac{a^3}{GM}} (\eta - \sin \eta), \quad \frac{dr}{dt} = \sqrt{\frac{GM}{a}} \frac{\sin \eta}{1 - \cos \eta}. \quad (1.15)$$

These three equations contain three unknowns, the eccentric anomaly η , semi-major axis a , and the sum of the MW and M31 masses M . Adopting $r = \Delta R$, $t = t_U$, $dr/dt = v_r$ the equations can be solved for these unknowns. For the current estimates of ΔR , t_U , and v_r this gives $M_{\text{MW}} + M_{\text{M31}} \approx 4.2 \pm 0.6 \times 10^{12} M_\odot$ (González, Kravtsov, and Gnedin 2014).

¹⁵This equation is the origin of the Kepler's period law. See, for example, §17.6 [here](#) for its derivation

Is there a way out of this argument? The errors of the radial velocity and distance measurements were sufficiently small even in 1959. The current measurements cannot resolve this discrepancy (in fact, it is somewhat larger than deduced by Kahn & Woltjer). Non-zero tangential velocity would make the orbital period longer, but would also make the orbital energy larger. This would thus require larger $M_1 + M_2$ to reverse expansion of the universe and bring the galaxies to their current separation. In any case, current observations constrain tangential velocity to be < 34.3 km/s at 68% confidence level (Sohn, Anderson, and van der Marel 2012). One could, in principle, argue that the picture of MW and M31 as two point masses evolving in isolation is too simplistic. However, recent checks against pairs of objects forming in realistic environments in cosmological simulations show that the lower limit obtained via the Kahn & Woltjer method is on average within a factor of two of real sum of masses (Li and White 2008, González, Kravtsov, and Gnedin 2014). Thus, the argument appears to be rock solid even though it is simple and relies on a single pair of galaxies. No wonder it withstood 50+ years of astronomical developments!

Nowadays, one can estimate the total mass of the Local Group using cosmological simulations and the Bayesian approach. In such approach halo pairs from cosmological simulations are used as the Monte Carlo sample of Local Groups, which corresponds to the prior that Λ CDM with Gaussian initial conditions and given cosmological parameters is correct. Constructing a likelihood based on select observed properties of real objects, allows one to estimate the posterior of their properties in the Bayesian framework (see Busha et al. 2011). Application of this method to the Local Group using measurements of ΔR and v_t and upper limit on v_t , give estimate of $M_{\text{MW}} + M_{\text{M31}} \approx 2.40^{+1.95}_{-1.05} \times 10^{12} M_\odot$, where errors indicate 90% credible region (González, Kravtsov, and Gnedin 2014), consistent with results of other recent approaches of this kind (Diaz et al. 2014, Peñarrubia et al. 2014).

1.9.4 Rotation curves of galaxies

In the same seminal paper in which Zwicky proposed to use galaxy velocities in clusters as a way to measure cluster masses (Zwicky 1937), he showed that rotation velocities of disk galaxies give lower limit on their masses. This is because rotation velocity of matter on a circular orbit at radius R – or the *circular velocity* – is set by the balance of the centrifugal force and radial gravitational acceleration $V_{\text{circ}}^2/R = g_r = \partial\Phi/\partial R$, so

$$V_{\text{circ}}(R) = \left(R \frac{\partial\Phi}{\partial R} \right)^{1/2}. \quad (1.16)$$

$V_{\text{circ}}(R)$ is called the circular velocity profile or rotation curve of a galaxy.

For a spherically symmetric system $\Phi = -GM(R)/R$ and so

$$\frac{\partial\Phi}{\partial R} = \frac{GM(R)}{R^2}, \quad \text{and} \quad V_{\text{circ}}^2(R) = \frac{GM(R)}{R}. \quad (1.17)$$

In particular, for point mass of mass M , $V_{\text{circ}} = \sqrt{GM/R} \propto R^{-0.5}$; in the literature the latter form is called *Keplerian rotation curve*.

In 1937 Zwicky estimated masses from rotation observed in galaxies that were two orders of magnitude larger than the estimate of their stellar mass. The excess was again overestimated due to incorrect distance ladder used back then, however the sign of the excess and argument itself are correct.

Rotation velocity profiles extending to large radii were measured for the Andromeda galaxy as early as 1939 (Babcock 1939), but it was the measurement by Rubin and Ford (1970), Roberts and Rots (1973), and then later for many other galaxies (e.g., Rubin, Thonnard, and Ford 1978, Rubin, Ford, and Thonnard 1980) that served as a watershed moment in terms of compelling astronomical community that evidence for dark matter in galaxies is real. These measurements showed rotational velocity of gas and stars depends only weakly on the distance to the center, even for distances exceeding the size of galaxy disks.

We know that typical surface density profile of stars in disk galaxies is well described by the exponential form (see §1.3 and Figure 1.4). Circular velocity profile in the mid-plane of a thin disk with exponential surface density profile is (see §2.6 in Binney and Tremain’s “Galactic dynamics” book):

$$V_{\text{circ,disk}}^2(R) = 4\pi G \Sigma_0 R_d y^2 [I_0(y)K_0(y) - I_1(y)K_1(y)], \quad (1.18)$$

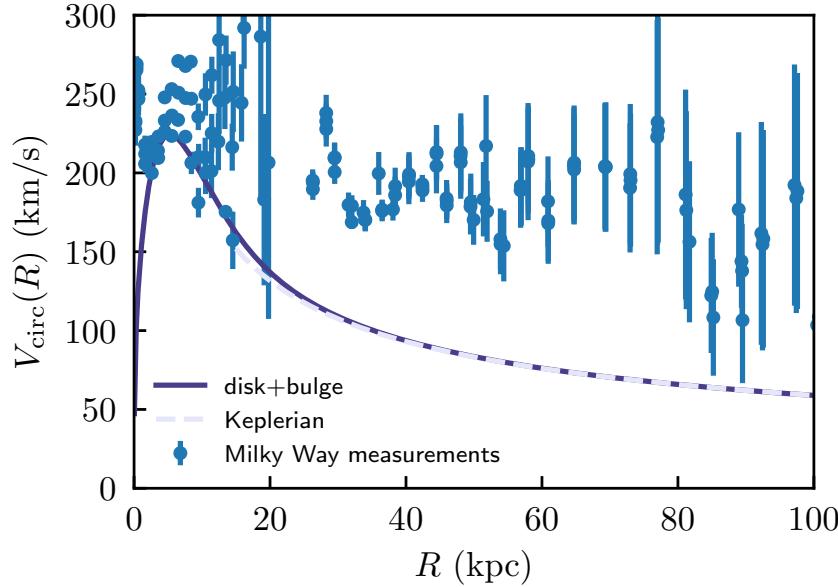


Figure 1.17: Circular velocity profile of the exponential disk (solid blue line) with mass $M_d = 7 \times 10^{10} M_\odot$ and disk scale length $R_d = 2.63$ kpc, given by eq. 1.18 with the bulge of mass $M_b = 0.15M_d$ with a spherically symmetric mass distribution described by the Hernquist (1990) profile. The light dashed line shows the Keplerian rotation curve, $V_{\text{circ},K} \propto R^{-0.5}$. The rotation velocity due to stars is predicted to decrease at large radii, while rotation velocities of gas and stars observed in the Milky Way shown by points with error bars stay approximately constant at these radii (measurements are from compilation in Table 2 of Bhattacharjee, Chaudhury, and Kundu 2014).

where I_n and K_n are the modified Bessel functions of the first and second kind, respectively, R_d is the disk scale length, and $y = R/(2R_d)$. Recall that the total mass of the exponential disk is $M_d = 2\pi\Sigma_0 R_d^2$.

Circular velocity as a function of distance to the galaxy center predicted for exponential disk is shown in Figure 1.17 for a galaxy with disk parameters close to those of the Milky Way, with a small contribution near the center from the bulge of mass $M_b = 0.15M_d$. The bulge is assumed to be spherically symmetric with the mass distribution described by the Hernquist (1990) profile.

The figure shows that rotation velocity due to disk and bulge is predicted to decrease at large radii $R \gtrsim 10$ kpc $\approx 3R_d$, while observed rotation velocities stay approximately constant at these radii. The difference between observed “flat” rotation curves and the predictions for observed distribution of stars and gas is attributed to the contribution of dark matter halo, in which luminous matter in galaxies is embedded. This difference is generic and is observed in all galaxies, in which such measurements were done.

Dark matter contribution to the rotation curve dominates that of the baryons at $R \gtrsim 1.5R_d \approx 5$ in the case of the Milky Way, while observed stars and gas dominate at smaller radii. The transition between baryon-dominated inner region and dark matter-dominated outer region is rather smooth – i.e., rotation curve does not show a strong feature in such transition. This behavior is called the *disk-halo conspiracy* and indicates that dark matter and observable matter are couple or “know” about each other. This coupling is manifested in the fact that one can construct a single functional form for rotation curves of galaxies of different luminosities – the so-called universal rotation curve – as a function of luminosity only (Persic, Salucci, and Stel 1996). In other words, the transition from the baryon-dominated inner regions to dark matter dominated outer regions is a function of galaxy luminosity, which implies that dark matter contribution as a function of radius “knows” about galaxy luminosity. This is one of the strongest arguments in the arsenal of the Modified Newtonian Dynamics (MOND) model advocates. This model explains flat rotation curves of galaxies by modification of gravity on cosmological scales, rather than by dark matter.

1.9.5 Evidence from disk stability

Additional argument based on the fact that self-gravitating disks are unstable to bar instability was presented by Ostriker and Peebles (1973). Although a sizeable fraction of disk galaxies do have bars, these are confined to inner regions and the overall disks are not undergoing violent instabilities. Simulations of self-gravitating disks by Ostriker and Peebles (1973) showed that secular stability of observed disks demands that a stabilizing massive spherical halo must exist around galactic disks.

This fact was then confirmed and quantified by a number of studies (e.g., Efstathiou, Lake, and Negroponte 1982, Christodoulou, Shlosman, and Tohline 1995, see also a summary and application in Mo, Mao, and White 1998). Numerical simulations showed that long term stability of disks requires that the following dimensionless parameter is larger than some threshold:

$$\epsilon_m = \frac{V_{\max}}{\sqrt{GM_d/R_d}}, \quad (1.19)$$

where $V_{\max} = \max V_{\text{circ,tot}}(r)$ is the maximum rotation velocity within the disk, M_d is mass of the disk and R_d is the exponential scale-length of the disk. Collisionless disks (when stars dominate mass) are stable for $\epsilon_m \gtrsim 1.1$ (Efstathiou, Lake, and Negroponte 1982), while gaseous disk (gas dominates mass) is stable when $\epsilon_m \gtrsim 0.9$ (see S 3.4.4 in Christodoulou, Shlosman, and Tohline 1995). Thin isolated exponential disk has $\epsilon_m \approx 0.63$ and thus additional contribution to V_{\max} from a dark halo is required to bring ϵ_m to the stable regime.

1.9.6 Evidence for dark matter from the cosmic microwave background

Nowadays, direct measurement of total cosmic density of matter can be made using anisotropies of temperature and polarization of the cosmic microwave background (CMB, e.g., Komatsu et al. 2011, Planck Collaboration et al. 2015). The CMB temperature anisotropies are a sensitive probe of dark matter because relative height of the acoustic oscillation peaks and their overall shape depend on the mean dark matter density. Note that the density is measured with h^2 factor, so the actual number depends on the Hubble constant.

However, CMB measurements were strongly (albeit indirectly) indicating dark matter long before the anisotropies were measured. In particular, by ~ 1980 upper limits on the temperature anisotropy have improved to $\Delta T/T \lesssim 10^{-4}$ (e.g., Melchiorri et al. 1981). With temperature fluctuations so low, it was impossible to rely only on baryons to explain formation of observable structures, such as galaxies and galaxy clusters. This is because prior to recombination baryons are coupled to photons and their perturbations do not grow, but oscillate in the so-called *baryonic acoustic oscillations*. Thus, in a purely baryonic universe small temperature fluctuations would indicate small initial density fluctuations. As we will see later, the rate at which perturbations were expected to grow from the redshift of recombination, z_r to a given z is given by $\approx (1+z)/(1+z_r)$. Given that $z_r \approx 1100$, if the initial density perturbation indeed had an amplitude of $\lesssim 10^{-4}$, galaxies and clusters would not have enough time to form by the present epoch, $z=0$.

The solution that was quickly proposed by several groups was to postulate existence of weakly interacting dark matter particles motivated by supersymmetry models of particle physics (Blumenthal, Pagels, and Primack 1982, Bond, Szalay, and Turner 1982, Peebles 1982). The *Cold Dark Matter* paradigm of structure formation was born.

1.10 Additional reading

For historical context on morphological classification, it is worth reading a brief three-page summary on Hubble's morphological classification in Hubble (1926a), or you may check the longer version Hubble (1926b) from the same year. Recent review by Sandage (2005) is an interesting window (in particular S 3 and S 4.2-4.4) into historical developments in galaxy morphological classification, as is the detailed paper by Hart and Berendzen (1971) about the history of how Hubble came up with his classification scheme and some of the rival scheme development.

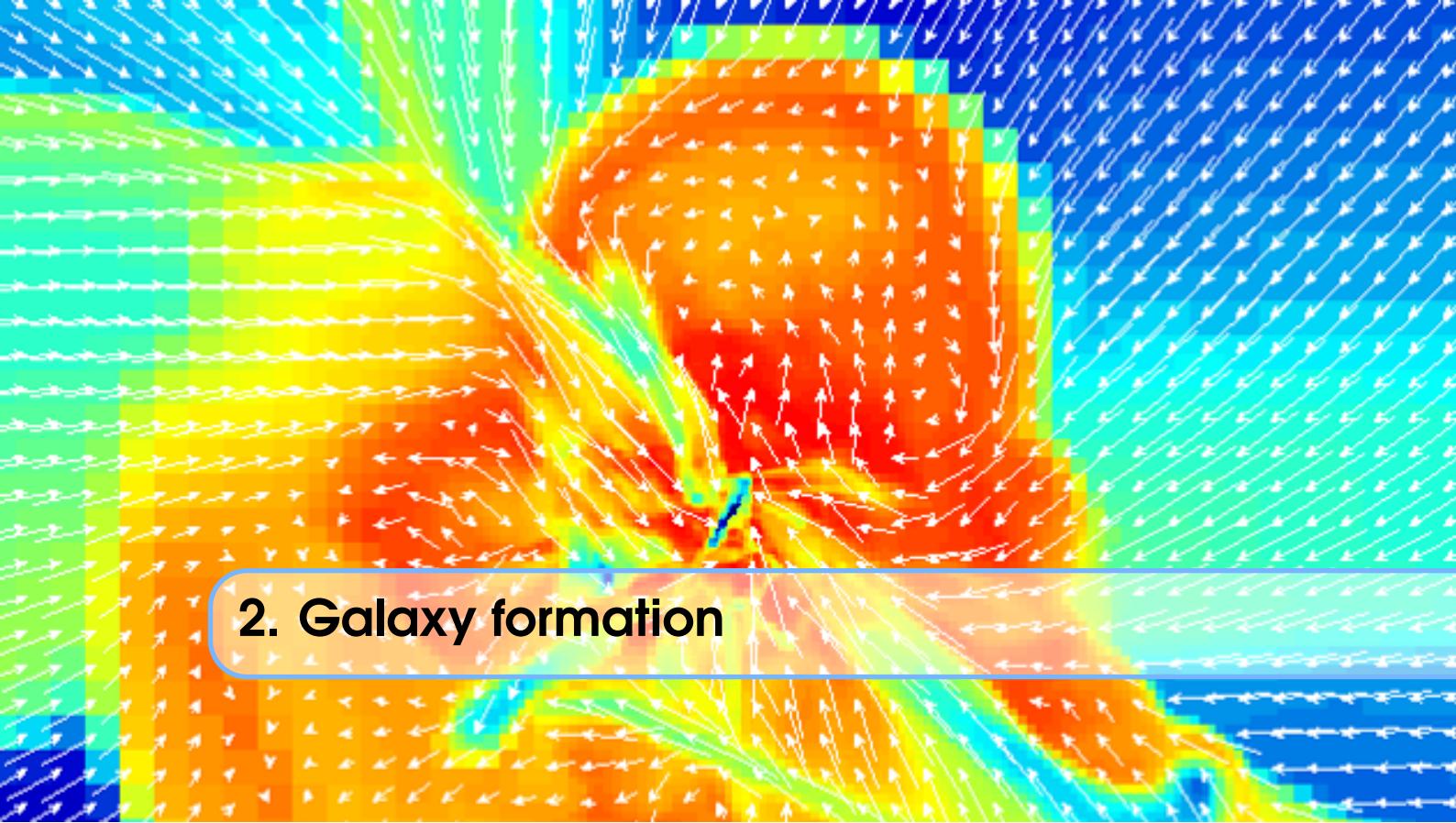
A good overview of galaxy properties at a level appropriate to both advanced undergraduate and graduate students can be found in S 1.1.1-1.1.2, 1.1.5, S 1.3-1.4 of “Galaxies in the Universe” book by Sparke and Gallagher (2007, hereafter GS). Sections S 2.1, 2.2 (pp 25-36), 2.3.1, 2.3.2 (pp 37-45), 15.2.2 (pp 659-660) of the “Galaxy formation and evolution” book of Mo, van den Bosch, and White (2010, hereafter MvdBW). Up-to-date, research level discussion of galaxy properties can be found in S 2.1-2.4 and S 3.1-3.2 of review by Blanton and Moustakas (2009). A comprehensive recent review of the Milky Way properties and its place in the realm of galaxies is presented by Bland-Hawthorn and Gerhard (2016).

For a review of history of large-scale structure mapping, see Thompson and Gregory (2011)¹⁶ and Einasto (2009).

1.11 Ideas for exercises and explorations

1. Load the SDSS DR8 main galaxy catalog provided to you (see the example ipython notebook). Select a sub-sample of galaxies within a certain redshift range and apparent magnitude. Draw objects randomly from this sub-sample and display their images as shown in example python scripts. Classify these random draws morphologically using Hubble’s classification scheme. Examine spectra of these objects and comment on the trends of spectral shapes with galaxy morphology.
2. Try to construct the Hubble’s tuning fork morphological diagram by such random drawings. Is it easy or difficult? What kind of galaxies are most common when you draw them randomly?
3. Try to make cuts on other galaxy properties discussed in this chapter (e.g., luminosity, color, concentration, surface brightness). Is it easier to identify galaxies of particular morphological type with such cuts? With what galaxy properties does morphology correlate best? What are the cuts in particular properties that allow you to select a given morphological type best?
4. The Milky Way half-light radius is $\approx 4.0 \pm 0.5$ kpc. For example, Bovy and Rix (2013) estimate the scale-length of the Milky Way disk to be $R_d = 2.15 \pm 0.14$ kpc, while McMillan (2011) estimates $R_d \approx 3.1 \pm 0.3$ kpc and half-light radius for exponential disk is $r_{50} \approx 1.68R_d$. Compute the r -band surface brightness of the Milky Way as it would be seen by an extragalactic observer (and show how you did it in intermediate steps). Plot the SDSS galaxies in the $\mu_r - M_r$ plot similar to the $\mu_r - m_r$ plot shown above and place your estimate for the Milky Way as a distinct point on the plot. Where is the Milky Way located relative to other disk galaxies in this diagram?
5. Select ten (or more) galaxies nearest to the Milky Way in half-light radius, M_r , and color discussed in §above and display a grid of thumbnail SDSS images of these objects (you can use example scripts in the ipython notebook that produced figures in this chapter). Examine images of these Milky Way analogues and describe their morphology. Do you notice any similarities? Compare morphologies of these analogues to typical morphologies of galaxies in the same region of the $(g - r) - M_r$ or $\mu_r - C_r$ diagram.

¹⁶Available as an online version here.



2. Galaxy formation

The header image above is a slice through cosmological simulation of galaxy formation. It shows gas flows into and out of a forming galaxy at $z = 4$. It may look scary: galaxy formation is too complicated and requires complex and computationally expensive simulations to figure out! This is true to some extent. However, the main features of galaxy formation are simpler than one might think. Therefore, in what follows we will forego simulations (which are outside the scope of this course anyway) and will instead focus on results that could be understood using simple, physical models.

As mentioned in §1.9.6, by ~ 1980 the upper limits on the temperature anisotropy of the Cosmic Microwave Background have improved to $\Delta T/T \lesssim 10^{-4}$. Such low level of anisotropies made it impossible to explain formation of observable structures within the age of the universe from solely baryonic seed density perturbations. This is because prior to recombination baryons are coupled to photons and their perturbations do not grow, but oscillate in the so-called *baryonic acoustic oscillations*.

The solution proposed by several groups was to postulate existence of weakly interacting dark matter particles motivated by supersymmetry models of particle physics (Blumenthal, Pagels, and Primack 1982, Bond, Szalay, and Turner 1982, Peebles 1982), the scenario that has become the *hierarchical Cold Dark Matter* (CDM) paradigm of structure formation. In this scenario, baryons co-evolve with dark matter. Density peaks in the latter collapse to form extended spheroidal *halos*, because it is assumed that dark matter cannot dissipate the kinetic energy it acquires during collapse. In contrast, normal “baryonic” matter (the baryons) can experience dissipation, which allows them to condense in the centers of dark matter halos forming dense regions characteristic of observed galaxies. Complicated nonlinear processes in these dense regions, such as star formation and stellar feedback, appear to shape galaxy formation and are still a subject of active investigations. It is clear, however, that the resulting properties of galaxies tightly correlate with the properties of their host dark matter halos, in particular their mass and formation history.

The CDM scenario still had to postulate the initial properties of perturbations, in particular their spectrum (see §B.2 for details of the definition of power spectrum of density perturbations used in cosmology and galaxy formation models). The guidance on the shape was provided by arguments for the *scale-free initial spectrum* (see next section) laid out in papers by Harrison (1970) and by Zeldovich (1972).

At about the same time, the concept of inflation was proposed as an outgrowth of exploration of consequences of supersymmetry models to the abundance of monopoles and to explain the observed flatness of the universe and its apparent uniformity of properties in seemingly causally disjoint regions (Starobinsky

1980, Guth 1981). After this concept was introduced, it was quickly realized that quantum fluctuations of the inflaton field could also provide natural seeds for structure formation (Guth and Pi 1982, Starobinskii 1983, Bardeen, Steinhardt, and Turner 1983). Remarkably, the spectrum of perturbations generically predicted by the inflationary models was close to the scale-free spectrum advocated by Harrison and Zeldovich. A brief review of the inflationary origin of perturbation and physics of their evolution in the linear regime is given below. Models of non-linear collapse are discussed next and are followed by discussion of processes that accompany formation of galaxies.

2.1 Inflationary origin of the seed density perturbations

In the current cold dark matter (CDM) paradigm of structure formation primordial density perturbations originate from the quantum fluctuations in the energy density of the scalar field(s) responsible for the early inflationary period of the expansion of the universe. These models generically predict that perturbations are close to Gaussian. This is because inflaton field is usually assumed to be weakly coupled, so its quantum fluctuations sum up to a Gaussian pdf via the central limit theorem. With stronger coupling, deviations from Gaussianity are expected, but are disfavored by current observations.

Inflation scenario explains flatness (geometry of observed universe is very close to Euclidian) and horizon problem (uniformity of the universe in different directions), if the inflation epoch lasts for $\approx 60 - 70$ e-foldings. The expansion during inflation is exponential (or close): i.e., $a(t) \propto \exp[H(t)t]$, where $H(t) = \dot{a}/a$ is the Hubble rate of expansion. To keep the universe exponentially inflating, $H(t)$ should be changing slowly ($H(t) \approx \text{const}$) so that e-folding time scale is nearly constant. This is usually achieved by assuming that potential of the inflaton field ϕ is quite flat, so that $\dot{\phi} \ll V(\phi)$. This results in the *slow-roll* evolution of the potential towards the minimum in $V(\phi)$, analogously to a solid ball slowly rolling down a slightly inclined plane.

Fast exponential expansion for $\sim 60 - 70$ e-foldings corresponds to the increase of scales by $> e^{60-70} \sim 10^{26} - 10^{30}$ so atomic scale ($\sim 10^{-10}$ cm) becomes pc ($\sim 10^{18} - 10^{20}$ cm) scale. Note that this scale is physical. The corresponding comoving scale is very large because universe has expanded by additional factor of $\sim 10^{30}$ after the end of inflation (Lineweaver 2005). Thus, even quantum scales get inflated to macroscopic and cosmological scales. The full consideration of quantization of the inflaton and amplitude evolution of its amplitude is well beyond the scope of this course, but if you are curious you can find details in Anthony Lewis's [notes](#) (see sections B and C). We can consider a hand waving approximate argument, which nevertheless captures the physics of what's going on.

A quantum fluctuation in the inflaton energy δE can be viewed as a wave packet of wavelength $\lambda \sim hc/\delta E$. The horizon scale is

$$R_h = a(t) \int_0^t \frac{cdt'}{a(t')} \sim ct \sim \frac{c}{H(t)} \approx \text{const}. \quad (2.1)$$

This means that during inflation ($H(t) \approx \text{const}$) $R_h \approx \text{const}$, but physical scales increase as $a(t) \propto \exp(Ht)$ and so most scales will become larger than the horizon with the smallest scales leaving the horizon last. When the wavelength of perturbation becomes equal to horizon $\lambda \sim R_h \propto 1/\delta E \sim \text{const}$ - i.e., amplitude of energy perturbations is approximately the same when they achieve horizon size. In the GR, energy perturbations source potential perturbations. Given that the latter have units of energy, they will scale in the same way - i.e., amplitude of potential fluctuations is the same when they reach horizon size: $\delta\Phi \approx \text{const}$ at t when $\lambda(t) \approx R_h$.

The potential fluctuations are related to the density fluctuations via the Poisson equation: $\nabla^2 \delta\Phi = 4\pi G \bar{\rho} \delta$, or in the Fourier space: $k^2 \delta_{\Phi,k} \propto \delta_{\rho,k}$. Thus, the Fourier modes of the density and potential fluctuations are related by a simple k^2 factor.

While inflation is ongoing the amplitude of density perturbations larger than horizon does not grow with time. This is because solution to the density perturbation evolution equation in this regime (see §B.4) is $\delta \approx c_1 + c_2 \exp(-2Ht)$, where c_i are constants. The second term is driven to tiny values quickly

during inflation and thus $\delta \approx c_1$ for super-horizon perturbations. Given that the perturbations of energy and potential have the same amplitude when their size reaches horizon scale and they do not evolve afterwards, during inflation the amplitude of the super-horizon potential fluctuations is independent of scale. Thus, the dimensionless power spectrum of potential fluctuations (see B.2 for the definitions of $\Delta^2(k)$ and $P(k)$) is also constant $\Delta_{\Phi,k} \equiv k^3 P_\Phi(k)/(2\pi^2) = k^3 \langle |\delta_{\Phi,k}|^2 \rangle \approx \text{const}$, where $\langle \rangle$ brackets indicate average per unite volume in the Fourier space which scales as k^{-3} . This is why it is often said that inflation predicts a nearly *scale-invariant perturbation spectrum*.

The power spectrum of density perturbations originating from inflation is thus close to $\Delta_\delta^2(k) \equiv k^3 P_\delta(k)/(2\pi^2) = k^3 \langle |\delta_{\delta,k}|^2 \rangle = k^3 \langle |k^2 \delta_{\Phi,k}|^2 \rangle \propto k^4$ and $P_\delta(k) \equiv \langle |\delta_k|^2 \rangle \propto k$. This spectrum is also called scale-invariant (which is somewhat confusing) or often the *Harrison-Zeldovich spectrum* (Harrison 1970, Zeldovich 1972).¹

The nearly scale-invariant spectrum of fluctuations is a generic prediction of inflation scenario, because the slow-roll period and the associated regime of $H \approx \text{const}$ are generic to this scenario, even as details of models vary. At the same time, these models predict that there is a small deviation from perfect scale invariance. The current constraints from the Planck observatory indicate initial density perturbation spectrum very close to the Gaussian (deviations from Gaussianity were not detected at a fairly stringent level) and the power spectrum is close to scale free, $P(k) \propto k^n$ with $n = 0.96 \pm 0.007$.

2.2 Evolution of density perturbations in the linear regime

After inflation ends, relevant perturbations that seeded formation of structures in the universe all have wavelengths larger than horizon and have the nearly scale-invariant spectrum, $P(k) \propto k^n$ with $n \approx 1$. The expansion starts to decelerate, i.e. $H(t)$ decreases, as there is no inflaton or vacuum energy driving accelerated expansion. The horizon scale, $R_h \sim c/H(t)$, thus grows with time, while the ratio $a(t)/R_h \sim \dot{a}$ decreases. The physical wavelengths of perturbations λ are stretched by expansion as $\propto a(t)$ and thus λ/R_h decreases. For some perturbations, there will be a time when their λ/R_h becomes less than unity – at this point such perturbations are said to *enter the horizon*.

Tremendous expansion during inflation adiabatically cooled the universe, while theory of primordial nucleosynthesis, initially proposed in the 1940s Alpher and Gamow, requires that later on universe was quite hot in order to explain observed abundances of hydrogen, helium and other light elements. Thus, at some point after inflation, the universe should have experienced the period of *reheating* during which the inflaton field oscillates around the minimum of $V(\phi)$ and damps its kinetic energy producing relativistic particles and radiation. At the end of this process universe is hot in the sense that all particles that fill it are relativistic and behave as radiation. This marks the beginning of the *radiation-domination era* of cosmic expansion and beginning of the *hot Big Bang*.

Without any assumptions about inflation, we know that *radiation-dominated era* preceded *matter-domination epoch*, because radiation energy-density scales $\propto a^{-4}$ and matter non-relativistic energy-density scales as $\propto a^{-3}$. Given the current values of radiation and matter densities, $\Omega_r \approx 8.2 \times 10^{-5}$ and $\Omega_m \approx 0.27$, the radiation must have dominated energy density at $a < a_{\text{eq}} \approx \Omega_r/\Omega_m \approx 0.0003$ ($z \approx 3300$). At $a > a_{\text{eq}}$ radiation becomes sub-dominant and energy-density of the universe is dominated by matter density for a while.

The evolution of perturbations depends on whether the universe is in the radiation or matter domination regime and on whether perturbation wavelength is larger or smaller than the current horizon scale.

¹Models of inflation appeared only in the late 1970s-early 1980s. But the scale-free primordial density fluctuation spectrum that naturally emerges from them, $P(k) \propto k^n$ with $n \approx 1$, was advocated a decade earlier by Harrison (1970) and by Zeldovich (1972). The details of the original arguments do not matter because they were not correct. However, the general sense of the arguments was that given the vast range of scales from the horizon to sub-kpc, if spectrum was too steep, there would be too much small-scale structure, and if it was too shallow, there would be too much large-scale structure relative to small-scales. It was clear already around 1970s that density fluctuations on super-cluster scales were smaller than on cluster and galactic scales, but not negligible.

2.2.1 Linear evolution of perturbations in the radiation domination regime

To understand what happens to the perturbations after their size becomes larger than horizon, we need to consider the equation that describes evolution of perturbations. In the Newtonian approximation such equation is derived from the continuity and Euler equation and considered in conjunction with the equation of state describing thermodynamics of the components of the universe (see, e.g., S 4.1 in the MvdBW book). In the general relativity (GR) theory a similar equation can be derived via a more involved derivation. These equations show that on super-horizon scales during inflation amplitude of perturbations does not grow. Thus, the constancy of the potential fluctuation amplitude when perturbation exits the horizon is preserved during inflation for superhorizon wavelengths (i.e., all relevant wavelengths).

Thus, after the inflation epoch, the perturbations that will be re-entering horizon at some later time will have a nearly scale-free spectrum.

2.2.2 Evolution of perturbations smaller than horizon

When the inflationary epoch ends, expansion starts to decelerate (i.e., $H(t)$ decreases), as there is no inflaton or vacuum energy driving accelerated expansion and expansion is decelerating. The horizon scale, R_h , thus grows, while the ratio $a(t)/R_h \sim \dot{a}$ decreases. This means that scales that previously left the horizon will begin to enter it, with the scales that left last entering first.

When perturbation enters the horizon ($\lambda < R_h$), baryons and radiation within it become subject to pressure forces.

When small scale perturbations enter the horizon during radiation domination era, photons and baryonic matter are coupled via the Compton scattering and their perturbations cannot grow due to very large pressure forces ($c_s \sim O(c)$). Thus, they oscillate. Weakly interacting non-baryonic dark matter is not coupled to the photon-baryon plasma, but cannot grow significantly because energy-density is dominated by radiation. So perturbation amplitude will be stalled (actually, it grows but only logarithmically). The factor by which amplitude will lag compared to perturbations above the horizon scale with the time since perturbation entered the horizon: smaller perturbations will experience larger amplitude lag. This results in *a characteristic bend in the power spectrum at the wavenumber corresponding to the horizon size at the matter-radiation equality epoch*. If the fluctuations did not grow at all after they enter the horizon, their power spectrum would lag by $(a/a_{\text{enter}}(k))^4$ (because power spectrum for superhorizon scales still grows as $P \propto D_+^2 \propto a^4$ during radiation domination), where $a_{\text{enter}} \propto \lambda \propto k^{-1}$ is the epoch when the mode entered the horizon. However, perturbations actually grow logarithmically, so the actual lag is smaller. Thus, the primordial power spectrum at $k > k_{\text{eq}}$ will be decreased by the k dependent factor that approaches k^4 asymptotically and so the result is $P(k) \propto k^{n_s}$ at $k \ll k_{\text{eq}}$ and $n_s \approx 1$ and $P(k) \propto k^\alpha$ where α approaches -3 asymptotically and logarithmically for $k \rightarrow \infty$.

After the epoch at which energy density of non-relativistic matter exceeds that of relativistic component, the expansion rate of the universe decelerates more rapidly because relativistic component now does not contribute much (so there is less energy-density to slow down the expansion). On most scales relevant for structure formation gravity would still be overwhelmed by pressure forces, and plasma would simply oscillate.

If the universe consistent only of baryons, the period between the perturbation entering horizon and the epoch of recombination would be "wasted" as perturbations would not grow, but oscillate. In this case, what we see in the temperature perturbations of the CMB would be the actual amplitude of perturbations at the recombination epoch $z \approx 1100$. As we discussed, by 1982 upper limits on the CMB temperature fluctuations became sufficiently stringent that it was clear that to explain formation of observed structures one needs to assume that they are seeded by a non-baryonic component.

Non-baryonic weakly interacting dark matter component would not be subject to pressure forces and thus $\omega^2 = -4\pi G\rho\delta_k$ in equation B.17 would be always negative. This then would result in non-oscillating growing and decaying solution modes. The growing mode would result in perturbation growth, which can happen between the end of inflation and the epoch of recombination (albeit at a logarithmic rate during radiation domination). This is what "saves" the ability of the CDM model to seed the observed structure formation with fluctuations of suitably large amplitude, despite the low amplitude of baryon density perturbations indicated

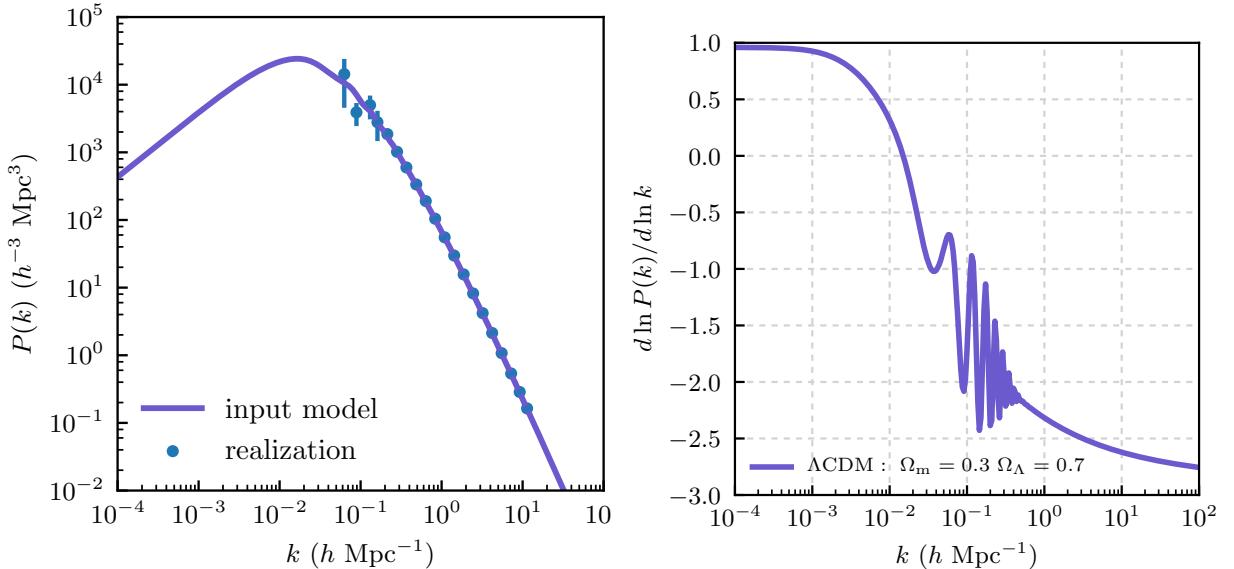


Figure 2.1: Power spectrum, $P(k)$, (left panel) and its logarithmic slope (right panel) at $z = 0$. The assumed cosmology is flat Λ CDM with $H_0 = 70 \text{ km/s/Mpc}$, $\Omega_{\text{m},0} = 0.27$, $\Omega_{\text{b},0} = 0.0457$, $\sigma_8 = 0.82$, $n_s = 0.96$. In the left panel, the power spectrum of the model is shown by the solid line and has the form $P(k) \propto k^{n_s}$ at $k \ll 10^{-2} h \text{ Mpc}^{-1}$, while at $k \gg 10 h \text{ Mpc}^{-1}$ $P(k)$ approaches slope of -3 . In addition, the points show the power spectrum of a random realization of the Gaussian field with such spectrum in a $100 h^{-1} \text{ Mpc}$ box, slice through which is shown in Figure 2.3 below. In a limited volume, the number of modes with wavelength close to the box size is limited and the power spectrum has large variance. At smallest scales, the largest k is simply determined by the size of the grid used to estimate the power spectrum.

by the CMB temperature fluctuations measurements.

2.2.3 Summary perturbation evolution in the linear regime

After reheating (beginning of the hot Big Bang), all scales relevant for structure formation are larger than the horizon. During the radiation-domination era, the amplitude of density perturbations grows as a^2 , while potential fluctuation amplitude remain constant until they enter the horizon. When perturbation wavelength becomes smaller than horizon, its different components behave differently: photon-baryon plasma oscillates due to the high radiation pressure, while non-baryonic dark matter perturbation amplitude grows logarithmically. The logarithmic growth of sub-horizon dark matter perturbations results in their lag relative to the super-horizon ones, which transforms the primordial $P(k) \propto k^{n_s}$ power spectrum into the power spectrum with a bend at k_{eq} corresponding to the scale that entered the horizon at $a = a_{\text{eq}}$ with asymptotic power law spectra: $P(k) \propto k^{n_s}$ for $k \ll k_{\text{eq}}$ and $P(k) \propto k^3$ for $k \gg k_{\text{eq}}$ with the logarithmically changing slope at intermediate k . The resulting power spectrum is shown in the left panel of Figure 2.2.

Still, the ability of dark matter fluctuations to grow while baryon fluctuations oscillate is what makes the formation of observed structures feasible in the limited age of the Universe. In other words, the observed amplitude of temperature fluctuations in the CMB is too small to ensure structure formation with the observed rate in 13.5 Gyr. The seed fluctuations are due to dark matter density perturbations, however, which have much larger amplitude because they could grow between the start of the hot Big Bang and recombination epoch.

With the onset of the matter domination era, the dark matter density perturbations on all scales grow as $\propto a$ (while potentials are constant), baryons still oscillate due to radiation pressure. After recombination baryonic matter decouples from radiation and is now able to collapse into potential wells created by dark matter. The amplitude of the baryon perturbations thus grows rapidly and catches up with the amplitude of

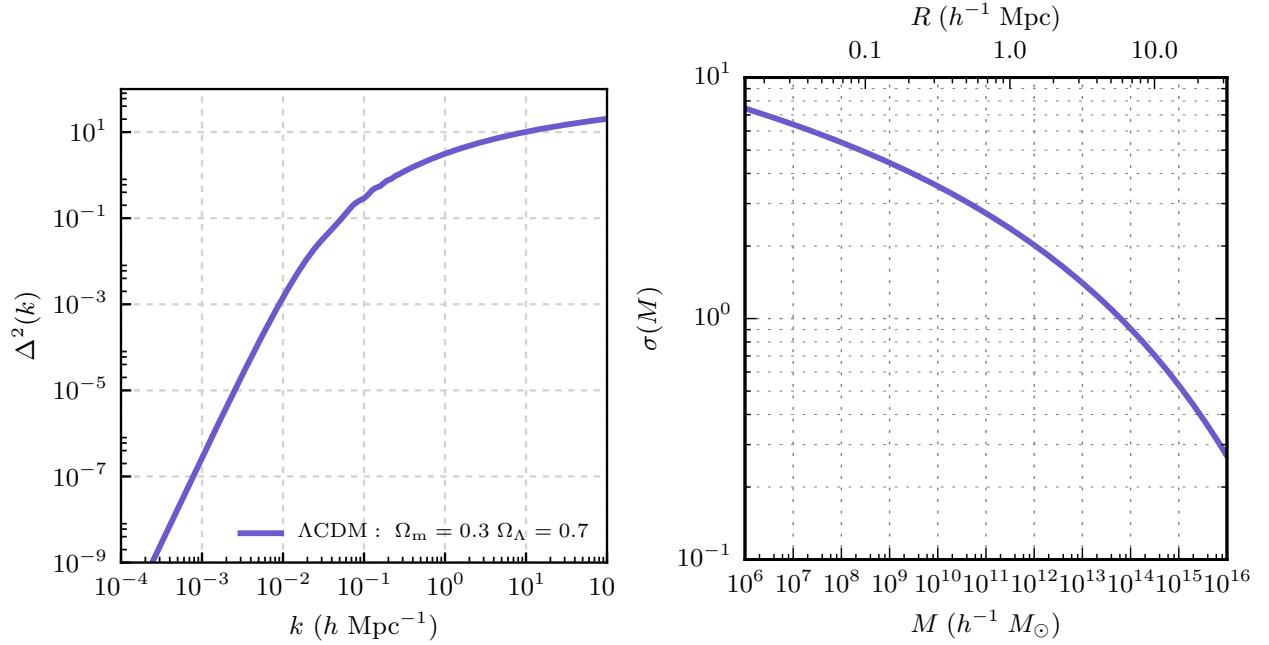


Figure 2.2: Dimensionless power spectrum, $\Delta^2(k) = k^3 P(k)/(2\pi^2)$, (left panel) and $\sigma(M)$ (right panel) at $z = 0$. In the right panel the top axis shows the comoving Lagrangian radius corresponding to mass M : $R = 3M/[4\pi\bar{\rho}_m(z = 0)]$. The assumed cosmology is flat Λ CDM with $H_0 = 70 \text{ km/s/Mpc}$, $\Omega_{m,0} = 0.27$, $\Omega_{b,0} = 0.0457$, $\sigma_8 = 0.82$, $n_s = 0.96$. In the left panel, the power spectrum of the model is shown by the solid line and has the form $P(k) \propto k^{n_s}$ at $k \ll 10^{-2} h \text{ Mpc}^{-1}$, while at $k \gg 10 h \text{ Mpc}^{-1}$ $P(k)$ approaches slope of -3 . In addition, the points show the power spectrum of a random realization of the Gaussian field with such spectrum in a $100 h^{-1} \text{ Mpc}$ box. In a limited volume, the number of modes with wavelength close to the box size is limited and the power spectrum has large variance. At smallest scales, the largest k is simply determined by the size of the grid used to estimate the power spectrum. In the right panel, the rms variance was calculated using the top-hat filter (see eq. B.8).

dark matter perturbations. Thereafter, both evolve as $\delta \propto a$ at scales larger than the Jeans scale corresponding to the temperature of the baryon gas. After reionization this temperature increases sharply and growth of perturbations on smaller scales is suppressed.

2.3 Evolution of perturbations in the non-linear regime

2.3.1 The spherical tophat collapse model

The *top-hat spherical collapse model* (Gunn and Gott 1972, Lahav et al. 1991), considers a spherically-symmetric density fluctuation of the initial *comoving* radius R_i , amplitude $\delta_i (< R_i)$ and mass $M = (4\pi/3)(1 + \delta_i)\bar{\rho}_{m0}R_i^3$, where R_i is comoving of the perturbation with the corresponding *physical* radius $r_i = a(t_i)R_i$, and $\bar{\rho}_{m0}$ is the mean *matter* density of the Universe at $z = 0$. In what follows we will assume that there is no dark energy like component of energy-density, $\Omega_\Lambda = 0$, but the differences for cosmologies with $\Omega_\Lambda > 0$ cosmologies will be discussed at the end.

Given that the perturbation is spherical, once the initial conditions are defined, the subsequent evolution of the perturbation is independent of the evolution of the Universe thanks to the first Newton's theorem, or its relativistic analogue – the Birkhoff theorem. However, it is still useful to contrast the evolution of the perturbation with the evolution of the universe on average.

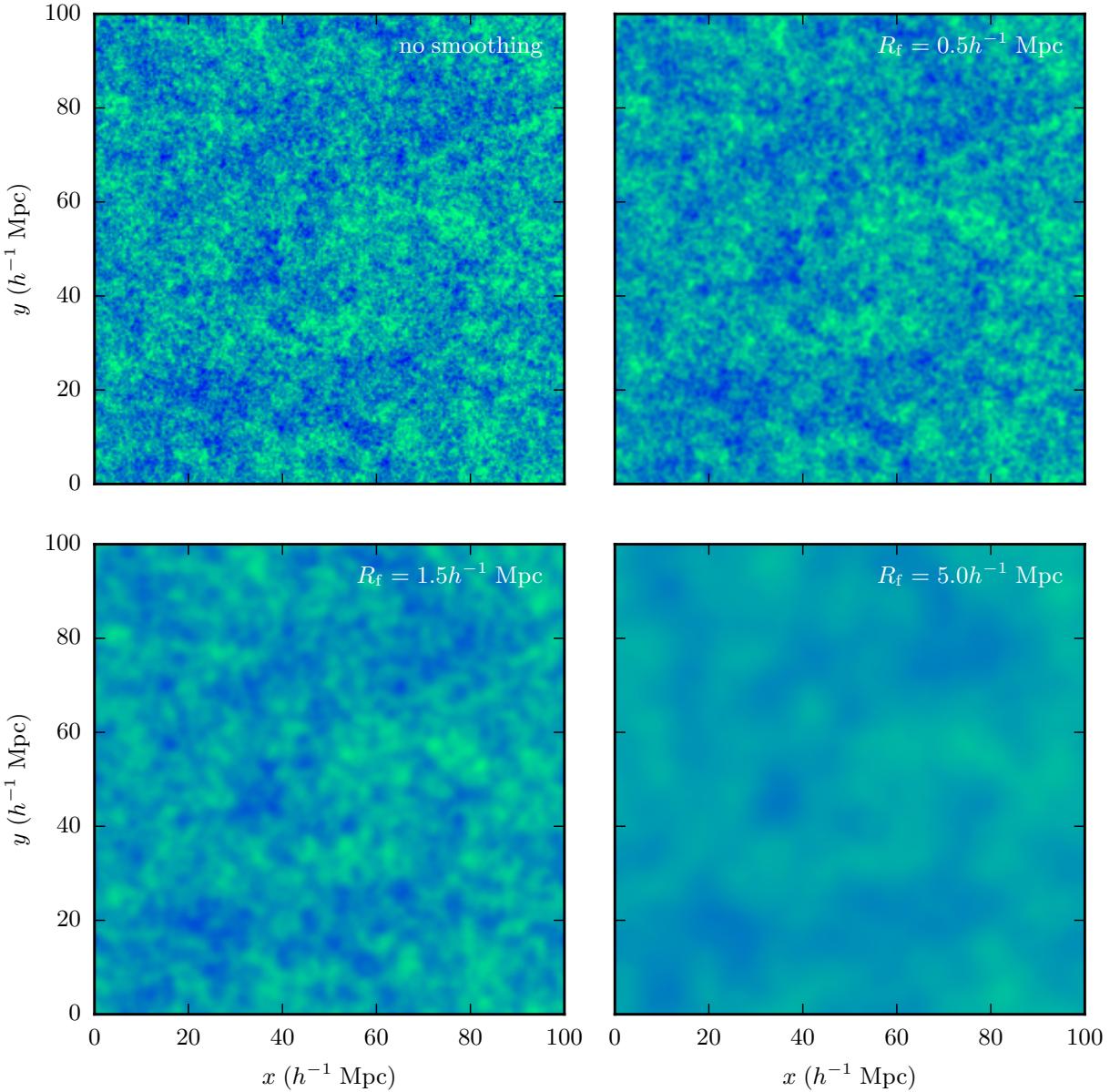


Figure 2.3: Overdensity, $\delta \equiv \rho/\rho_m - 1$, in a slice through a realization of linear Gaussian field in $100h^{-1}$ Mpc cube on a 256^3 grid with the power spectrum shown in Figure 2.2, but at $z = 30$. The top left panel shows δ field without any smoothing, while the other panels show the same field smoothed with the top-hat filter of radii 0.5 , 1.5 , and $5h^{-1}$ Mpc. In all panels the field is plotted on the same color stretch, and one can see that the peaks in smoothed fields of larger R_f have smaller amplitudes.

Initial conditions and conditions for collapse.

The density within the perturbation will evolve as $\rho = 3M/(4\pi r^3)$, while the mean density of the universe evolves as $\bar{\rho} = \bar{\rho}_{m0}a^{-3}$. Thus,

$$1 + \delta(t) = \frac{\rho}{\bar{\rho}} = \frac{3M}{4\pi\bar{\rho}_{m0}} \frac{a^3}{r^3} = R_L^3 \frac{a^3}{r^3}, \quad (2.2)$$

where $R_L = [3M/(4\pi\bar{\rho}_{m0})]^{1/3}$ is the comoving Lagrangian radius of the perturbation, which is independent of time. The initial physical radius of the perturbation is thus:

$$r_i = R_L a(t_i) (1 + \delta_i)^{-1/3}. \quad (2.3)$$

Here $a(t_i) = (1 + z_i)^{-1}$ is the expansion factor at the initial time t_i .

The initial velocity, in turn, can be obtained by calculating time derivative \dot{r}_i :

$$\begin{aligned} \dot{r}_i = v_i &= R_L \left[H(t_i) a(t_i) (1 + \delta_i)^{-1/3} - \frac{a(t_i)}{3} \frac{\delta_i}{(1 + \delta_i)^{4/3}} \frac{\dot{D}_+(t_i)}{D_+(t_i)} \right], \\ &= H(t_i) r_i - \frac{r_i}{3} \frac{\delta_i}{1 + \delta_i} \frac{\dot{D}_+(t_i)}{D_+(t_i)} \end{aligned} \quad (2.4)$$

where $H(t_i) = \dot{a}(t_i)/a(t_i) = \sqrt{8\pi G \rho_{\text{crit}}(t_i)/3}$ and we used eq. 2.3 above and the fact that $\delta_0 = \delta_i/D_+(t_i)$ where D_+ is the linear growth factor normalized to unity at $z = 0$.

Given that $\delta_i > -1$ by definition and all of the quantities in the second term of eq. 2.4 are always positive, except for δ_i , the equation shows that underdensities $\delta_i < 0$ expand faster than the Hubble flow expansion velocity $H(t_i)r_i$, while overdensities $\delta_i > 0$ expand slower than the Hubble expansion.

To collapse the overdensity must be *gravitationally bound* - i.e., its total specific energy (i.e., energy per unit mass) must be negative:

$$E = \frac{1}{2}v_i^2 - \frac{GM}{r_i} < 0 \quad \text{or} \quad v_i^2 < \frac{2GM}{r_i}. \quad (2.5)$$

We can re-write the latter condition as

$$v_i^2 < \frac{2GM}{r_i} = \frac{1}{r_i} \frac{8\pi G}{3} \bar{\rho}_{m0} R_L^3 = \frac{8\pi G}{3} \bar{\rho}_{m0} a_i^{-3} r_i^2 (1 + \delta_i) = H^2(t_i) r_i^2 \Omega_m(t_i) (1 + \delta_i), \quad (2.6)$$

where by definition $\Omega(t_i) = \rho_m(t_i)/\rho_{\text{crit}}(t_i)$ and $H^2(t_i) = 8\pi G \rho_{\text{crit}}(t_i)/3$.

Given equation 2.4, a perturbation is bound if $\Omega_m(1 + \delta_i) > 1$. Thus, for the Einstein-de Sitter universe, $\Omega_m(t) = 1$, the perturbation is bound for any $\delta_i > 0$. For cosmologies with $\Omega_m < 1$, the perturbation is bound for $\delta_i > 1/\Omega_m(t_i) - 1$.

Evolution

Evolution of a spherical top-hat perturbation is governed by equation

$$\ddot{r} = -\frac{GM}{r^2}. \quad (2.7)$$

Mass M within the perturbation is constant and the solution of the equation can be written in parametric form as

$$r = \frac{r_{\text{ta}}}{2} (1 - \cos \eta); \quad t = \frac{t_{\text{ta}}}{\pi} (\eta - \sin \eta), \quad (2.8)$$

where r_{ta} is the turnaround radius, the maximum physical radius that will be reached by the edge of the perturbation, t_{ta} is the time of the turnaround and η is the eccentric anomaly parameter which ranges from 0 to 2π during the evolution until the collapse.

The velocity is zero at the turnaround radius and given that total specific energy E is conserved we have $E = -GM/r_{\text{ta}}$ and $r_{\text{ta}} = GM/|E|$. The turnaround time, on the other hand, corresponds to half of the period, which is given by the free fall time of the uniform top-hat sphere (assuming $\Omega_\Lambda = 0$):

$$t_{\text{ta}} = t_{\text{ff}} = \left(\frac{3\pi}{32G\rho} \right)^{1/2} = \left(\frac{\pi^2 r_{\text{ta}}^3}{8GM} \right)^{1/2} = \pi \frac{GM}{(2|E|)^{3/2}}. \quad (2.9)$$

Given that $M = 4\pi(1 + \delta_i)\bar{\rho}_{m0}R_i^3/3$, r_{ta} and t_{ta} depend on the initial comoving perturbation radius, R_i , density contrast δ_i , and total energy $|E|$.

Figure 2.4 shows evolution of radius and density contrast $\delta(< r)$ of the perturbation of $R_i = 1.5h^{-1}$ Mpc, initial density contrast $\delta_i \approx 0.044$, which corresponds to $\approx 1\sigma$ fluctuation of the density field at $z_i = 100$, mass $M \approx 1.75 \times 10^{12} M_\odot$ in the $\Omega_{m0} = 0.3$, $\Omega_\Lambda = 0$ cosmology ($H_0 = 70 \text{ km s}^{-1} \text{ Mpc}$). Initially, $r(t)$ increases as $r(t) \approx a(t)$ as the amplitude of the perturbation is small and it expands almost at the expansion rate of the Universe. However, $r(t)$ steadily decelerates the perturbation radius evolves progressively slower than the overall expansion of the universe. At the turnaround time, t_{ta} , the radius reaches its maximum value, r_{ta} , and subsequently $r(t)$ at $t > t_{ta}$ decreases until the perturbation collapses to $r = 0$. Likewise, density contrast $\delta(< r)$ initially evolves as the linear growth factor, $\delta(< r) \propto D_+(t)$, but before the turnaround, when density contrast reaches values of $\delta \gtrsim 0.1$, its evolution starts to noticeably deviate from the linear prediction. The perturbation evolution enters the *nonlinear regime*. In this regime the evolution of $\delta(t)$ is a much stronger function of time and in fact accelerates as time approaches t_{coll} . At $t \approx t_{coll}$, δ has singularity formally.

Virialization, virial density contrast, and linear density contrast at collapse

In reality, the matter always has some angular momentum, which prevent it from going to zero radius. Moreover, even for purely radial motion, in 3D *radial orbit instability* (Polyachenko and Shukhman 1981, MacMillan, Widrow, and Henriksen 2006) prevents collapse of most mass elements to $r = 0$. What happens when radius reaches $r \approx 0$ is rapid relaxation, called “virialization” in which different mass elements quickly interact with each other gravitationally, potential fluctuates violently, and energy acquired during collapse is converted into energy of random motions that stabilize the collapsed perturbation. If we assume that collapsed halo achieves *virial equilibrium* its kinetic energy is related to the potential energy via the virial relation: $2K = -W$ and the total energy is thus $E = W/2$. Given that at the turnaround $K = 0$ as velocity is zero and that the total energy is conserved, we have $E = W_{ta}$. Thus, $GM^2/r_{ta} = GM^2/(2r_f)$ and the final radius of the virialized perturbation is $r_f = r_{ta}/2$. This is true in cosmologies with $\Omega_\Lambda = 0$. In cosmologies with $\Omega_\Lambda > 0$ vacuum energy acts as an anti-gravity force during late stages of evolution and the initial total energy of the perturbation is not conserved. Even in this case, however, the above approximation is accurate during early stages when $\Omega_\Lambda \approx 0$.

The final mean internal density of a collapsed object in a $\Omega_m = 1$ Universe (with negligible relativistic component and $\Omega_\Lambda = 0$) can be estimated by noting that in a $\Omega_\Lambda = 0$ Universe the time interval $t_{coll} - t_{ta} = t_{ta}$ should be equal to the free-fall time of a uniform sphere $t_{ff} = \sqrt{3\pi/(32G\rho_{ta})}$, which means that the mean density of perturbation at turnaround is $\rho_{ta} = 3\pi/(32Gt_{ta}^2)$ and $\rho_{coll} = 8\rho_{ta} = 3\pi/(Gt_{coll}^2)$. These densities can be compared with background mean matter densities at the corresponding times to get mean internal density contrasts: $\Delta = \rho/\bar{\rho}_m$. In the Einstein-de Sitter model ($\Omega_m = 1$, $\Omega_\Lambda = 0$), background density evolves as $\bar{\rho}_m = \bar{\rho}_{m0}a^{-3}$; for $\Omega_m = 1$ $\bar{\rho}_{m0} = 3H_0^2/(8\pi G)$, while $a = (t/t_U)^{2/3}$, where $t_U = 2/(3H_0)$ is the current ($a = 1$) age of universe. These equations combine to give $\bar{\rho}_m = 1/(6\pi Gt^2)$, which means that density contrast after virialization is

$$\Delta_{vir} \equiv \frac{\rho_{coll}}{\bar{\rho}_m} = 18\pi^2 = 177.653. \quad (2.10)$$

For other cosmologies density contrast can be computed by estimating ρ_{coll} and $\bar{\rho}_m(t_{coll})$ in a similar fashion. For lower Ω_m models, fluctuation of the same mass M and δ has a larger initial radius and smaller physical density and, thus, takes longer to collapse. The density contrasts of collapsed objects therefore are larger in lower density models because the mean density of matter at the time of collapse is smaller. Accurate (to $< 1\%$ for $\Omega_m = 0.1 - 1$) approximations for Δ_{vir} in open ($\Omega_\Lambda = 0$) and flat Λ CDM ($1 - \Omega_\Lambda - \Omega_m = 0$) cosmologies are given by eq. 6 in Bryan and Norman (1998). For example, for the concordance Λ CDM cosmology with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ (Komatsu et al. 2011, Planck Collaboration et al. 2015), density contrast at $z = 0$ is $\Delta_{vir} \approx 358$.

Note that if the initial density contrast δ_i would grow only at the linear rate, $D_+(z)$, then the density contrast at the time of collapse would be more than a hundred times smaller. Its value can be derived starting from the density contrast linearly extrapolated to the turn around epoch, δ_{ta} . This epoch corresponds to the

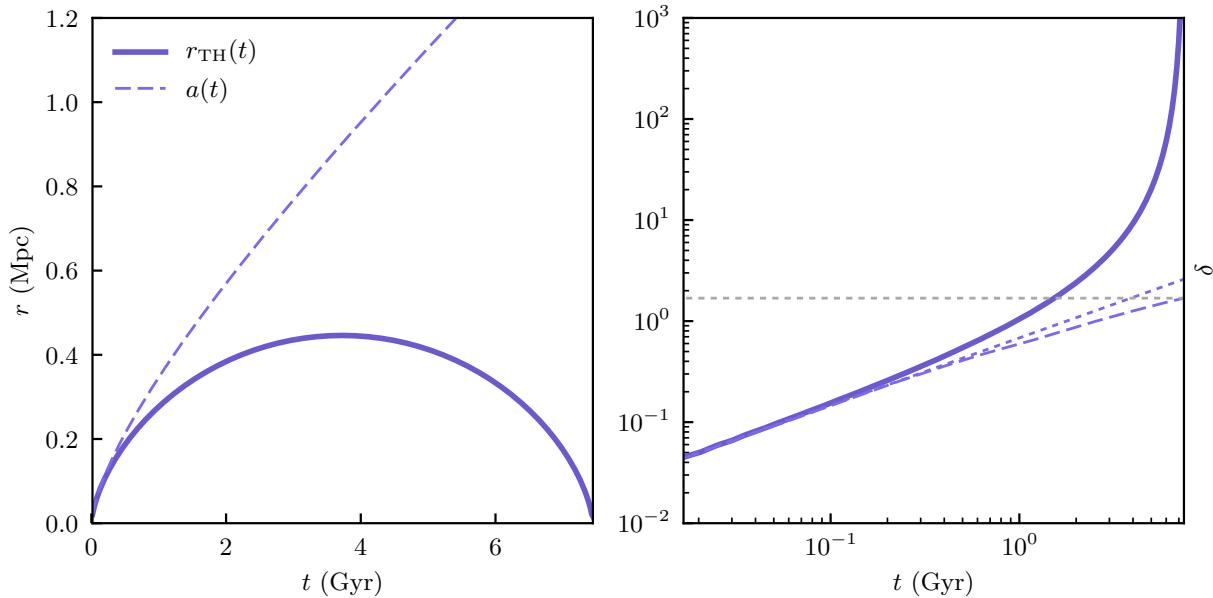


Figure 2.4: Evolution of the physical radius (left) and density contrast $\delta(< r)$ (right) as a function of time for a perturbation with comoving initial radius $R_i = 1.5 h^{-1}$ Mpc, mass $M \approx 1.75 \times 10^{12} M_\odot$ in the $\Omega_{m0} = 0.3$, $\Omega_\Lambda = 0$ cosmology ($H_0 = 70 \text{ km s}^{-1} \text{ Mpc}$). In the left panel the evolution of the top-hat perturbation size, $r_{\text{TH}}(t)$ (solid line), is compared to the evolution of the scale factor $R_i a(t)$ (dashed line). In the right panel the evolution of the density contrast $\delta(< r)$ is compared to the linear evolution $\delta_i D_+(t)/D_+(t_i)$ (dashed line) correct for this cosmology and to $\delta_i(t/t_i)^{2/3}$ expected linear growth of the $\Omega_m = 1$ cosmology (dark dotted line). The horizontal gray dotted line shows $\delta_c = 1.69$ – one can see that $\delta_i D_+(t)/D_+(t_i) = 1.69$ at the time of collapse ($r_{\text{TH}} = 0$).

time at which perturbation enters in the non-linear regime and detaches from the Hubble expansion, so that $\delta_{\text{ta}} \sim 1$ is expected. Because $t_{\text{coll}} = 2t_{\text{ta}}$, further linear evolution for $\Omega_m(z) = 1$ until the collapse time gives $\delta_c = \delta_{\text{ta}} D_+(t_c)/D_+(t_{\text{ta}}) = a(t_c)/a(t_{\text{ta}}) = (t_c/t_{\text{ta}})^{2/3} = 2^{2/3} \approx 1.6$

In fact, the exact calculation of the evolution equations in the case of $\Omega_m(z) = 1$ at the redshift of turnaround gives $\delta_{\text{ta}} = 1.062$ (Gunn and Gott 1972), which gives $\delta_c = \delta_{\text{ta}} 2^{2/3} \approx 1.686$. In the case of $\Omega_m \neq 1$ we expect that δ_{ta} should have different values. For instance, for $\Omega_m < 1$ density contrast at turn-around should be higher to account for the higher rate of the Hubble expansion. However, linear growth from t_{ta} to t_{coll} is smaller due to the slower redshift dependence of $D_+(z)$. As a matter of fact, these two factors nearly cancel, so that δ_c has a weak dependence on Ω_m and Ω_Λ (e.g., Percival 2005). For the concordance Λ CDM cosmology at $z = 0$, for example, $\delta_c \approx 1.675$. The exact dependence of δ_c on cosmological parameters is thus often neglected, but one should be aware that in some cases this may matter (see, e.g., Courtin et al. 2011).

2.3.2 Nonlinear mass and peak height

The linear value of the collapse overdensity δ_c is useful in predicting whether a given initial perturbation $\delta_i \ll 1$ at initial z_i collapses by some later redshift z . The collapse condition is simply $\delta_i D_{+0}(z) \geq \delta_c(z)$ and is used extensively to model the abundance and clustering of collapsed halos, as we discuss below. As discussed in the beginning of this chapter, the distribution of peak amplitudes in the initial Gaussian overdensity field smoothed over mass scale M is given by the Gaussian pdf with a rms value of $\sigma(M)$, computed by convolving the power spectrum with a filter window of size $R = [3M/(4\pi\bar{\rho})]^{1/3}$ (see eq. B.8 in §B.2).

The peaks in the initial Gaussian overdensity field smoothed at redshift z_i over mass scale M can be characterized by the ratio $v = \delta_i/\sigma(M, z_i)$, where $\sigma(M, z_i) = \sigma(M, z=0)D_+(z)/D_+(z=0)$, called the *peak height*. Peak height is analogous to signal-to-noise in experiments: it quantifies by how much the peak stands

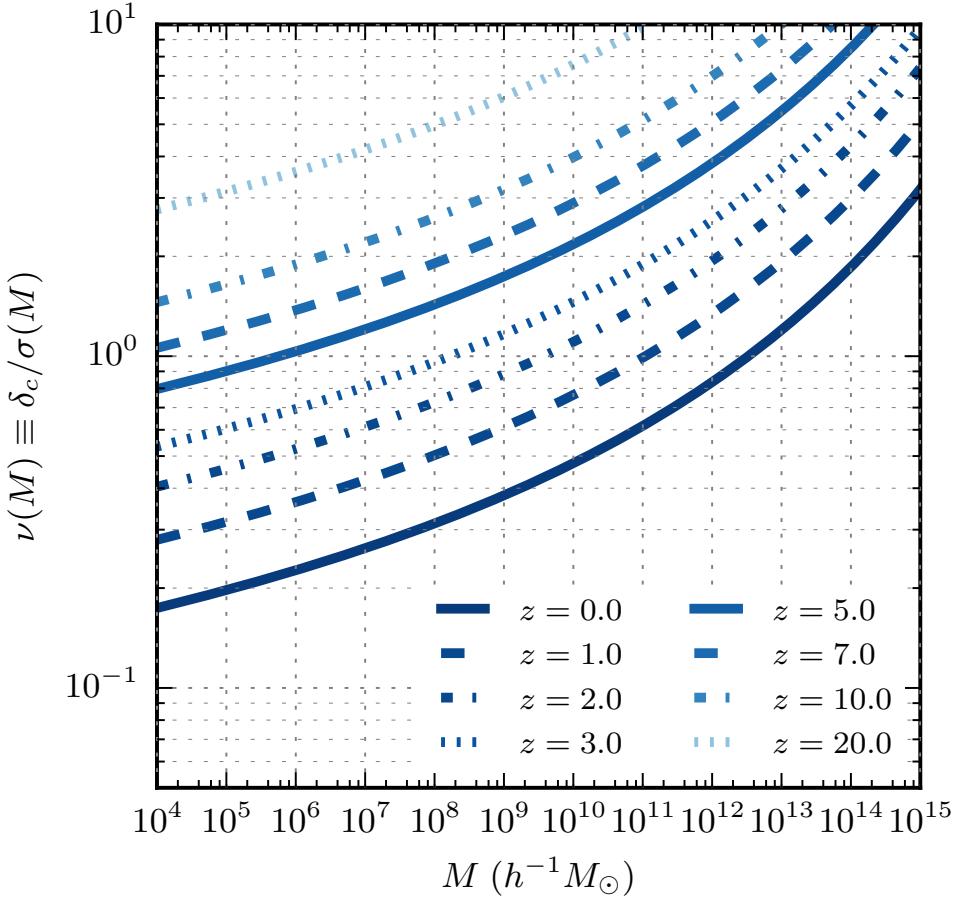


Figure 2.5: Peak-height, $v \equiv \delta_c/\sigma(M, z)$, as a function of M for different redshifts. The assumed cosmology is flat Λ CDM with $H_0 = 70$ km/s/Mpc, $\Omega_{\text{m},0} = 0.27$, $\Omega_{\text{b},0} = 0.0457$, $\sigma_8 = 0.82$, $n_s = 0.96$.

out compared to typical “noise” fluctuation of amplitude σ . For a Gaussian field it also specifies how rare such peaks are because abundance of peaks is suppressed exponentially as $e^{-v^2/2}$ (see eq. B.11).

For a given mass scale M , the peaks collapsing at a given redshift z according to the spherical collapse model have the peak height given by:

$$v \equiv \frac{\delta_c(z)}{\sigma(M, z)}. \quad (2.11)$$

Given that $\delta_c(z)$ is a very weak function of z (changing by $\lesssim 1 - 2\%$ typically), whereas $\sigma(M, z) = \sigma(M, z=0)D_{+0}(z)$ decreases strongly with increasing z , the peak height of collapsing objects of a given mass M increases rapidly with increasing redshift, as can be seen in Figure 2.5, which shows $v(z, M)$ for the concordance Λ CDM cosmology.

Using equation for v we can define the characteristic mass scale for which a typical peak ($v = 1$) collapses at redshift z :

$$\sigma(M_{\text{nl}}, z) = \sigma(M_{\text{nl}}, z=0)D_{+0}(z) = \delta_c(z). \quad (2.12)$$

The nonlinear mass M_{nl} can be obtained from this equation by solving it numerically. In Figure 2.5 the nonlinear mass at different redshifts can be read off the plot, as the mass at which the horizontal $v = 1$ line intersects the $v(M, z)$ line for corresponding z .

2.3.3 The ellipsoidal collapse model

A number of studies have considered collapse of aspherical peaks (e.g., Monaco 1995, Eisenstein and Loeb 1995, Bond and Myers 1996, Audit, Teyssier, and Alimi 1997, Lee and Shandarin 1998, Sheth and Tormen 2002).

The simplest model of this kind is the model, in which peaks are approximated as ellipsoids of constant density (Monaco 1995, Eisenstein and Loeb 1995, Bond and Myers 1996). Elliptical peaks are elongated along major axis and take longer to collapse due to stronger torques from surrounding halos. This means that lower- v , more elliptical halos have higher density threshold to collapse by a given epoch z than the more spherical high- v peaks.

This means that threshold for collapse is no longer constant, but depends on the ellipticity of peaks. The average ellipticity changes from larger for low v to low for large v . This is because high v peaks result from constructive superposition of many Fourier modes, which are likely to give a more spherical peak compared to low v peaks formed by superposition of a few modes. Calibration of the collapse threshold dependence on peak height using cosmological simulations gives:

$$\delta_{\text{ec}} = \delta_c (1 + \beta v^{-2\gamma}), \quad (2.13)$$

where $\delta_c \approx 1.69$ is the top-hat collapse threshold, $\beta \approx 0.41$ and $\gamma \approx 0.81$ (see Table 1 and Figure 3 of Robertson et al. 2009). The formula above shows that δ_{ec} approaches δ_c for large v , while for $v = 1$, the mean threshold for collapse is quite a bit larger: $\delta_{\text{ec}} \approx 2.4$.

2.3.4 Mass function of halos

Contrasting predictions for the abundance and clustering of collapsed objects with the observed abundance and clustering of galaxies, groups, and clusters has been among the most powerful validation tests of structure formation models (e.g., Press and Schechter 1974, Blumenthal et al. 1984, Kaiser 1984, Kaiser 1986).

The abundance of collapsed halos is quantified as a function of their mass, i.e., in terms of the mass function $dn(M, z)$ defined as the comoving volume number density of halos in the mass interval $[M, M + dM]$ at a given redshift z . Below I review theoretical models for halo abundance and underlying reasons for its approximate universality.

The Press-Schechter ansatz

The first statistical model for the abundance of collapsed objects as a function of their mass was developed by (Press and Schechter 1974). The main powerful principle underlying this model is that the mass function of objects resulting from nonlinear collapse can be tied directly and uniquely to the statistical properties of the initial linear density contrast field $\delta(\mathbf{x})$.

Statistically, one can define the probability $F(M)$ that a given region within the initial overdensity field smoothed on a mass scale M , $\delta_M(\mathbf{x})$, will collapse into a halo of mass M or larger:

$$F(M, a) = \int_{-1}^{\infty} p(\delta) C_{\text{coll}}(\delta, a) d\delta, \quad (2.14)$$

where $p(\delta)d\delta$ is the PDF of $\delta_M(\mathbf{x})$ for the Gaussian initial density field smoothed on mass scale M (see §B.2), and C_{coll} is the probability that any given point \mathbf{x} with local overdensity $\delta_M(\mathbf{x})$ will actually collapse by a given epoch a . The mass function can then be derived as a fraction of the total volume collapsing into halos of mass $(M, M + dM)$, i.e., dF/dM , divided by the comoving volume within the initial density field occupied by each such halo, i.e., $M/\bar{\rho}$:

$$\frac{dn(M)}{dM} = \frac{\bar{\rho}_m}{M} \left| \frac{dF}{dM} \right|. \quad (2.15)$$

In their pioneering model, Press and Schechter (1974) have adopted the ansatz motivated by the spherical top-hat collapse model that any point in space with $\delta_M(\mathbf{x})D_{+0}(z) \geq \delta_c$ will collapse into a halo of mass $\geq M$ by

redshift z : i.e., $C_{\text{coll}}(\delta) = \Theta(\delta - \delta_c)$, where Θ is the Heaviside step function. Note that $\delta_M(x)$ used above is not the actual initial overdensity, but the initial overdensity evolved to $z = 0$ with the linear growth rate. One can easily check that for a Gaussian initial density field this assumption gives $F(M) = \frac{1}{2}\text{erfc}[\delta_c/(\sqrt{2}\sigma(M, z))] = F(v)$. This line of arguments and assumptions thus leads to an important conclusion that *the abundance of halos of mass M at redshift z is a universal function of only their peak height $v(M, z) \equiv \delta_c/\sigma(M, z)$* . In particular, the fraction of mass in halos per logarithmic interval of mass in such a model is:

$$\frac{dn(M)}{d\ln M} = \frac{\bar{\rho}_m}{M} \left| \frac{dF}{d\ln M} \right| = \frac{\bar{\rho}_m}{M} \left| \frac{d\ln v}{d\ln M} \frac{\partial F}{\partial \ln v} \right| \equiv \frac{\bar{\rho}_m}{M} \left| \frac{d\ln v}{d\ln M} \right| g(v) \equiv \frac{\bar{\rho}_m}{M} \psi(v). \quad (2.16)$$

Clearly, the shape $\psi(v)$ in such models is set by the assumptions of the collapse model. In the Press-Schechter ansatz the mass function of halos has the form

$$\frac{dn_{\text{PS}}}{d\ln M} = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}}{M} v \frac{d\ln v}{d\ln M} \exp\left(-\frac{v^2}{2}\right). \quad (2.17)$$

Note that this function has units of inverse *comoving* volume.

It is instructive to consider prediction of this ansatz for the self-similar model with $\Omega_m = 1$ and power law power spectrum, $P(k) \propto k^n$ or $\sigma^2(R) \sim k^3 P(k)|_{k \sim 1/R} \propto R^{-(3+n)}$. Fluctuations in the matter-dominated regime in $\Omega_m = 1$ cosmology grow as $\delta \propto a(t)$ and thus $\sigma \propto a(t)R^{-(3+n)/2} \propto a(t)M^{-(3+n)/6}$ or $v = \delta_c/\sigma \propto a^{-1}M^{(3+n)/6}$, where $M \propto (1 + \delta)R^3$ is mass of the perturbation, where δ can be neglected in the linear regime $\delta \ll 1$. Thus, for $v \ll 1$, $\exp(-v^2/2) \approx 1$ and $dn/d\ln M \propto (3+n)/(6a)M^{(3+n)/6-1}$. At small masses n is close to $n \approx -3$ and thus $dn/d\ln M \propto M^{-\alpha}$ with $\alpha \approx -1$ (albeit somewhat smaller than -1, so that mass converges logarithmically).

At $v \gtrsim 1$ the abundance of halos is suppressed exponentially. The exponential suppression shifts to smaller masses and is less sharp at higher z because $v = 1$ shifts to smaller mass halos, where v is a very shallow function of mass because $n \approx -3$ (see also Figure 2.5 above). The apparent lack of evolution of $dn/d\ln M$ at $M \sim 10^6 - 10^6 M_\odot$ in Figure 2.6 is due to a near cancellation of the increase due to a^{-1} factor and the increasing of exponential term at fixed M (i.e., increasing $v(M)$) with increasing redshift. Physically, this behavior reflects the nearly self-similar behavior of peak collapse and merging in this nearly power law spectrum regime: evolution of halos to larger masses due to mass accretion and mergers is compensated almost exactly by the corresponding growth of halos of smaller masses, so that comoving number density of small mass halos stays almost the same.

Mass function in the ellipsoidal collapse model

Numerical studies (e.g., Gross et al. 1998) based on cosmological simulations have eventually revealed that the shape $\psi_{\text{PS}}(v)$ predicted by the PS ansatz deviates by $\gtrsim 50\%$ from the actual shape measured in cosmological simulations.

A number of modifications to the original ansatz have been proposed, which result in $\psi(v)$ that more accurately describes simulation results. Such modifications are based on the collapse conditions that take into account asphericity of the peaks in the initial density field discussed above in §2.3.3, and stochasticity due to the dependence of the collapse condition on peak properties other than v or shape.

The most commonly used modification is that of Sheth, Mo, and Tormen (2001), hereafter SMT, in which it is taken into account that peaks in the Gaussian field are non-spherical with ellipticity decreasing with increasing peak height (Doroshkevich 1970). Elliptical peaks are elongated along major axis and take longer to collapse due to stronger torques from surrounding halos. This means that lower- v , more elliptical halos have higher density threshold to collapse by a given epoch z than the more spherical high- v peaks. The approximate threshold for collapse derived by SMT is:

$$\delta_{\text{ec}} = \delta_{\text{sc}} (1 + \beta v^{-\gamma})$$

with $\beta = 0.47$ and $\gamma = 0.615$. This collapse threshold results in the following mass function

$$\frac{dn_{\text{SMT}}}{d\ln M} = A \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}}{M} \left(1 + \frac{1}{(av)^{2q}} \right) av \frac{d\ln v}{d\ln M} \exp\left(-\frac{(av)^2}{2}\right), \quad (2.18)$$

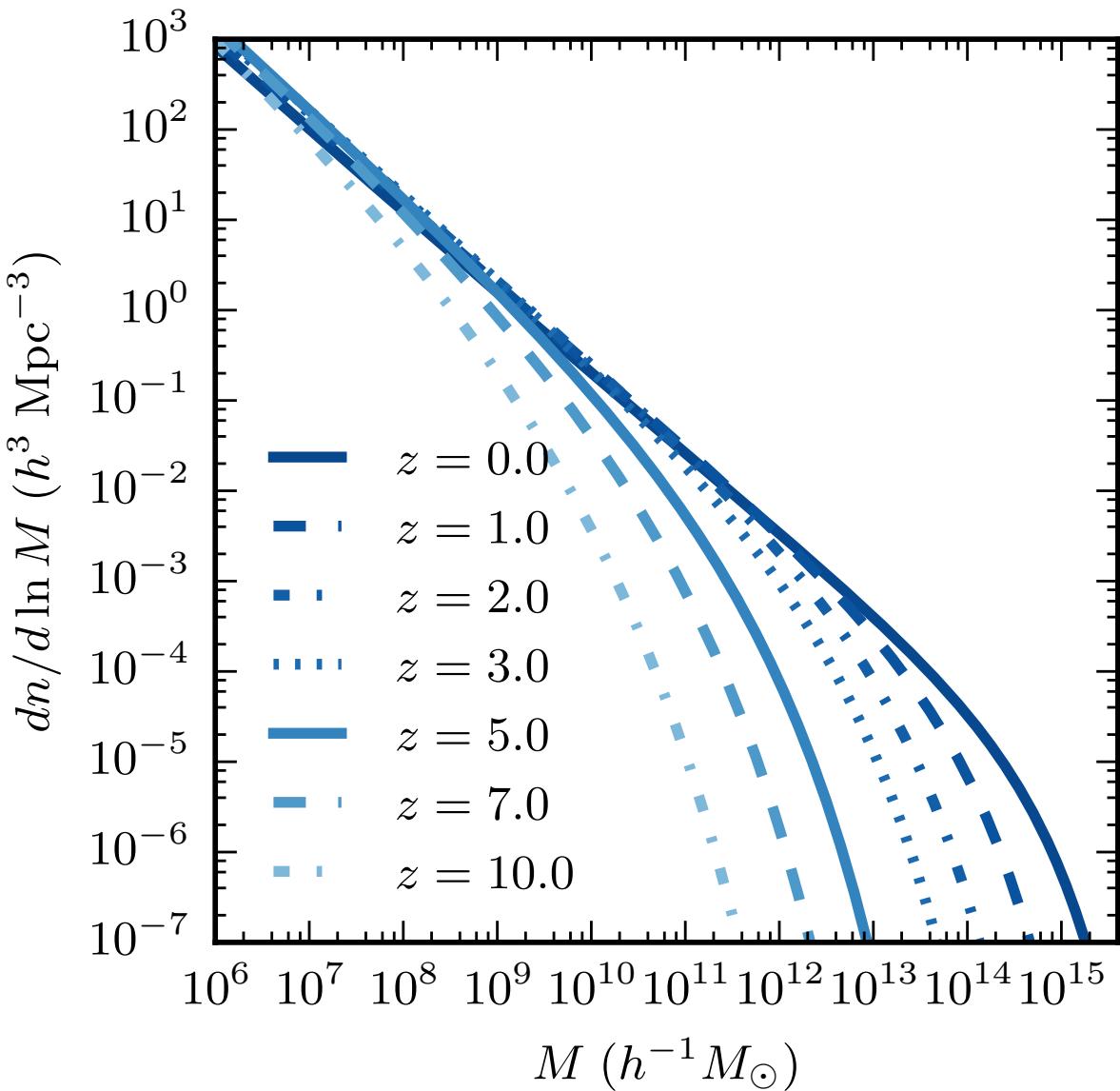


Figure 2.6: Evolution of halo mass function in flat Λ CDM cosmology ($\Omega_m = 1 - \Omega_\Lambda = 0.27$, $h = 0.7$, $\sigma_8 = 0.82$, $n_s = 0.96$). The mass functions are computed using the approximation based on the ellipsoidal collapse and given by eq. 2.18 with $A = 0.3222$, $a = 0.707$, and $q = 0.3$.

with $A = 0.3222$, $a = 0.707$, and $q = 0.3$ (Sheth, Mo, and Tormen 2001). Figure 2.6 shows mass function in this approximation at different z .

2.3.5 Collapse of peaks of non-uniform density

How well does the top-hat profile approximate real peaks in a Gaussian field with CDM power spectrum? The statistics of peaks in Gaussian random fields is well characterized (e.g., Bardeen et al. 1986) and can actually be computed fairly straightforwardly for a given Gaussian field with a known power spectrum. However, the result depends on how the profiles are centered and whether or not the Gaussian field is smoothed on some scale.

Figure 2.7 shows the density profiles of peaks in a Gaussian field computed in different ways from the study of Dalal, Lithwick, and Kuhlen (2010). The dotted line shows the average profile of peaks in a Gaussian field smoothed on scale $R_{\text{TH}} = [3M/(4\pi\bar{\rho})]^{1/3}$, where $M \approx 2 \times 10^{12} M_\odot$ (i.e., mass close to the mass of Milky Way halo). This is the profile given by the mean of the conditional pdf $p(\delta(< r)|\delta_{\text{TH}}, \delta'_{\text{TH}})$,

in which probability of a given δ is conditioned by the value of δ_{TH} for $r = R_{\text{TH}}$ and by the derivative $\delta'_{\text{TH}} = d\delta/dr|_{r=R_{\text{TH}}}$. The first condition is to ensure that we have a peak that will actually collapse by a given epoch. The derivative condition is needed to ensure that we have an isolated peak. For example, $\delta' > 0$ would indicate that the peak is local and is a part of a larger peak that would collapse at an earlier epoch. The conditional probability of a Gaussian variable is also Gaussian, but with modified mean and variance (see §B.3), which can be computed for a given power spectrum.

The resulting profile is quite flat (shallow) near the peak center, but as R approaches R_{TH} the density decreases sharply with increasing radius. Such profile can be fairly reasonably approximated by the top-hat profile: i.e., $\delta = \text{const}$ for $r < R_{\text{TH}}$ and $\delta \leq 0$ for $r > R_{\text{TH}}$. The main difference from the top-hat collapse is that collapse of such peak would be somewhat extended in time, due to curvature of the actual profile compared to the instantaneous collapse in the top-hat model.²

However, this is not quite the right profile to consider when we want to understand evolution of a halo. Gaussian fields with a CDM power spectrum have power on all scales relevant for galaxy formation. The peaks smaller than R_{TH} have been smoothed in the field used to construct the average profile shown by dotted line, but they are actually present in the density field. Their density is higher than the mean density of the smoothed peak and they would thus collapse first, well before the collapse of the main peak happens. This can be taken into account by constructing a profile that takes into account existence of sub-peaks within a given Gaussian peak using supremum statistics (see §4 in Dalal, Lithwick, and Kuhlen 2010). If x is a Gaussian random variable with pdf $p_G(x)$ of zero mean and unit variance and we have N samples $\{x_i\}$, the probability that a given y is larger than any one x_i is

$$P_1(y) = \int_{-\infty}^y \frac{dP_G}{dx} dx = 1 - \frac{1}{2} \text{erfc}\left(\frac{y}{\sqrt{2}}\right), \quad (2.19)$$

where $\text{erfc}(x)$ is the complementary error function.

If we assume that x_i samples are uncorrelated (which is not quite correct, so this is an approximation), the probability for y to exceed all N samples is:

$$P_N(y) = [P_1(y)]^N. \quad (2.20)$$

For a given peak of mass M and scale $R_{\text{TH}} = [3M/(4\pi\bar{\rho})]^{1/3}$, when we consider scale r , we can think of $N \approx (R_{\text{TH}}/r)^3$ sub-peaks fitting within it. Thus, at a given radius r from the peak center the probability to have a given $\bar{\delta}$ that is larger than central density of any Gaussian sub-peaks within r shell is given by the above equation, in which $p(x) = p_G(\bar{\delta}(< r/r_{\text{TH}}), |\delta_{\text{TH}}, \delta'_{\text{TH}})$ is the Gaussian field with the mean density of the smoothed peak conditioned to have a given overdensity and its slope at $r = R_{\text{TH}}$ (i.e. δ of the dotted line in Figure 2.7). Profile computed in this way is shown by the blue solid line in Figure 2.7 along with the profile of the actual Λ CDM halo of $M \approx 2 \times 10^{12} M_\odot$, simulated in the “Via Lactea 2” simulation, centered on the most massive progenitor at $z \approx 17$ shown by the black solid line (see Dalal, Lithwick, and Kuhlen 2010).

We can see that the profile computed from supremum statistics matches the actual profile around progenitor of Λ CDM halo. This profile is considerably steeper than the profile of a peak smoothed on scale R_{TH} and is quite different from the uniform density of the top-hat profile.

Thus, even in spherical symmetry, a more realistic model of peak collapse must consider collapse of spherical shells of different $\delta(< r)$, decreasing with increasing shell radius r . The models of this kind have been developed and their main features are discussed in below. The main difference from the top-hat collapse model is that different radial shells collapse at different times due to varying density. This is the reason why collapse of real halos is extended in time, in contrast with the instantaneous collapse assumed in the top-hat collapse model.

²Note that in the top-hat collapse model the virialization process cannot be instant, even if collapse is instantaneous. Relaxation processes establishing the Jeans and virial equilibrium operate on $\sim 1 - 3$ dynamical time scales. The dynamical time scale for halos is ~ 1 Gyr and thus establishing equilibrium will take a substantial fraction of the age of the universe. The assumption of instantaneous virialization is thus a significant idealization.

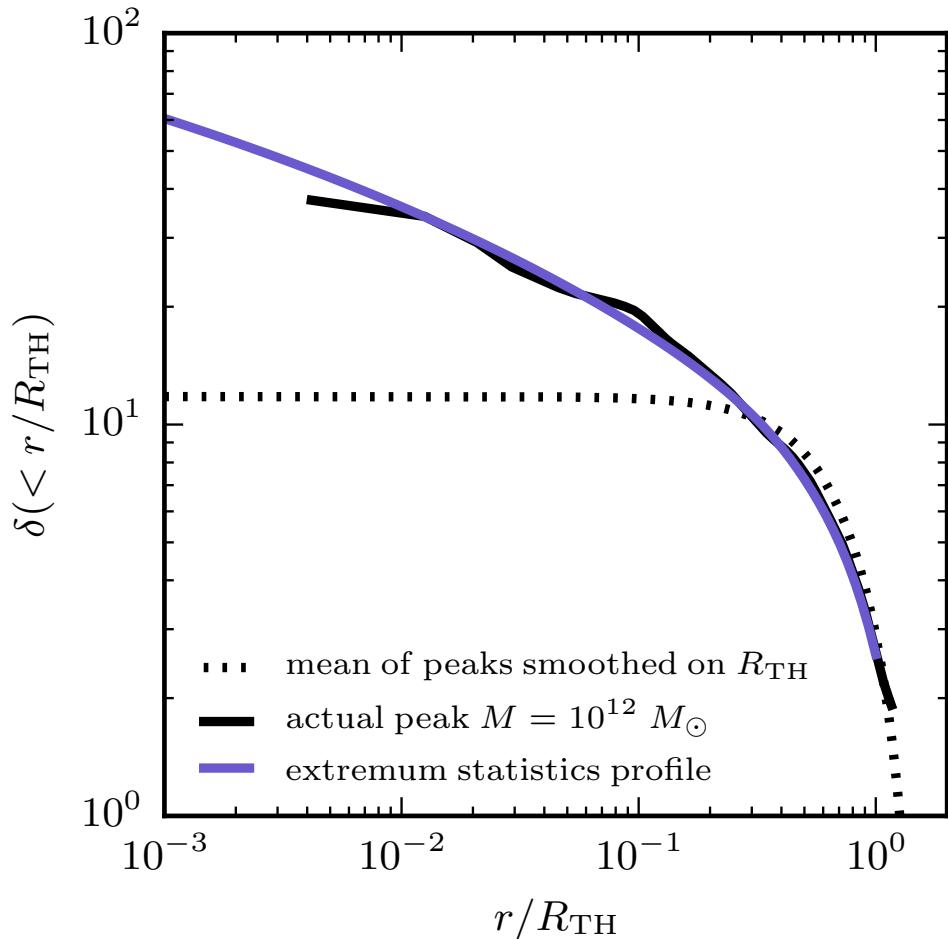


Figure 2.7: Enclosed overdensity profiles of peaks in the linear Gaussian density field linearly extrapolated to $z = 0$: $\delta(< r) = \delta_i(< r)D_+(z=0)/D_+(z_i)$, where $D_+(z_i)$ is the linear growth factor. The dotted line shows that mean profiles of peaks in the Gaussian field with Λ CDM smoothed on scale $R_{\text{TH}} = [3M/(4\pi\bar{\rho})]^{1/3}$, where $M \approx 10^{12} M_\odot$ with the top-hat filter. The blue solid line shows the density profile computed using supremum statistics, as described in the text. This profile takes into account existence of smaller peaks within the main peak. The solid black line shows the profile of the actual $M \approx 2 \times 10^{12}$ halo formed in the Via Lactea 2 Λ CDM simulation centered on its main progenitor at $z = 17$. The profiles are from Dalal, Lithwick, and Kuhlen (2010) and were kindly provided by Neal Dalal.

Collapse of real peaks, therefore, is characterized by a *mass accretion history* (MAH), rather than by a single collapse redshift z_c . In the next section, we will consider MAHs resulting from a given density peak profile and as we will see the results of the top-hat collapse calculation are quite useful in determining collapse time of individual shells.

2.3.6 Mass accretion histories of halos

Let us represent an arbitrary density profile of initial density peak by power law *locally*:

$$\delta_i(< r) = (M_i/M_*)^{-\varepsilon} \propto r^{-3\varepsilon}, \quad (2.21)$$

where logarithmic slope $3\varepsilon(r) = d \ln \delta / d \ln r$ is a function of r , $M_i(< r_i)$ is the mass perturbation corresponding to the density perturbation and M_* is just some specific physical mass scale (say, mass $M = M(< R_{\text{TH}})$ of the peak). In the following I will assume that δ_i has amplitude linearly extrapolated to $z = 0$, so that its

amplitude at any earlier redshift $z > 0$ is $\delta = \delta_i D_+(z)$, where $D_+(z)$ is the linear growth factor normalized to unity at $z = 0$.

Collapse of a given shell r_i can be assumed to occur when its linearly extrapolated overdensity reaches the value $\delta_c \approx 1.68$ derived above in the context of the top-hat collapse model: $\delta_i(r_i)D_+(z) = \delta_c$ or $M(< r) = M_*(D_+(t)/\delta_c)^{1/\varepsilon}$. Taking time derivative of the mass profile gives

$$\frac{dM}{dt} = \frac{1}{\varepsilon} \frac{M_*}{\delta_c} \dot{D}_+ \left(\frac{D_+}{\delta_c} \right)^{1/\varepsilon-1}, \quad (2.22)$$

and

$$\frac{d \ln M}{d \ln t} = \frac{t}{M} \frac{dM}{dt} = \frac{t}{\varepsilon} \frac{\dot{D}_+}{D_+}. \quad (2.23)$$

For $\Omega_m = 1$ cosmology in the matter-dominated regime: $D_+(t) = a(t) = (t/t_U)^{2/3}$, where t_U is the age of the universe, so that $\dot{D}_+/D_+ = 2/(3t)$, so the above equation gives $d \ln M / d \ln t = 2/(3\varepsilon)$. This explicitly shows that the rate of mass accretion at a given epoch is directly related to the slope of the initial density perturbation. Shallower slopes, $\varepsilon \ll 1$, in the inner radii of a density peak result in large accretion rate, while steep slopes of the outer profile at $r \sim R_{TH}$ result in slower accretion rates.

Figure 2.8 shows comparison of prediction by eq. 2.23 and $\varepsilon(r)$ given by the slope of the density profile shown by the blue line in Figure 2.7 to the results of fits to the actual MAHs obtained for halos of $M \approx 2 \times 10^{12} M_\odot$ formed in cosmological simulations (Diemand et al. 2008). These approximations are from Krumholz and Dekel (2012) (see their §2.1.1):

$$\left\langle \frac{dM}{dt} \right\rangle = -6.28 \times 10^{11} M_\odot \left(\frac{M}{10^{12} M_\odot} \right)^{1.14} \frac{d}{dt} \left(\frac{\delta_c}{D_+(t)} \right), \quad (2.24)$$

and Fakhouri, Ma, and Boylan-Kolchin (2010):

$$\left\langle \frac{dM}{dt} \right\rangle = 46.1 M_\odot \text{yr}^{-1} \left(\frac{M}{10^{12} M_\odot} \right)^{1.1} \times (1 + 1.11z) \sqrt{\Omega_{m,0}(1+z)^3 + \Omega_\Lambda}. \quad (2.25)$$

We can see that prediction from the simple peak collapse model matches results of cosmological simulations quite well. Although it predicts evolution of mass that is faster than the simulation fits at early times (i.e., $d \ln M / d \ln t$ somewhat steeper), there are actually reasons to think that peak collapse prediction is more correct in this regime. This is because the mass evolution on this mass scale is not probed reliably in simulations at these high redshifts because number of particles in halos is small and the halo masses can be easily overestimated. Also, the MAH from the peak collapse was constrained to have a shape that matched a peak of a particular halo in the simulation, while there is a range of MAHs for the same mass. A recent calibration of MAH in simulation using parametrization based on extended Press-Schechter modelling was presented by Correa et al. (2015):

$$\left\langle \frac{dM}{dt} \right\rangle = 71.6 M_\odot \text{yr}^{-1} \left(\frac{M}{10^{12} M_\odot} \right) \left(\frac{h}{0.7} \right) (0.75(1+z) - 0.24) \sqrt{\Omega_{m,0}(1+z)^3 + \Omega_\Lambda}. \quad (2.26)$$

This calibration is also shown in Figure 2.8 and is steeper than the previous calibrations and the peak collapse prediction.

Overall, the agreement between models is quite reasonable at $t \gtrsim 2$ Gyrs. This is not surprising because this is the range of epochs where they actually were calibrated at. At higher z , these models are extrapolations. The performance of the peak collapse model is remarkable, given how simple it is compared to running full cosmological simulations.

Thus, the fact that peak collapse model predicts more or less correct evolution of mass means that the *MAHs of halos measured in simulations reflect the initial overdensity profiles of the peaks from which they form*.

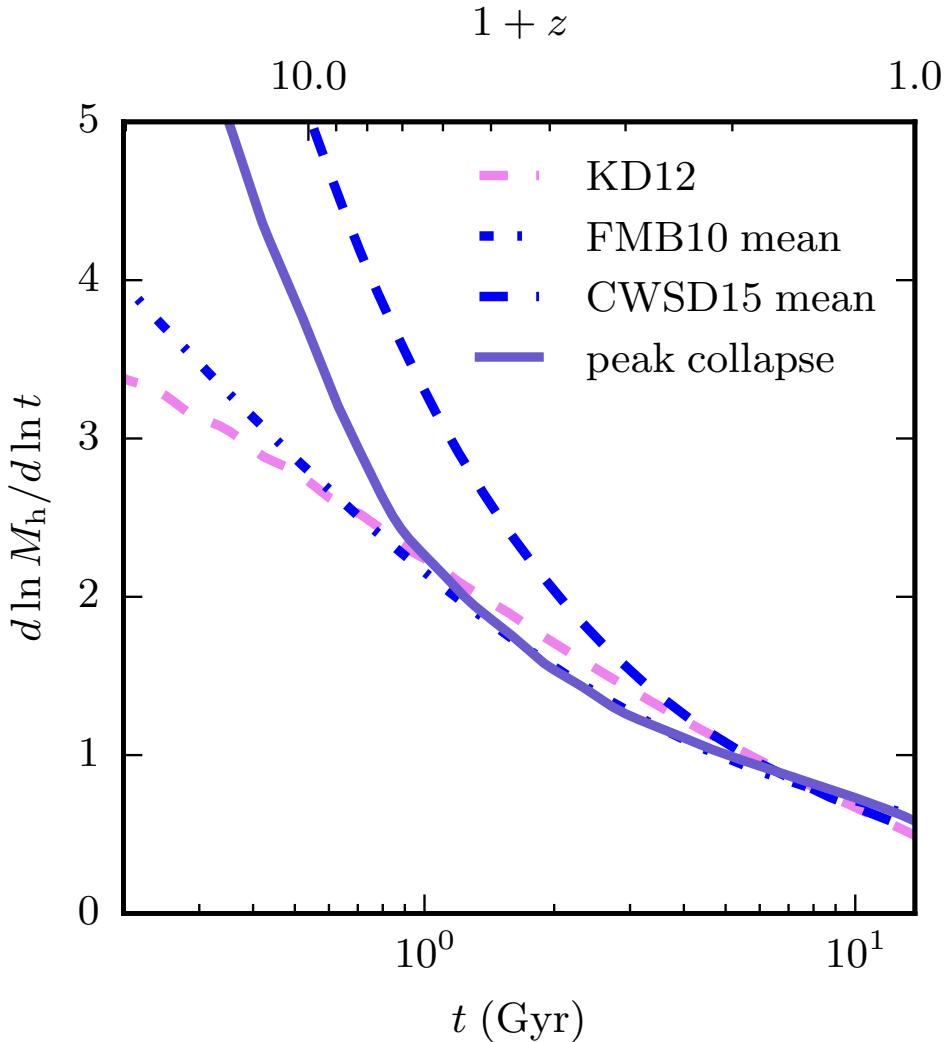


Figure 2.8: Logarithmic slope of the mass accretion history, $d \ln M / d \ln t$, as a function of time in the peak collapse model (solid line) for peak collapsing into $M \approx 2 \times 10^{12} M_\odot$ halo compared to fits of MAHs of halos of similar mass in cosmological simulations of Λ CDM cosmology. We see that the peak collapse model reproduces the shape of MAHs observed in cosmological simulations quite well. This implies that the MAHs simply reflect the shape of initial density peaks in the Gaussian field from which halos collapse.

2.3.7 Characteristic “formation” time of halos

Although the mass accretion history of halos is extended, it is often useful to define a characteristic epoch that characterizes a given MAH. Different choices are used in the literature. A frequent choice is to define formation epoch as the epoch when main progenitor of a halo had mass equal to half (or some smaller fraction) of the present mass. This definition is reasonable, but in practice individual MAHs have jumps associated with merger events and $z_{1/2}$ defined in this way may be quite noisy due to such jumps.

Another choice is to approximate MAHs by a smooth function and measure characteristic time or redshift from such function. For example, the mass accretion histories can be approximated by the exponential function: $M(z) = M_0 \exp(-\alpha z)$ (Wechsler et al. 2002), where α is the quantity related to the formation redshift, in whatever way one chooses to define it, and M_0 is mass at $z = 0$. One possible choice is to define formation epoch as the redshift or expansion factor at which logarithmic slope of the mass accretion rate is

equal to 2: $\Gamma = d \log M / d \log a = 2$: $a_f = \alpha a_0 / 2$. Then we can write the mass accretion history as:

$$M(a) = M_0 \exp [-2a_f(a_0/a - 1)]. \quad (2.27)$$

In practice, the MAHs of halos are better described by the exponential multiplied by power law $(1+z)^\beta$ (Tasitsiomi et al. 2004, McBride, Fakhouri, and Ma 2009). Regardless of the actual smooth fitting function used, one can still define the characteristic formation using definition like the one above.

2.3.8 Secondary infall model

Model of collapse of peaks with a given density profile was first considered by Gunn and Gott (1972), with follow up in Fillmore and Goldreich (1984) and Bertschinger (1985) and is called the *secondary infall model* because it considers a collapse of a spherical shell onto a pre-existing overdensity. We can think of using the tophat collapse model to figure out collapse of the first innermost shell and then consider collapse of all other shells onto the collapsed first shell in the secondary infall model. Although the full collapse can be solved numerically for arbitrary $\delta(< r)$ profile, analytic models were also developed which use a simplifying assumption of power law $\delta(< r) \propto r^{-\varepsilon}$ profile with $\varepsilon = \text{const}$, which allows one to obtain self-similar solutions of dynamical equations. The useful thing about these models is that they predict not only the mass accretion history, but also density profile around collapsed halo. Let us consider an example of such calculation.

The collapse of each shell follows the same cycloid parametric solution for its $r_s(t)$ as the radius of the tophat overdensity discussed above, but now with enclosed mass and corresponding enclosed overdensity $\delta_i = \Delta M / M$ differing for each shell. Let's consider how properties of turnaround will scale for each shell.

We will assume that the initial enclosed overdensity profile is a power law: $\delta_i \propto r_i^{-\varepsilon}$. At turnaround $\eta = \pi$ the shell's radius is $r = 2A = r_{\text{ta}}$. Given that turnaround occurs when the enclosed overdensity is $\delta_{\text{ta}} = \delta_i a_{\text{ta}} / a_i \approx 1$, $a_{\text{ta}} \approx \delta_i^{-1} \propto r_i^\varepsilon$. Until shell cross after collapse, the enclosed mass M is constant. At turnaround

$$\frac{3M}{4\pi\bar{\rho}(a_{\text{ta}})r_{\text{ta}}^3} = \delta_{\text{ta}} + 1 \approx 2. \quad (2.28)$$

Thus, using a similar formula at a_i we get:

$$r_{\text{ta}}^3 \approx \frac{1}{2}(1 + \delta_i) \left(\frac{a_{\text{ta}}}{a_i} \right)^3 r_i^3 \quad (2.29)$$

or assuming $\delta_i \ll 1$:

$$r_{\text{ta}} \approx \frac{1}{2^{1/3}} \frac{a_{\text{ta}}}{a_i} r_i \propto r_i^{1+\varepsilon}. \quad (2.30)$$

From the parametric solution $r_i a = 2A \propto r_i^{1+\varepsilon}$. Then the constant $B = (A^3/GM)^{1/2} \propto r_i^{3(\varepsilon+1)/2-3/2} \propto r_i^{3\varepsilon/2}$, where I used $M \propto (1 + \delta_i)r_i^3 \propto r_i^3$. Now we have a way to rank-order constants corresponding to different shells using their Lagrangian (i.e., initial) radius r_i .

We can then write solutions for radius and time for each shell in units of turnaround radius and time of the shell that is just at turnaround at time $t = t_{\text{ta}}$ (it has label $A \propto r_i^{1+\varepsilon}$) as:

$$\frac{r_s(t)}{r_{\text{ta}}} = \left(\frac{r_i}{r_{i,\text{ta}}} \right)^{1+\varepsilon} \frac{1 - \cos \eta}{2} \quad (2.31)$$

$$\frac{t}{t_{\text{ta}}} = 1 = \left(\frac{r_i}{r_{i,\text{ta}}} \right)^{3\varepsilon/2} \frac{\eta - \sin \eta}{\pi} \quad (2.32)$$

Substituting r_i/r_{ta} from the second equation into the first we get:

$$\frac{r_s(t)}{r_{i,\text{ta}}} = \left(\frac{\pi}{\eta - \sin \eta} \right)^{2(1+\varepsilon)/3\varepsilon} \frac{1 - \cos \eta}{2} \quad (2.33)$$

Now we have a way to connect r_i , η , and r , and using the mass profile $M(< r_i) \propto r_i^3$ convert it into the current density profile via $\rho(r) \propto d^3\rho(r_i)/d^3\rho(r) \propto (r_i/r)^2 dr_i/dr$. For a given r_i we compute η , for that η calculate r using the equations above. The solution can be derived in analytic form relative to the background mean density, which for $\Omega_m = 1$ is $\bar{\rho} = 1/(6\pi G t^2)$. For example, for $\varepsilon = 3$:

$$\frac{\rho(r,t)}{\bar{\rho}(t)} = \frac{9}{2} \frac{(\eta - \sin \eta)^2}{(1 - \cos \eta)^3} \left[4 - \frac{9}{2} \frac{\sin \eta (\eta - \sin \eta)}{(1 - \cos \eta)^2} \right]^{-1} \quad (2.34)$$

This solution describes the density profile at radii at which the shells did not cross yet, so that enclosed mass M can be assumed constant. At radii where the shells did cross, one needs to abandon this assumption and integrate orbit for each shell taking into account changing enclosed mass for each time step. However, we can get insight by considering a simple ansatz due to Gunn (1977), in which the final apocenter of shell orbit after it enters the shell crossing regime is assumed to be a fixed fraction of shell's turnaround radius: $r \propto r_{\text{ta}}$. Given that the turnaround radius of each shell is $r_{\text{ta}} \propto r_i/\delta_i \propto r_i^{1+\varepsilon}$, we can connect the initial density profile $\rho(r_i)$ to the final density profile via $r \propto r_{\text{ta}} \propto r_i^{1+\varepsilon}$ and using again

$$\rho(r) \propto d^3\rho(r_i)/d^3\rho(r) \propto (r_i/r)^2 dr_i/dr \propto r^{-2+3/(1+\varepsilon)-1} \propto r^{-3\varepsilon/(1+\varepsilon)}. \quad (2.35)$$

Thus, if $\varepsilon = 3$, which corresponds to the point mass as the seed overdensity, the profile formed by accreting shells onto such perturber will be $\rho \propto r^{-9/4}$; numerical integration gives the constant of proportionality of $\rho(r)/\bar{\rho} = 2.79(r_{\text{ta}}/r)^{9/4}$, although this is only the average profile.

The tophat initial overdensity profile corresponds to $\varepsilon = 0$, has constant final density profile, while large values ε will result in steeper density profiles with asymptotic slope for $\varepsilon \gg 1$: $\rho(r) \propto r^{-3}$. This can be seen in the left panel of Figure 2.11, which shows density profiles predicted by different collapse models for the initial profile shown by the dotted line (realistic profile of initial Gaussian peaks). The red solid curve shows prediction of the above approximate secondary infall ansatz with assumption of $r \propto r_{\text{ta}}$ of Gunn (1977) discussed above. As can be seen, the flat region in the linear profile leads to the flat region in the collapsed halo profile in this model, while the steep region in the linear profile produces the roughly r^{-3} slope in the outer halo.

Within the simple framework of the secondary infall model, therefore, we can expect for a generic Gaussian density peak, the shells near the center of the peak where profile is shallow, produce a shallow density profile after collapse, while the shells in the outskirts of the peak where profile is steep will produce steep profile after collapse. This qualitatively explains the features of the density profile observed in cosmological simulations, shown by the dashed line in the left panel and solid line in the right panel of Figure 2.11. This *indicates a tight connection between the shape of the initial density profile of the peak and its density profile after collapse*.

2.3.9 “Virial” mass definition

In the context of the spherical collapse of a density fluctuation with the top-hat profile (i.e., uniform density, sharp boundary), collapse proceeds on the same time scale at all radii and the collapse time, “virial radius,” R_{vir} , and “virial density contrast”, Δ_{vir} , are well defined by the spherical collapse model equations. This is used to motivate definition of halo boundary as the sphere of R_{vir} , which, by definition encloses mass

$$M_{\text{vir}} = \frac{4\pi}{3} \Delta_{\text{vir,m}} \rho_m(z) R_{\text{vir}}^3(z), \quad (2.36)$$

where $\Delta_{\text{vir,m}}$ is defined as the contrast relative to the mean density of the universe, $\rho_m(z)$, at the redshift z at which mass is defined (for $\Omega_m = 1$ $\Delta_{\text{vir}} \approx 178$; for concordance ΛCDM $\Delta_{\text{vir}} \approx 360$).

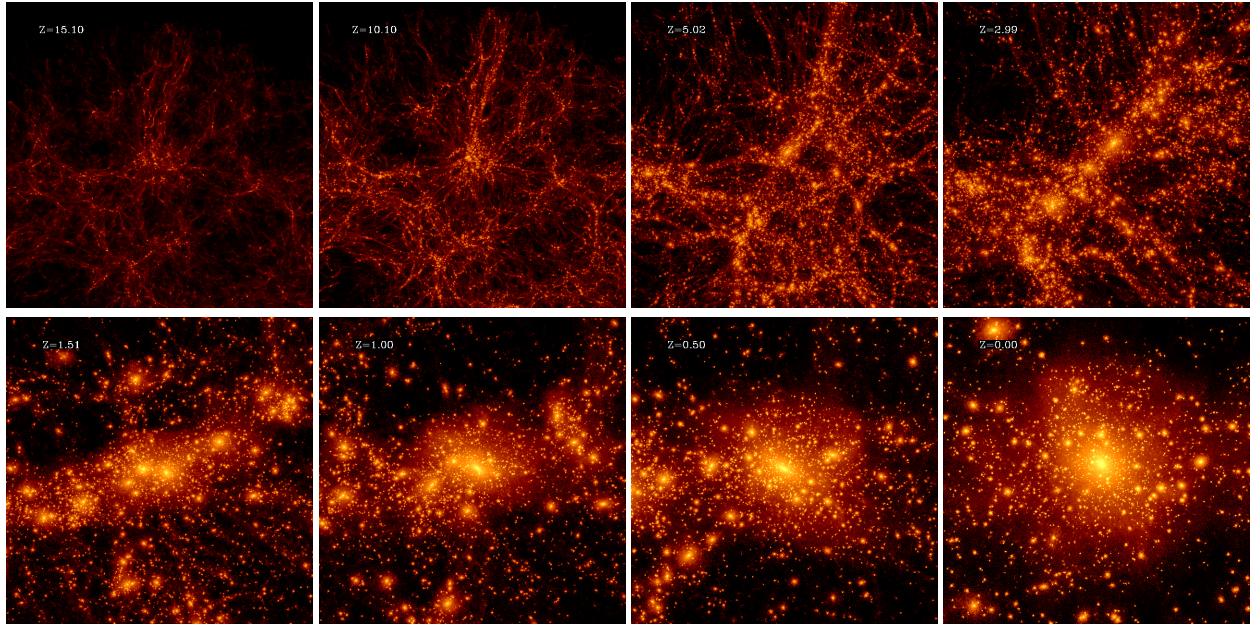


Figure 2.9: Formation of a Milky Way-sized dark matter halo in a cosmological simulation of flat Λ CDM cosmology ($\Omega_m = 1 - \Omega_\Lambda = 0.3$, $h = 0.7$, $\sigma_8 = 0.9$). The panels show an evolutionary sequence at nine redshifts (from left to right and from top to bottom) focusing on the most massive progenitor of the main halo at each epoch (redshift of each epoch is shown in the left upper corner). The rendering shows the dark matter particles with intensity indicating the local matter density on a logarithmic stretch. The build-up of the halo proceeds through a series of spectacular mergers, particularly frequent in the early stages of evolution. Many of the merging clumps survive until the present epoch ($z = 0$) in the form of "substructure". The size of the region shown is about 3 comoving Mpc at $z = 15$, monotonically zooming in to a scale of ≈ 1 comoving Mpc across $z = 0$.

However, as discussed above the peaks in the initial density field are not uniform in density, are not spherical, and do not have a sharp boundary. Existence of a density profile results in different times of collapse for different radial shells. Note also that even in the spherical collapse model the virial density contrast formally applies only at the time of collapse; after a given density peak collapses its internal density stays constant while the reference (i.e., either the mean or critical) density changes merely due to cosmological expansion. The actual overdensity of the collapsed top-hat initial fluctuations will therefore grow larger than the initial virial overdensity at $t > t_{\text{collapse}}$. The boundary based on the virial density contrast is, thus, only loosely motivated by theoretical considerations.

Moreover, the density profile around the “virial radius” is often smooth and does not show a well-defined boundary. The absence of a well-defined boundary of collapsed objects makes the definition of the halo boundary and the associated enclosed mass ambiguous. This explains, at least partly, the existence of various halo boundary and mass definitions in the literature. The commonly used definitions use density contrast of $\Delta = 200$ with respect to either critical density of the universe or relative to the mean density.

2.3.10 Collapse of real dark matter peaks

The tophat collapse and secondary infall models provide useful approximate guideline for the time scale of halo collapse and have proven to be useful tools in developing approximate statistical models for the formation and evolution of halo populations. Such a simple model and their extensions (e.g., ellipsoidal collapse model) do, however, miss many important details and complexities of collapse of the real density peaks. Such complexities are usually explored using three-dimensional numerical cosmological simulations. Techniques and numerical details of such simulations are outside the scope of this course, but can be found in

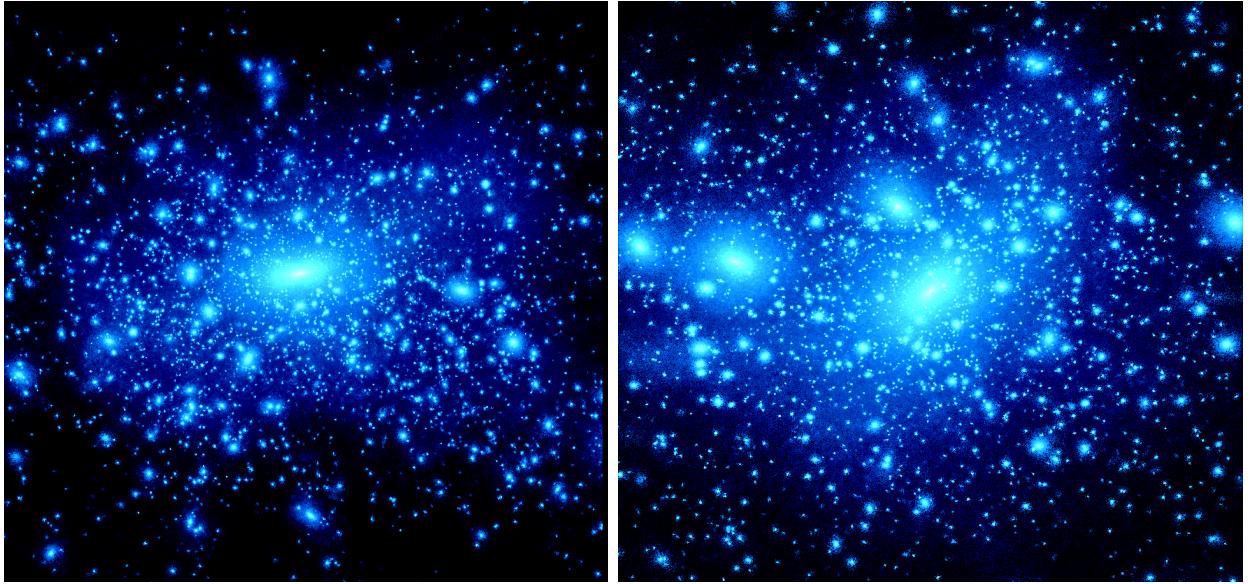


Figure 2.10: Comparison of two $z = 0$ halos of masses $3 \times 10^{14} M_{\odot}$ and $3 \times 10^{12} M_{\odot}$ formed in flat Λ CDM cosmology ($\Omega_m = 1 - \Omega_{\Lambda} = 0.3$, $h = 0.7$, $\sigma_8 = 0.9$). In each case the mass distribution around the center of the halo is shown to approximately two virial radii from the center of each halo. Both objects were resolved with similar number of particles and similar spatial resolution relative to the virial radius of the halo in their respective simulations. I leave it as an exercise to the reader to guess the mass of the halo shown in each panel.

reviews on this subject (Bertschinger 1998, Borgani and Kravtsov 2011). Here, we simply discuss the main features of gravitational collapse learned from analyses of such simulations.

The collapse of realistic CDM peaks is different from the top-hat collapse in several ways. First of all, real peaks in the primordial field do not have the constant density or sharp boundary of the top-hat, but have a certain radial profile and curvature (Bardeen et al. 1986, Dalal et al. 2008). As a result, different regions of a peak collapse at different times so that the overall collapse is extended in time and the peak does not have a single collapse epoch (see, e.g., discussion of simulation results in Diemand, Kuhlen, and Madau 2007). Consequently, the distribution of matter around the collapsed peak can smoothly extend to several virial radii for late epochs and small masses (Cuesta et al. 2008). This creates ambiguity about the definition of halo mass and results in a variety of mass definitions adopted in practice.

Second, the peaks in the smoothed density field, $\delta_R(\mathbf{x})$, are not isolated but are surrounded by other peaks and density inhomogeneities. The tidal forces from the most massive and rarest peaks in the initial density field shepherd the surrounding matter into massive filamentary structures that connect them (Bond, Kofman, and Pogosyan 1996). Accretion of matter onto clusters at late epochs occurs preferentially along such filaments, as can be clearly seen in Figure 2.9.

Finally, the density distribution within the peaks in the actual density field is not smooth, as in the smoothed field $\delta_R(\mathbf{x})$, but contains fluctuations on all scales. Collapse of density peaks on different scales can proceed almost simultaneously, especially during early stages of evolution in the CDM models when peaks undergoing collapse involve small scales, over which the power spectrum has an effective slope $n \approx -3$. Figure 2.9 shows that at high redshifts the proto-cluster region contains mostly small-mass collapsed objects, which merge to form a larger and larger virialized system near the center of the shown region at later epochs. Nonlinear interactions between smaller-scale peaks within a cluster-scale peak during mergers result in

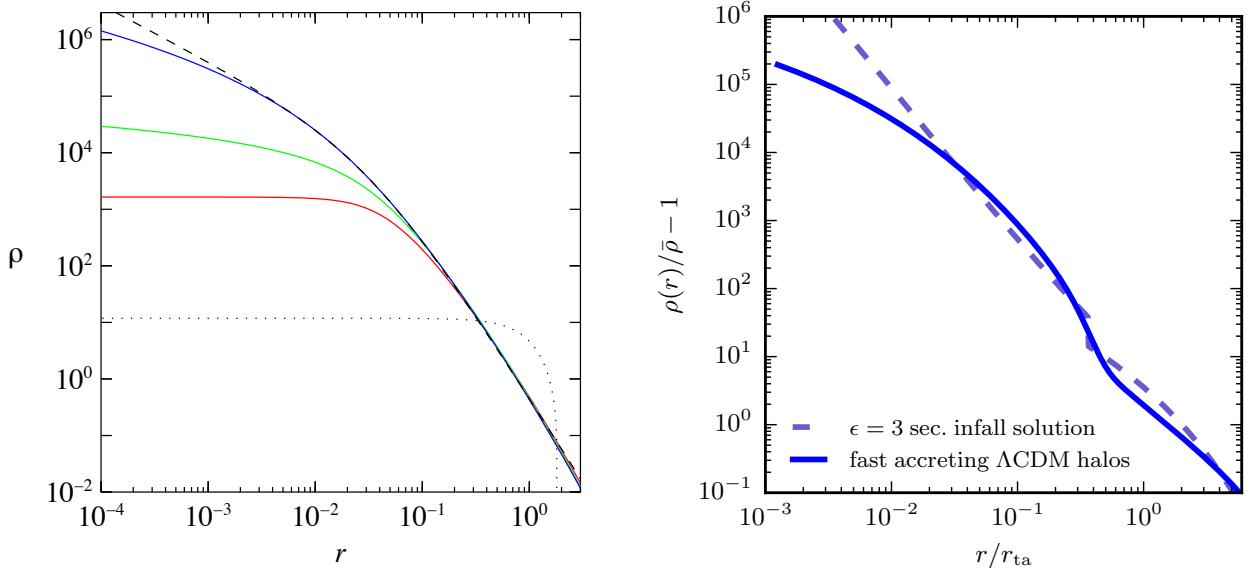


Figure 2.11: Left panel: the density profiles of collapsed peaks predicted by models of peak collapse, adopted from Dalal, Lithwick, and Kuhlen (2010). The dotted black line shows the initial overdensity profile of a peak in the Gaussian field, corresponding to stacked initial peak profiles from which low-mass halos formed in the N -body simulations of Dalal et al. (2008). The red curve shows the prediction of the frozen collapse model. The green curve shows the profile if each shell lays down interior profiles with minimal tails, without any resulting adiabatic contraction. The blue curve shows the prediction of the minimal contraction model, while the dashed line shows NFW profile typical for CDM halos. Right panel: comparison of the real density profile characteristic of rapidly accreting halos in simulations represented by the profile model of Diemer and Kravtsov (2014) and prediction of the simple secondary infall model for $\epsilon = 3$ (dashed line).

relaxation processes and energy exchange on different scales, and mass redistribution. Although the processes accompanying major mergers are not as violent as envisioned in the violent relaxation scenario (Valluri et al. 2007), such interactions lead to significant redistribution of mass (Kazantzidis, Zentner, and Kravtsov 2006) and angular momentum (Vitvitska et al. 2002), both within and outside of the virial radius.

Collapse of peaks of different size and mass proceeds very similarly and the resulting halos are qualitatively similar as well, as shown in Figure 2.10. The figure shows distribution of dark matter out to approximately two virial radii around the centers of two CDM halos of masses different by two orders of magnitude. It is clear that it is not easy to tell the mass of the halo by simply examining the overall mass distribution or by counting the number of subhalos. This is a visual manifestation of approximate self-similarity of CDM halos of different mass. The differences in structure of halos of different mass do exist due to the differences in the local slope and normalization of the power spectrum of perturbations, which result in somewhat different merger histories and formation times. However, these differences are relatively subtle.

If we would compare similar images of distribution of luminous matter around galaxies and clusters, the difference would be striking. Thus, *processes of galaxy formation break the self-similarity inherent in gravitational collapse*.

2.3.11 Density profiles of CDM halos

CDM halos in cosmological simulations exhibit profiles with steadily steepening logarithmic slope from slopes of ≤ -1 at $r < 0.01R_{200}$ to the slopes of $-3 \div -4$ at $r \sim R_{200}$ (Dubinski and Carlberg 1991, Katz 1991, Navarro, Frenk, and White 1995, Navarro, Frenk, and White 1996, Navarro, Frenk, and White 1997, Cole and Lacey 1996). The generic profile shape is "universal" in the sense that it is virtually independent of

the shape of power spectrum and background cosmology and even for non-cosmological evolution of the initial tophat profile (Navarro, Frenk, and White 1997, Huss, Jain, and Steinmetz 1999), which is a non-trivial fact. Such a generic form of the profile also arises when small-scale structure is suppressed and the collapse is smooth, as is the case for halos forming at the cut-off scale of the power spectrum (Moore et al. 1999, Diemand, Moore, and Stadel 2005). However, from the arguments of the secondary infall model, we could see that the slope is generically expected to "roll" from shallow to steep for typical density profiles of peaks in a Gaussian field. This, and additional bit of physics - adiabatic contraction of particle orbits - is what explains the "universal" shape of density profiles of CDM halos (see Figure 2.11 and Dalal, Lithwick, and Kuhlen 2010, Lithwick and Dalal 2011).

The density profiles of CDM halos within measured in dissipationless simulations at $r < R_{200}$ are most commonly approximated by the "NFW" form proposed by Navarro, Frenk, and White (1995) based on their simulation of cluster formation:

$$\rho_{\text{NFW}}(r) = \frac{4\rho_s}{x(1+x)^2}, \quad x \equiv r/r_s, \quad (2.37)$$

where r_s is the scale radius, at which the logarithmic slope of the profile is equal to -2 and ρ_s is the characteristic density at $r = r_s$. Overall, the slope of this profile varies with radius as $d \ln \rho / d \ln r = -[1 + 2x/(1+x)]$, i.e., from the asymptotic slope of -1 at $x \ll 1$ to -3 at $x \gg 1$, where the enclosed mass diverges logarithmically: $M(< r) = M_\Delta f(x)/f(c_\Delta)$, where M_Δ is the mass enclosing a given overdensity Δ , $f(x) \equiv \ln(1+x) - x/(1+x)$ and $c_\Delta \equiv R_\Delta/r_s$ is the concentration parameter. Accurate formulae for the conversion of mass of the NFW halos defined for different values of Δ are given in the appendix of Hu and Kravtsov 2003; the conversion routines are also available in the `colossus` code.

The key finding of Wechsler et al. (2002) was that the characteristic formation time a_f of equation 2.3.6 above is tightly related to halo concentration c (i.e. to its profile) as:

$$a_f \approx c_{\min}/c_\Delta. \quad (2.38)$$

Here $c_{\min} \approx 3 - 4 \approx \text{const}$ is the typical concentrations halos have in the fast accretion regime, while c_Δ is current concentration of halo measured relative to the outer radius defined to the corresponding concentration $c_\Delta = R_\Delta/r_s$.

These correlations reflect the fact that initial conditions and cosmology determine statistical properties of the initial Gaussian field. As we discussed in §2.3.6, the initial peak density profile controls the mass accretion history of halos. At the same time, density profile of resulting halo also reflects the initial density profile. This is a manifestation of limited relaxation during formation of halos, which allows them to retain very good memory of their initial conditions (amplitude and shape of the density peaks from which they formed). Thus, *density profiles of halos correlate with their MAHs because both reflect the density profile of initial peak*.

Subsequent simulations (Navarro et al. 2004, Merritt et al. 2006, Graham et al. 2006) showed that the Einasto (1965) profile and other similar models designed to describe the de-projection of the Sérsic profile (Merritt et al. 2006) provide a more accurate description of halo density profiles arising during cosmological halo collapse, as well as profiles of bulges and elliptical galaxies (Cardone, Piedipalumbo, and Tortora 2005). The Einasto profile is characterized by the logarithmic slope that varies as a power law with radius:

$$\rho_E(r) = \rho_s \exp \left[\frac{2}{\alpha} (1 - x^\alpha) \right], \quad x \equiv r/r_s, \quad (2.39)$$

where r_s is again the scale radius at which the logarithmic slope is -2 , but now for the Einasto profile, $\rho_s \equiv \rho_E(r_s)$, and α is an additional parameter that describes the power-law dependence of the logarithmic slope on radius: $d \ln \rho_E / d \ln r = -2x^\alpha$.

Note that unlike the NFW profile and several other profiles discussed in the literature, the Einasto profile does not have an asymptotic slope at small radii. The slope of the density profile becomes increasingly shallower at small radii at the rate controlled by α . The parameter α varies with halo mass and redshift:

at $z = 0$ galaxy-sized halos are described by $\alpha \approx 0.16$, whereas massive cluster halos are described by $\alpha \approx 0.2 - 0.3$; these values increase by ~ 0.1 by $z \approx 3$ (Gao et al. 2008). Although α depends on mass and redshift (and thus also on the cosmology) in a non-trivial way, Gao et al. (2008) showed that these dependencies can be captured as a universal dependence on the peak height $v = \delta_c/\sigma(M, z)$ (see section on nonlinear mass above): $\alpha = 0.0095v^2 + 0.155$.

Finally, unlike the NFW profile, the total mass for the Einasto profile is finite due to the exponentially decreasing density at large radii. A number of useful expressions for the Einasto profile, such as mass within a radius, are provided by Cardone, Piedipalumbo, and Tortora (2005), Mamon and Łokas (2005), and Graham et al. (2006).

The functional profile forms discussed above describe radial density profiles out to R_{vir} . Diemer and Kravtsov (2014) showed that at radii $r \gtrsim 1 - 2R_{200m}$ the profiles are not described by these profiles in general. First, the logarithmic slope of the profile steepens sharply near the splashback radius, R_{sp} , the location of which depends on halo mass accretion rate, as expected in the secondary infall models (see Adhikari, Dalal, and Chamberlain 2014). Second, at $r > R_{\text{sp}}$ the profile has a considerably shallower slope that changes with radius. This behavior can be seen in the profile shown by the solid line in the right panel of Figure 2.11. Thus, when one cares about the profile at $r > R_{\text{vir}}$ an accurate functional form proposed by Diemer and Kravtsov (2014) should be used. This profile is implemented in the `colossus` code along with the profiles discussed above.

The splashback radius and mass

In the previous section we have considered the approximate solution for the inner profile of collapsed shells and outer profile before shell-crossing occurs. These solutions do not join smoothly, but discontinuously. The discontinuity physically corresponds to the *density caustic* orbital apocenter of the shell that has just reached its first apocenter after it went through the peak center. This radius is called the *splashback* radius and is observed in the CDM halos (Diemer and Kravtsov 2014, Adhikari, Dalal, and Chamberlain 2014) and now also in the galaxy distribution around observed galaxy clusters (More et al. 2016).

The splashback radius, R_{sp} , physically separates the infall region from the inner halo regions, where shells have crossed. It is thus a natural physical boundary of halos (Diemer and Kravtsov 2014, More, Diemer, and Kravtsov 2015), which can be used to define halo mass $M_{\text{sp}} = M(< R_{\text{sp}})$. In practice, however, the shape of the splashback surface is not spherical due to asphericity of halo collapse. Thus, such “spherical” definition of the splashback mass is only a rough approximation. This is a relatively new area of research, but the results so far indicate that for overall galaxy formation process the difference in the mass accretion rates due to mass definition should be of secondary importance compared to the primary processes driving galaxy formation (see, e.g., detailed discussion in More, Diemer, and Kravtsov 2015).

2.3.12 Physical model for CDM halo profiles and concentrations

The origin of the generic form of the density profile has recently been explored in detail by Lithwick and Dalal (2011), who show that it arises due to two main factors: (a) the density and triaxiality profile of the original peak and (b) approximately adiabatic contraction of the previously collapsed matter due to deepening of the potential well during continuing collapse. Without adiabatic contraction the profile resulting from the collapse would reflect the shape of the initial profile of the peak, as can be seen from the Figure ?? Dalal, Lithwick, and Kuhlen (2010).

For example, if as we considered in the secondary infall model above, the initial profile of the enclosed overdensity within radius r around the peak can be described as $\delta_i \propto r_i^{-\epsilon}$, then the resulting differential density profile after collapse behaves as $\rho(r) \propto r^{-\gamma}$, where $\gamma = 3\epsilon/(1+\epsilon)$ (Fillmore and Goldreich 1984). Typical profiles of initial density peaks are characterized by shallow slopes, $\gamma \sim 0 - 0.3$ at small radii, and very steep slopes at large radii (e.g., Dalal et al. 2008), which means that resulting profiles after collapse should have slopes varying from $\gamma \approx 0 - 0.7$ at small radii to $\gamma \approx 3$ at large radii. This can be seen in Figure 2.11 from Dalal, Lithwick, and Kuhlen (2010).

The figure shows the initial density profile of a real CDM peak as the dotted line, which corresponds

to stacked initial peak profiles from which low-mass halos formed in the N -body simulations of Dalal et al. (2008). The red curve shows prediction of the above approximate secondary infall ansatz with assumption of $r \propto r_{\text{ta}}$ of Gunn (1977) discussed above in §2.3.8. The green curve is yet another approximation but still neglecting adiabatic contraction effect. The central density predicted by this model is somewhat larger than in the frozen model, but the profile still changes slope much more sharply than the profiles of real CDM halos discussed above. The solid blue line is result of a model which takes into account effect of adiabatic contraction, which reflects the conservation of radial action $J_r \equiv \int v_r dr \propto [rM(< r)]^{1/2}$. This quantity should stay approximately constant for orbits if gravitational potential changes on time scales longer than its orbital period. After shell accretes, the mass within its apocenter r grows during further mass accretion and thus r shrinks. This "minimal contraction model" matches the profile of the "NFW" profile shape characteristic of real halos (black dashed line) rather well. This model predicts at small radii, the slope of the minimally contracted profile rolls over more quickly than the NFW slope and does not actually achieve an asymptotic slope: the slope instead continues to get shallower with decreasing radius, similarly to the behavior of the Einasto profile.

Overall, in a more detailed study Lithwick and Dalal (2011) showed that contraction of particle orbits during subsequent accretion of mass interior to a given radius r leads to a much more gradual change of logarithmic slope with radius, such that the regime within which $\gamma \approx 0 - 0.7$ is shifted to very small radii ($r/r_{200} \lesssim 10^{-5}$), whereas at the radii typically resolved in cosmological simulations the logarithmic slope is in the range of $\gamma \approx 1 - 3$, so that the radial dependence of the logarithmic slope $\gamma(r) = d \ln \rho / d \ln r$ is in good qualitative agreement with simulation results shown above.

This contraction occurs because matter that is accreted by a halo at a given stage of its evolution can deposit matter over a wide range of radii, including small radii. The orbits of particles that accreted previously have to respond to the additional mass, and they do so by contracting. For example, for a purely spherical system in which mass is added slowly so that the adiabatic invariant is conserved, radii r of spherical shells must decrease to compensate an increase of $M(< r)$. This model thus elegantly explains both the qualitative shape of density profiles observed in cosmological simulations and their universality. The profiles of peaks in the Gaussian density perturbation field are universal. The contraction process crucial to shaping the form of the profile should operate under general collapse conditions, in which different shells of matter collapse at different times.

Although the functional form of the density profile arising during halo collapse is generic for a wide variety of collapse conditions and models, initial conditions and cosmology do significantly affect the physical properties of halo profiles such as its characteristic density and scale radius (Navarro et al. 1997). These dependencies are often discussed in terms of halo concentrations, $c_\Delta \equiv R_\Delta / r_s$. Simulations show that the scale radius, r_s , is approximately constant during late stages of halo evolution (Bullock et al. 2001, Wechsler et al. 2002), but evolves as $r_s \approx c_{\min} R_\Delta$ during early stages, when a halo quickly increases its mass through accretion and mergers (Zhao et al. 2003, Zhao et al. 2009). The minimum value of concentration is $c_{\min} = \text{const} \approx 3 - 4$ for $\Delta = 200$ with respect to the critical density of the universe $\rho_{\text{crit}}(z)$. For massive cluster halos, which are in the fast growth regime at any redshift, the concentrations are thus expected to stay approximately constant with redshift or may even increase after reaching a minimum.

The actual evolution is somewhat more complicated, but an accurate universal model for concentrations that works across a wide range of cosmologies, redshifts, and masses can be worked out. This was done in our recent paper (Diemer and Kravtsov 2015) and the corresponding model for concentration is implemented in the `colossus` code. The model captures the dependence of the initial Gaussian peak density profile on its peakheight v and on the slope of the power spectrum at the halo scale, which affects abundance of substructure and merger history of a given peak.

2.3.13 Halo angular momentum

Given that the peaks are not spherical in general, the gravitation torque exerted by surrounding matter distribution will lead to angular momentum generation during collapse of a peak (Doroshkevich 1973). Angular momentum changes due to nonlinear interactions during mergers of halos (Vitvitska et al. 2002).

Overall, the angular momentum grows linearly with time in the linear regime and then fluctuates randomly in the nonlinear regime. These random fluctuations are multiplicative in the specific angular momentum that is characterized by the dimensionless spin parameter $\lambda \equiv J/(\sqrt{2}MVR)$, where M , V , and R are “virial” mass, circular velocity, and radius of halo. Thus, λ specifies dimensionless specific angular momentum relative to the angular momentum which would be close to that required for rotational support at the halo virial radius R . So values of $\lambda \sim 0.5 - 1$ indicate matter that would be rotationally supported near the virial radius, while $\lambda \ll 1$ indicate that matter has to collapse to a radius $r \ll R$, before it can be rotationally supported.

The multiplicative random walk in λ during nonlinear stages of evolution leads to log-normal pdf of λ via the central limit theorem. Thus, distribution of λ measured within halo radii found in cosmological simulations to be log-normal with the median of $\lambda_{\text{med}} \approx 0.045$ and width of $\sigma_{\ln \lambda} \approx 0.55$. These parameters do not depend in detectable way on halo mass, redshift, or cosmology.

2.3.14 Scaling relations of halos

As we saw in the previous chapter, observed galaxies exhibited a number of tight equilibrium scalings. What about halos? First of all, the mass of halos and their radii are usually defined using each other, so that they are not independent: $M = 4\pi/3\Delta\bar{\rho}R^3$. This is loosely motivated by the top-hat collapse model and a strong assumption of instantaneous virialization, as discussed in §2.3.9. This relation changes with z : the mean density of the universe is larger in the past, $\bar{\rho} \propto a^{-3}$ and so for a fixed M halo radius R is smaller at higher z (smaller a).

The circular velocity at R is $V = (GM/R)^{1/2}$. If circular velocity profile $V(r)$ was constant with r , we could equate V with the rotation velocity measured in the inner regions of halos via gas and stellar motions, V_{rot} . This would be the case if mass distribution within halo had an “isothermal” profile: $\rho \propto r^{-2}$. However, as discussed above, the actual mass distribution has the Einasto profile with a rolling slope. Such profiles have characteristic scale, r_s , at which $d \ln \rho / d \ln r = -2$ and in the vicinity of which the mass distribution is thus roughly isothermal. Location of r_s relative to R is characterized by halo concentration $c = R/r_s$. For a given profile, one can derive relation between maximum circular velocity and V , which depends on halo concentration. For example, for the NFW profile: $V_{\text{max}} \approx 0.465Vc/f(c)$, where $f(c) = \ln(1+c) - c/(1+c)$. In practice, $c/f(c)$ is a weak function of mass and $V \propto V_{\text{max}}^{1+\beta}$, where $\beta \approx 0.1$ on galaxy-sized halo scales.

The relation between halo mass and rotation velocity of galaxy is thus

$$M \propto a^{3/2}V_{\text{max}}^{3(1+\beta)} \approx a^{3/2}V_{\text{max}}^{3.3}. \quad (2.40)$$

Note that the scaling with expansion factor a in the above equation arises simply due to the way M and R are defined and does not have a deep physical meaning.

Physically, the relations should be set around the halo formation epoch. Given that halo collapse is extended in time, we can associate formation epoch with the time when a typical halo of mass M reaches the linear collapse threshold predicted by the spherical collapse model: $\delta_c \approx 1.69$ (equivalently, we can say that formation epoch is the time when halos of mass M have peak height of $v \approx 1$). The typical amplitude of peaks of mass M is given by the power spectrum amplitude $\delta \propto D_+(a)[\Delta^2(k)]^{1/2} \propto [k^3 P(k)]^{1/2}$. If we approximate the power spectrum at the scales of perturbations which form galaxy-sized halos as a power law, $P(k) \propto k^n$, $\delta \propto D_+(a)k^{(3+n)/2} \propto D_+(a)R^{-(3+n)/2} \propto D_+(a)M^{-(3+n)/6}$. Assuming $\Omega_m \approx 1$, which is a reasonable assumption during formation epochs of most galaxies, $D_+(a) \propto a$, and we have $\delta_c \approx 1.69 \propto aM^{-(3+n)/6}$ or

$$a_f \propto M^{(3+n)/6}. \quad (2.41)$$

For typical galaxy-sized perturbations n varies from ≈ -2.2 to ≈ -2.8 , so we can take $n = -2.5$ as the representative value which gives:

$$a_f \propto M^{0.08}, \quad (2.42)$$

which shows that formation epoch depends rather weakly on halo mass, but nevertheless more massive halos will collapse later (larger a_f). Halo concentration, which is set at a_f scales as $c \propto a_f^{-1} \propto M^{-0.1}$ (e.g., Wechsler et al. 2002).

Relation between halo mass and rotation velocity and halo radius can then be obtained by substituting a_f in eq. 2.41 instead of a in eq. 2.40:

$$M \propto V^{12/(1-n)} \propto R^{6/(5+n)} \propto V_{\max}^{12(1+\eta)/(1-n)}. \quad (2.43)$$

For $n = -2.5$ and $\eta = 0.1$, $12(1+\eta)/(1-n) \approx 3.77$. This is not too far from the slope of the baryonic Tully-Fisher relation discussed in §??, if we had M_{bar} scaling approximately linearly with M .

2.4 Baryonic processes affecting galaxy formation

Although baryons are collisional and can experience pressure forces and shocks, the actual collapse until the shell crossing zone (i.e. until $r \sim R_{\text{splashback}}$) is governed by the dark matter potential and is actually quite similar to what we considered for collisionless matter above. Consequently profiles are almost identical with the physical difference that baryons experience shock approximately at $r \sim R_{\text{splashback}}$ and do not orbit freely within that radius as dark matter does. The actual equilibrium profile within $R_{\text{splashback}}$, however, is very similar for the gas and dark matter, except in the central $\sim 10\%$ of halo virial radius, where baryon density profile is shallower. A reasonably accurate model of gas density profile for a halo with a given dark matter profile can be derived with a couple of simple assumptions, as described by Patej and Loeb (2015).

In the self-similar secondary infall models discussed above, solution for the gas can be obtained by solving 1d ordinary differential equations (see S 8.3.2 in MvdBW or the original paper by Bertschinger 1985). Cosmological simulations confirm similarity of dark matter and gas density profiles (e.g., Lau et al. 2015).

2.4.1 The filtering scale in baryon perturbations after recombination

The baryon perturbations also grow as $\propto a$ once they catch up in amplitude to the dark matter perturbations on all scales larger than roughly the Jeans' length. The latter becomes very small after recombination as the universe cools adiabatically, but becomes larger again and relevant for formation of dwarf galaxies after the universe is reionized and the intergalactic medium is heated to $T \sim 1000 - 20000$ K. The mathematical formalism describing effect of such re-heating on the evolution of perturbations is presented in Gnedin (2000). The quantitative effect on accretion of baryons on a collapsing peak is calibrated using numerical simulations of galaxy formation (e.g., Hoeft et al. 2006, Okamoto, Gao, and Theuns 2008). It predicts that baryon fractions in halos of mass M at redshift z is modified from a constant value \bar{f}_b as:

$$f_b(M, z) = \bar{f}_b \left\{ 1 + (2^{\alpha/3} - 1) \left[\frac{M}{M_c(z)} \right]^{-\alpha} \right\}^{-3/\alpha}, \quad (2.44)$$

with $\alpha = 2$. The cutoff mass M_c is derived from simulations of galaxy formation. For example, $M_c(z)$, in figures of Okamoto, Gao, and Theuns (2008) is well approximated by the following formula:

$$M_c(z) = 6 \times 10^9 h^{-1} M_\odot \exp(-0.63z), \quad \text{for } z < 9, \quad (2.45)$$

while M_c can be assumed to be $\approx 10^5 - 10^6 h^{-1} M_\odot$ at $z > 9$, because temperature of the gas is much smaller before reionization.

2.4.2 Thermodynamics of baryon plasma

Unlike dark matter, when gas collapses and settles into an equilibrium profile it can then radiate away the internal energy it acquired during gravitational collapse thanks to a plethora of radiative (i.e. cooling) processes. On such process is the Compton cooling, in which the electrons Compton scatter CMB photons which have low energy due to redshifting at the redshifts relevant for galaxies. This process depends only on

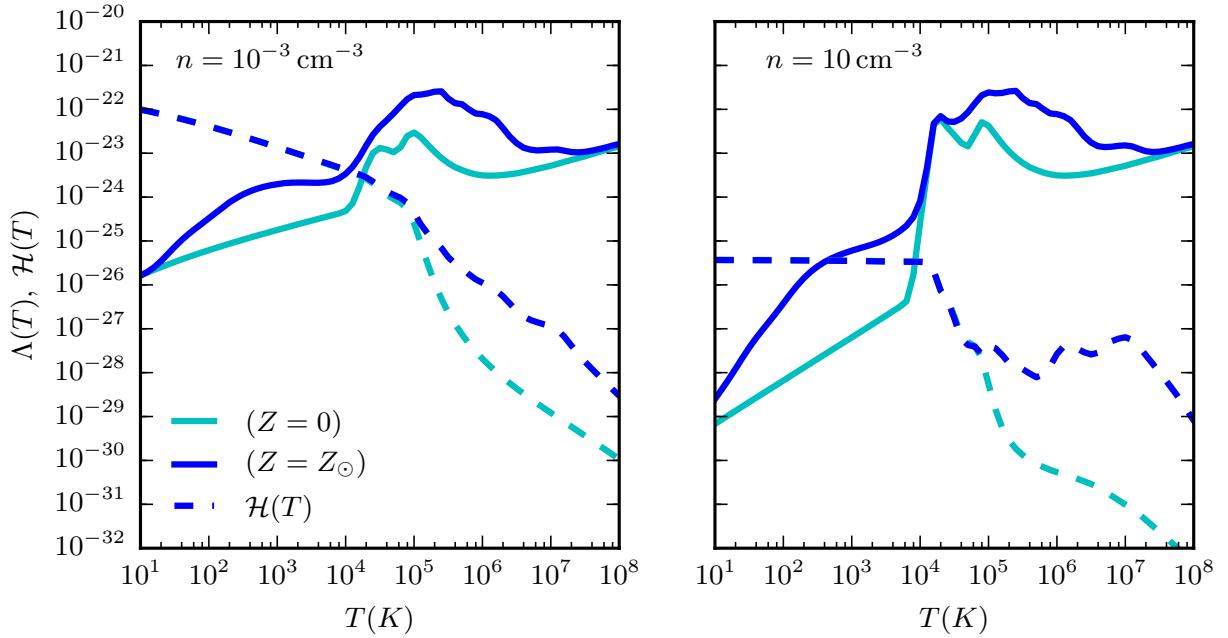


Figure 2.12: Examples of the cooling (solid) and heating (dashed) functions of gas of low (left panel) and high (right panel) density are shown for primordial (cyan) and solar (blue) metallicities, assuming ambient UV flux equal to the cosmic mean at $z = 0$. The plotted functions were generated using accurate approximations of Gnedin and Hollon (2012) for $Z = 0$ and $Z = Z_{\odot}$ to the actual calculations with the Cloudy plasma emission code.

the plasma density and temperature and energy density of the CMB photons (which is set by observations). It only operates in ionized gas (free electrons are needed). It is also the only major cooling process with rate that depends linearly on plasma density (see S 8.1.2 in MvdBW book):

$$\mathcal{L}_{\text{Compon}} = \frac{4kT_e}{m_e c^2} c \sigma_T n_e a T_{\text{CMB}}^4 \quad (2.46)$$

The *cooling time* defined the thermal energy divided by the cooling rate, for the Compton cooling and ideal monatomic gas is:

$$t_{\text{Compton}} \approx \frac{3}{2} \frac{n_e k T_e}{\mathcal{L}_{\text{Compton}}} = \frac{3 m_e c}{4 \sigma_T a_r T_{\text{CMB}}^4}, \quad (2.47)$$

i.e., it is independent of density. The Compton cooling is only shorter than or comparable to the Hubble time (current age of the universe) at $z > 5 - 6$.

The other relevant processes all involve two particles - i.e., electron and ion, ion-ion, or electron-ion. For neutral or partially ionized gas it could be neutral atoms, neutral-electron, or electron-ionized atoms. For the hydrogen-helium plasma, these processes can be written down in a page or a couple of pages (e.g., see Fig 1 in Weinberg, Hernquist, and Katz 1997 or Appendix A.2 in Gnedin and Kravtsov 2011).

As gas gets enriched with heavy elements, the number of different ions present and possible processes of their mutual interactions is large. So the usual practice is not to follow all these via equations (although see exceptions like recent work by Richings, Schaye, and Oppenheimer 2014), but use a plasma emission code, such as Cloudy to tabulate the cooling rates as a function of gas metallicity, density, and ionized radiation (e.g., public releases of such tabulations in papers by Wiersma, Schaye, and Smith 2009 and Gnedin and Hollon 2012: see [here](#) and [here](#)).

Given that cooling rates depend trivially on n^2 , the actual tabulated quantity is so-called *cooling function*: $\Lambda(T, Z, \ell) \equiv \mathcal{L}/n^2$.

Examples of the cooling functions of gas of low and high density are shown in Figure 2.12 for primordial and solar metallicities. The plotted functions were generated using accurate approximations of Gnedin and Hollon (2012) for $Z = 0$ and $Z = Z_\odot$ to the actual calculations with the [Cloudy](#) plasma emission code.

Cooling time is a characteristic time scale over which plasma loses its internal energy: $\dot{e} = e/t_{\text{cool}}$, where $\dot{e} = H - n^2 L$ is the net rate of heating/cooling. The cooling time also depends on the physical *gas* density: $n(r) = \rho_{\text{gas}}(r)/(\mu m_p) = f_{\text{gas}} \rho_{\text{tot}}/(\mu m_p)$, where the mean molecular weight for ionized plasma is $\mu \approx 0.6$ and $m_p = 1.67262 \times 10^{-24}$ g is the mass of proton:

$$t_{\text{cool}} = \frac{e}{\dot{e}} = \frac{n k T}{(\gamma - 1)(H - n^2 \Lambda)}, \quad (2.48)$$

where $\gamma = 5/3$ for ionized plasma.

2.4.3 Baryon cooling and condensation

Given the density dependence, and the fact that halos have density profiles, the cooling time will increase with increasing radius. If cooling proceeds in a normal way, the radius at which local free fall time defined as the free-fall time for a uniform sphere:

$$t_{\text{ff}} \equiv \left(\frac{3\pi}{32G\rho_{\text{tot}}} \right)^{1/2} \quad (2.49)$$

is roughly equal to the cooling time, $t_{\text{cool}}(r_{\text{cool}}) = t_{\text{ff}}(r_{\text{cool}})$, will be the radius within which we can expect most gas to lose its internal energy, associated thermal pressure support, and condense to the center of halo. Note that at a given *physical* radius (i.e., the radius in say kpc without scaling) the physical density for halos of different mass will be different and it is the physical density that enters into equation 2.49.

Even if $t_{\text{cool}} > t_{\text{ff}}$ locally, gas may be *thermally unstable* locally. This means that small perturbations due to turbulence, motion of satellite galaxies, potential changes, etc. can trigger cooling and condensation in local regions, which will lead to formation of dense, cold clouds in pressure equilibrium with ambient hotter gas. It turns out that this process operates efficiently when $t_{\text{cool}} \lesssim 10t_{\text{ff}}$ (Sharma et al. 2012). *Precipitation* of gas via clouds forming by such thermal instability may be as efficient as runaway cooling of all gas in condensing gas out of halo, although the time scale for gas condensation may be somewhat longer. The role of thermal instability induced precipitation is to extend the range of radii at which plasma may be subject to cooling and condensation towards the halo center.

2.4.4 Radius of angular momentum support

As baryons cool they lose pressure support and have to condense towards the center of the potential well of the halo. This process would continue until baryons go to $r = 0$ if there was no angular momentum. However, as mentioned above both dark matter and baryons acquire angular momentum due to gravitational torques from surrounding large scale structures which spin the collapsing peak. The amount of angular momentum imparted by this process is small compared to what would be needed to make the entire collapsed halo rotationally supported (typical halo spin is $\lambda \approx 0.05$ so only few per cent of that needed for rotational support on the halo scale, see §2.3.13), but it is not zero. This is sufficient to stop gas by the rotational support (i.e. centrifugal force) at some radius $r \sim \lambda R_{200}$. This is because at the virial radius by definition specific angular momentum is $j = \lambda \sqrt{2} V_{200} R_{200}$, and if the specific angular momentum is conserved, we have at the radius of rotational support $j = r V_c(r) = \lambda \sqrt{2} V_{200} R_{200}$, so $r = \lambda \sqrt{2} R_{200} [V_{200}/V_c(r)]$. For typical halos circular velocity profile is almost flat, $V_c \approx V_{200} \approx \text{const}$ and constant is only weakly dependent on halo mass. So, approximately $r = \eta \lambda \sqrt{2} R_{200} R_{200}$, where η is of order unity.

This was one of the key ideas in the early models of disk galaxy formation by Fall and Efstathiou (1980) and by Ryden and Gunn (1987). If the angular momentum is parametrized via dimensionless spin, as described above, the detailed calculations of rotationally supported disk size for baryons condensing in

gravitational potential corresponding to the NFW profile is predicted to be (Mo, Mao, and White 1998; hereafter MMW98):

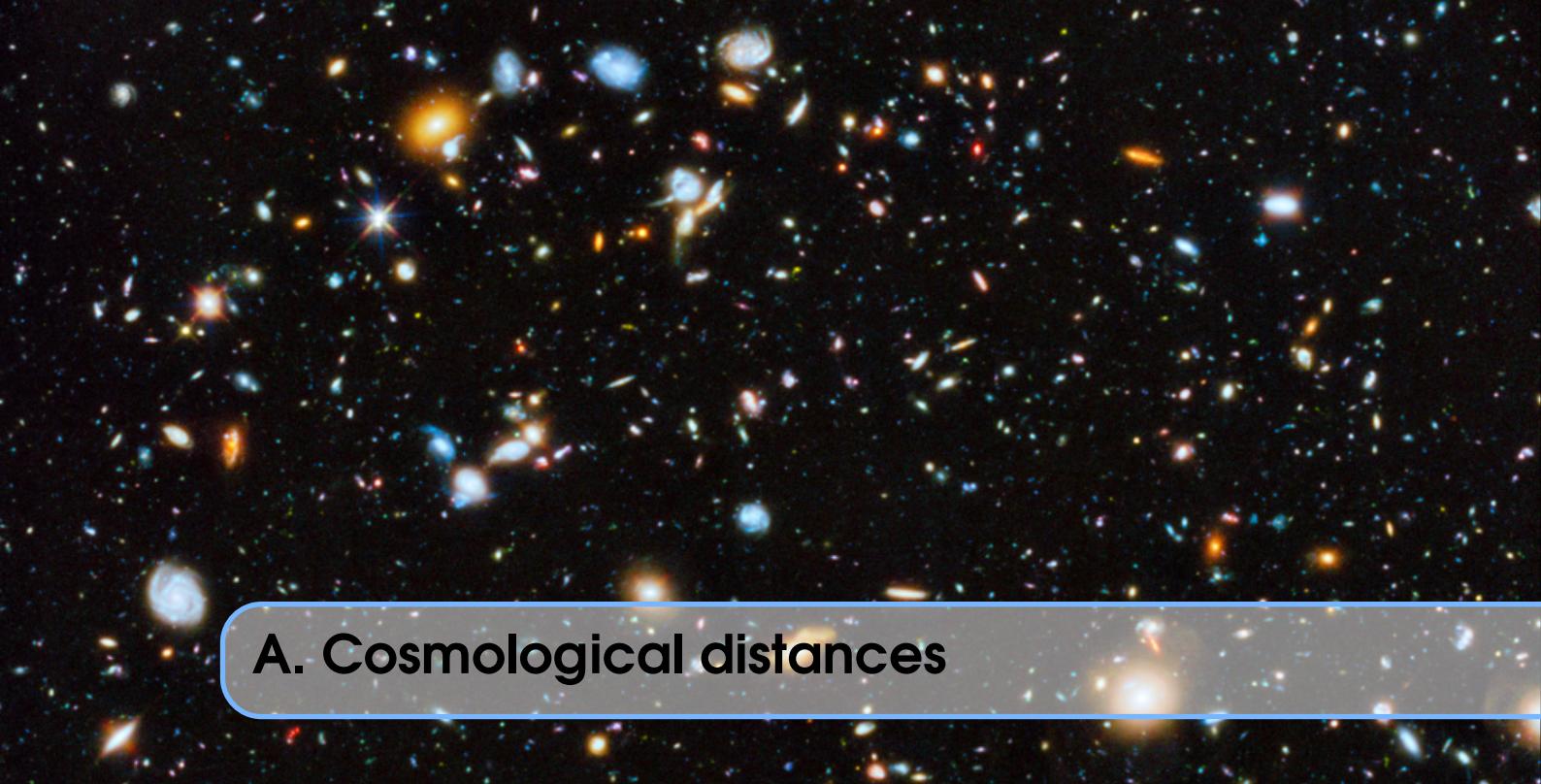
$$r_{1/2} = 1.678R_d = 1.187(j_d/m_d)f_c^{-1/2}f_R\lambda R_{200}, \quad (2.50)$$

where $r_{1/2}$ is disk half mass radius, R_d is disk exponential scale length, R_{200} is the radius enclosing density contrast of 200 relative to the critical density at the redshift of interest, $c = c_{200}$ is concentration of halo profile (see above), j_d and m_d are fractions of baryon angular momentum and mass within halo in the central disk (e.g., if disk mass is $M_d = 5 \times 10^{10} M_\odot$ and $M_h = 10^{12} M_\odot$, $m_d = 5 \times 10^{10}/10^{12} = 0.05$), $f_c \approx 0.667 + (c_{200c}/21.5)^{0.7}$ is a function of halo concentration (see eq. 23 in MMW98 for the exact expression), and f_R is a function that can be used to take into account baryonic contraction of halo in response to halo formation (see S 2.3 of MMW98 and approximate formula for this quantity in their eq. 32). Assuming $j_d/m_d = 1$, $m_d = 0.05$, and typical spin $\bar{\lambda} = 0.045$ and halo concentration of $c_{200} = 10$, gives $f_c \approx 1.25$ and $f_R \approx 0.66$, and thus *linear relation between galaxy size and halo virial radius*: $r_{1/2} = 0.032R_{200c}$.

The scatter in size at fixed R_{200c} in such model is mostly due to scatter in λ – the spin of parent halo. As discussed above, λ has log-normal pdf with the width of $\sigma_{\ln \lambda} \approx 0.55$ (e.g., Vitvitska et al. 2002).

2.5 Additional reading

The origin and evolution of density perturbations. The origin of density fluctuations in the universe and their evolution is covered in detail in Ch. 4 of MvdBW book, Anthony Lewis's [notes](#) on generation and evolution of perturbations during inflation. Chapter 7 of Scott Dodelson's "Modern Cosmology" book and Chapter 10 of Steven Weinberg's "Cosmology" book offer even more complete treatment. I highly recommend Lineweaver (2005) as a high-level introduction into inflation scenario and associated issues.



A. Cosmological distances

Reading: Hogg (1999)

More pedagogical reading: Dodelson, "Modern cosmology" pp. 33-36, Mo, van den Bosch & White "Galaxy formation", §3.1.1-3.1.4, 3.1.6

Distances are one of the most difficult quantities to measure in astronomy. Although a number of good methods, which constitute rungs of the *distance ladder* (e.g., see review by Freedman and Madore 2010 and an older but methodologically excellent review by Jacoby et al. 1992), exists, such measurements are difficult to do for large numbers of galaxies identified and observed in modern galaxy surveys, especially at distances $> 50 - 100$ Mpc. Therefore, most commonly the Hubble law (Hubble, 1929) is exploited to convert redshift into distance. Redshifts for such measurements are usually measured from spectra of galaxies of suitable resolution. However, now increasingly the *photometric redshifts* are measured from broad band photometry of galaxies. These are not particularly accurate individually but can be used to measure an accurate redshift of a galaxy cluster or for problems where high distance accuracy is not critical.

A good collection of practical information about calculation of distances given a redshift of an object in expanding universe is provided by Hogg (1999, see also Dodelson, "Modern cosmology" pp. 33-36). What follows is a brief summary of the information relevant for galaxies that we will use lifted from that paper with minor modifications and addition of section on calculation of surface brightness by me. This info is for the standard Λ CDM cosmology, if you will ever need dark energy cosmology (with constant or varying w), you need S 2.1 and 2.2 in Frieman, Turner, and Huterer (2008).

Expansion of the universe leads locally to the Hubble law scaling between the recession velocity, v , implied by the redshift of spectral lines in spectrum, $z = \lambda_{\text{obs}}/\lambda_{\text{em}} - 1$, and distance, d :

$$v = cz = H_0 d, \quad (\text{A.1})$$

where the Hubble constant H_0 is the proportionality constant. This constant has units of inverse time and its inverse can be used to define expansion time scale called *the Hubble time*:

$$t_H \equiv \frac{1}{H_0} = 9.78 \times 10^9 h^{-1} \text{ yr} = 3.09 \times 10^{17} h^{-1} \text{ s.} \quad (\text{A.2})$$

H_0 is often expressed in units of km/s/Mpc because it is measured locally from the Hubble expansion law $H_0 = v/d$ with recession velocity v expressed in km/s and distance d expressed in Mpc. It is also often parameterized as $H_0 = 100h$ km/s/Mpc, where h is dimensionless Hubble constant.

A.1 h -scalings

Given that H_0 is not yet measured precisely, different people use somewhat different values of h and these assumed values then affect any quantity that is estimated using distances. When comparing such quantities from different papers, one has to be aware of the assumed h values and how different quantities depend on them in order to correct the difference in the assumed Hubble constants properly. Historically, these scalings were useful because values of h assumed in different papers could be different by a factor of two. Expressing quantities computed for $h = 1$ with explicit h -scalings was a useful way to allow people to rescale these quantities to the value of h they prefer. This logic is still retained, although h is now known much better and is within 5% of $h = 0.7$.

Here I review the main origin of such scalings for quantities for which they are most commonly quoted. Distances measured using redshift, scale inversely with h : $d = cz/H_0 \propto h^{-1}$. Luminosities of galaxies depend on measured flux, f , and inverse square of the distance: $L = 4\pi f d_L^2$. Thus, the h -scaling of luminosity is $L \propto f d^2 \propto h^{-2}$, while the absolute magnitude scales as $M \propto \dots + 5 \log_{10} h$, so that absolute magnitudes are commonly plotted in units of $\text{mag} - 5 \log_{10} h$.

Physical sizes or observables defined within a given aperture scale with h because distances are used to convert observed angular scale, θ , to physical size within which an “observable” is defined, $R = \theta d_A(z) \propto \theta h^{-1}$. Thus, if the total mass M of a galaxy or galaxy cluster is measured using the hydrostatic equilibrium equation and measurement of the temperature of velocity dispersion profile, we have $M_{\text{HE}} \propto \sigma R$ or $\propto T R \propto d_A \propto h^{-1}$. The same scaling is expected for the mass derived from the weak lensing shear profile measurements.

If the gas mass is measured from the X-ray flux from a volume $V \propto R^3 \propto \theta^3 d_A^3$, which scales as $f = L_X/(4\pi d_L^2) \propto \rho_{\text{gas}}^2 V / d_L^2 \propto M_{\text{gas}}^2 / (V d_L^2) \propto M_{\text{gas}}^2 / (\theta d_L^2 d_A^3)$ and where f and θ are observables, gas mass then scales with distance as $M_{\text{gas}} \propto d_L d_A^{3/2} \propto h^{5/2}$. This dependence can be exploited to constrain cosmological parameters, as in the case of X-ray measurements of gas fractions in clusters.

A.2 Redshift space distortions

There is usually a difference between an object’s measured redshift z_{obs} and its *cosmological redshift* z_{cos} due purely to the expansion of the universe. The difference is due to its (radial) peculiar velocity v_{pec} ; i.e., we define the cosmological redshift as that part of the redshift due solely to the expansion of the Universe, or *Hubble flow*. The peculiar velocity is related to the redshift difference by

$$v_{\text{pec}} = c \frac{(z_{\text{obs}} - z_{\text{cos}})}{(1+z)} \quad (\text{A.3})$$

where it was assumed $v_{\text{pec}} \ll c$. To estimate v_{pec} requires distance measurement independent of redshift, which is difficult but possible for limited number of galaxies. Thus, there are studies that estimate peculiar velocity field of galaxies locally. However, for most galaxies in the survey like the SDSS independent measurements are not available. From here on, we assume $z = z_{\text{cos}}$ when estimating distances. We should just keep in mind that in some cases we need to take into account the fact that redshift has two components to it, when we examine spatial distribution of galaxies in *redshift space*, as redshift derived distances will not correspond to the true spatial distances due to peculiar motions of galaxies relative to the uniformly expanding background. These motion lead to what’s called *redshift space distortions*.

A.3 Cosmological distances

In what follows we will discuss distances from an observer to an object at redshift z under the assumption that peculiar velocities can be neglected, i.e. redshift is given by the pure uniform cosmological expansion.

Distances in expanding universe depend both on the expansion history, quantified via the *expansion factor* $a(t)$ measuring how distances between objects in the Hubble flow change with time and normalized so that $a(t_{\text{now}}) = 1$. In terms of cosmography, the cosmological redshift z is directly related to the scale factor $a(t)$:

$$1 + z = \frac{a(t_0)}{a(t_e)} \quad (\text{A.4})$$

where $a(t_0)$ is the scale factor at the time the light from the object is observed, and $a(t_e)$ is the scale factor at the time it was emitted.

The expansion history $a(t)$ is derived from the solution of the General Relativity equations and depends on the energy and density components that fill the universe. In what follows, we will assume that the main components of the universe during epochs when we compute distances is matter and vacuum energy.¹ The latter is parametrized by the cosmological constant Λ , which is assumed to be constant in time. The matter density, on the other hand, changes with time as the volume filled by a given mass expands with the Hubble flow: $\rho_m \propto a^{-3} = \rho_{m,0}a^{-3}$.

The density of matter and vacuum energy can be expressed as dimensionless density parameters Ω_M and Ω_Λ by normalizing by the *critical density of the universe*, $\rho_{\text{cr}}(z) = 3H^2(z)/8\pi G$, that would be required to make it spatially flat:

$$\Omega_m \equiv \frac{8\pi G \rho_m(z)}{3H^2(z)} = \Omega_{m,0} \frac{(1+z)^3}{E^2(z)}, \quad (\text{A.5})$$

where $E(z) \equiv H(z)/H_0$ is the dimensionless Hubble constant. Likewise,

$$\Omega_\Lambda \equiv \frac{\Lambda c^2}{3H^2(z)} = \frac{\Omega_{\Lambda,0}}{E^2(z)} \quad (\text{A.6})$$

(Peebles, 1993, pp. 310–313), where the subscripted zeroes indicate that the quantities are evaluated at $z = 0$. The present-day value of the critical density is

$$\rho_{\text{cr},0} = \frac{3H_0^2}{8\pi G} \approx 2.77 \times 10^{11} h^2 M_\odot \text{Mpc}^{-3}. \quad (\text{A.7})$$

A third density parameter Ω_k measures the “curvature of space” and can be defined by the relation

$$\Omega_M + \Omega_\Lambda + \Omega_k = 1 \quad (\text{A.8})$$

These parameters completely determine the geometry of the Universe if it is homogeneous, isotropic, and radiation-like components are negligible, as will be assumed to be the case in our study of galaxies.

The best current estimates of these parameters from the CMB anisotropies, Baryonic Acoustic Oscillation feature in galaxy clustering, etc. can be found in Table 1 of Planck Collaboration (2016). The precise values of some of the key parameters - e.g., Ω_m , H_0 , and σ_8 are still under debate at the $\sim 5 - 10\%$ level, as different observational probes of these parameters are in tension with each other.

A.3.1 Comoving line-of-sight distance

When we observe an object that exhibits a cosmological redshift z of its spectrum, the redshift was accumulated during the time light travelled to us after it was emitted by the object. The emitted photons had to play catch up: in addition to covering the initial *physical* distance between us and the object at the time light was

¹There is also radiation, but it does not contribute significantly at the redshifts where we observe galaxies.

emitted, the photons had to travel additional distance due to expansion of the universe between emission and observation.

The redshift of the photons during this travel was accumulated via a sequence of small redshifts $dz = H(z)\delta d/c$. Thus, the total distance traveled by photons to accumulate redshift $z = \int_0^z dz$ is

$$d_c = \int_0^z \delta d = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}, \quad (\text{A.9})$$

where $E(z) = H(z)/H_0$ is the dimensionless Hubble rate:

$$E(z) \equiv H(z)/H_0 = \sqrt{\Omega_{\text{m},0}(1+z)^3 + \Omega_k(1+z)^2 + \Omega_{\Lambda,0}} \quad (\text{A.10})$$

The ratio c/H_0 is often denoted as D_H – the Hubble distance:

$$D_H \equiv \frac{c}{H_0} = 3000 h^{-1} \text{ Mpc} = 9.26 \times 10^{27} h^{-1} \text{ cm.} \quad (\text{A.11})$$

is the Hubble distance and $E(z)$ is dimensionless Hubble parameter:

In cosmology, the actual physical separation between two objects at a given redshift is called *proper* or *physical* distance, while distance between these objects at $z = 0$, assuming they were moving only with the Hubble expansion flow, is called *comoving* distance. The way d_c distance is defined above, it is clearly the distance between observed object and us today ($z = 0$) along the line-of-sight to the object. Thus, d_c is called the *line-of-sight comoving distance*. It is often also denoted by χ in cosmology textbooks.

Given that $a = (1+z)^{-1}$, $dz = -da(1+z)^2 = -a^{-2}da$, while da is related to dt via the Hubble rate $H(a) \equiv a^{-1}da/dt$, so $dt = da/(aH)$, d_c can be expressed as

$$d_c = D_H \int_a^1 \frac{da}{a^2 H(a)} = D_H \int_t^{t_U} \frac{dt'}{a(t')}, \quad (\text{A.12})$$

where t_U is the age of the universe at $z = 0$. When this distance is evaluated for $a = 0$ or $t = 0$ this gives the largest distance the light could have travelled since the Big Bang:

$$d_h = D_H \int_0^{t_U} \frac{dt'}{a(t')}. \quad (\text{A.13})$$

Such distance defines the size of the space regions that can be causally connected and is called *comoving horizon distance*.

A.3.2 Angular diameter distance

In astronomy we often want to evaluate physical size of an object or a region that has redshift z and subtends an angle on the sky θ . Let us first consider how this distance is related to the comoving line-of-sight distance in the case of flat cosmological models ($\Omega_k = 0$), when usual Euclidian geometry applies, and then generalize to $\Omega_k \neq 0$ models.

The *angular diameter distance* d_A is defined as the ratio of an object's *physical* transverse size, s , to its angular size θ (in radians): $s \equiv \theta d_A$, where we assume small-angle regime. Let us consider 3 photons emitted at the same time from the center of the region, and from the distances $\pm s/2$ from the center. As these photons travel towards us, the distance between them grow due to expansion of the universe, so that by the time they are observed by us the distance between off-center photons appears to be s/a , while the central photon travelled distance d_c to us, as discussed above. Thus, the angle subtended by the emitting region will be $\theta = s/a/d_c \equiv s/d_A$, where $d_A = ad_c = d_c/(1+z)$ is an effective distance for which the usual small-angle formula applies.

A curious property of d_A is that it does not increase monotonically with increasing z but turns over at $z \sim 1$ and thereafter more distant objects actually appear larger in angular size.

When $\Omega_k \neq 0$, Euclidian geometry does not apply and relation between θ and s has to take into account the curved paths of the photons reflecting the curvature of space. In general case, angular diameter distance by analogy to the above is written in relation to the *transverse comoving distance*, d_M as

$$d_A = \frac{d_M}{1+z}, \quad (\text{A.14})$$

where

$$d_M = \begin{cases} D_H \frac{1}{\sqrt{\Omega_k}} \sinh [\sqrt{\Omega_k} d_C / D_H] & \text{for } \Omega_k > 0 \\ d_c & \text{for } \Omega_k = 0 \\ D_H \frac{1}{\sqrt{|\Omega_k|}} \sin [\sqrt{|\Omega_k|} d_c / D_H] & \text{for } \Omega_k < 0 \end{cases} \quad (\text{A.15})$$

A.3.3 Luminosity distance

In astronomy it is often required to estimate intrinsic luminosity of an object, L , given its observed flux, f . In the non-expanding space, these are related by the distance: $L = 4\pi d^2 f$. In the expanding space, expansion has effect on the rate at which photons are received and on the photon energy via redshift.

For simplicity, we will consider luminosity of photons of a given wavelength λ ; luminosity of photons of other wavelengths can be treated in the same way. Luminosity is energy emitted per unit time, while flux is energy received by the detector per unit area per unit time. Distance between two photons emitted at times separated by Δt_e will initially be $c\Delta t_e$, but will grow as the universe expands by a by the time the photons are observed. Thus, the photons will be received with larger time interval, $\Delta t_o = \Delta t_e/a$, i.e. their flux is reduced by a factor a . In addition, the wavelength of the photons, and thus their energy, decreases as they propagate to us by a . Thus, the flux through the sphere of radius d_c – the distance photons travelled to us – is reduced by a^2 compare to no expansion case: $f = La^2/(4\pi d_c^2) \equiv L/(4\pi d_L^2)$, where $d_L = d_c/a = d_c(1+z)$ is the *luminosity distance* – i.e. the distance relating flux and luminosity in the expanding space, in the way similar to the non-expanding space.

For general cosmological model it is given by

$$d_L = (1+z)d_M = (1+z)^2 d_A. \quad (\text{A.16})$$

Incidentally, this relation implies that the surface brightness, S , of a receding object of luminosity L that subtends a given solid angle Ω will decrease with increasing redshift as $S = \Omega^{-1}L/d_L^2 \propto 1/[(1+z)^4 d_A^2] \propto 1/[(1+z)^4 A]$, where $A = d_A^2 \Omega$ is proper area of the object. This is known as $(1+z)^4$ *cosmological surface brightness dimming* and is one of the main factors it makes it difficult to observe galaxies at high redshifts (even though their sizes do not change much).

A.4 Apparent and absolute magnitudes and k -correction

The *apparent magnitude* m of an astronomical source in a photometric bandpass is defined to be the ratio of the apparent flux of that source to the apparent flux of the bright reference stars through that bandpass. The *distance modulus* DM is defined by

$$DM \equiv 5 \log \left(\frac{D_L}{10 \text{ pc}} \right) \quad (\text{A.17})$$

because it is the magnitude difference between an object's observed bolometric flux and what it would be if it were at 10 pc (this was once thought to be the distance to Vega which was the most commonly used standard star until recently).

The absolute magnitude M is the astronomer's measure of luminosity, defined to be the apparent magnitude the object in question would have if it were at 10 pc, so

$$m = M + DM + K(z) \quad (\text{A.18})$$

where K is the so called *k-correction*:

$$K(\lambda_0) = 2.5 \log \left[(1+z) \frac{\int_0^\infty f(\lambda_0) S(\lambda) d\lambda}{\int_0^\infty f[\lambda_0/(1+z)] S(\lambda) d\lambda} \right], \quad (\text{A.19})$$

where λ_0 is the wavelength of the filter at z_0 to which we would like to correct the magnitudes (Oke and Sandage 1968, see §4 in Blanton and Roweis 2007 for a more complete treatment).

The k-correction depends on the spectrum of the object in question, and is unnecessary only if the object has spectrum $v L_v = \text{constant}$. For any other spectrum the differential flux S_v is related to the differential luminosity L_v by

$$S_v = (1+z) \frac{L_{(1+z)v}}{L_v} \frac{L_v}{4\pi d_L^2} \quad (\text{A.20})$$

where z is the redshift, the ratio of luminosities equalizes the difference in flux between the observed and emitted bands, and the factor of $(1+z)$ accounts for the redshifting of the bandwidth. Similarly, for differential flux per unit wavelength,

$$S_\lambda = \frac{1}{(1+z)} \frac{L_{\lambda/(1+z)}}{L_\lambda} \frac{L_\lambda}{4\pi d_L^2}. \quad (\text{A.21})$$

Given the galaxy absolute magnitude M_f estimated for a given filter using equations above, we can compute galaxy luminosity in units of solar luminosity in the same filter as

$$L_f = 10^{0.4(M_{\odot,f} - M_f)}, \quad (\text{A.22})$$

where $M_{\odot,f}$ is the absolute magnitude of the Sun in the same band (see, e.g., eq. 14 in Blanton et al. 2003 for the SDSS bands and here for many other commonly used bands).

A.5 Surface brightness

If a given patch in a galaxy at redshift z has the surface brightness S_b in some band b in units of $L_{\odot,b} \text{ pc}^{-2}$ (where $L_{\odot,b}$ is Sun's luminosity in the same band) in the small angle approximation the luminosity of the patch corresponding to the solid angle of 1 arcsec 2 will be $L_1 = S_b \Omega_1 D_A^2(z)$, where $\Omega_1 = 2.35044 \times 10^{-11}$ is the solid angle in steradians corresponding to one square arcsecond.

The absolute magnitude of the galaxy will be

$$M_b = M_{\odot,b} - 2.5 \log_{10} \frac{L_1}{L_{\odot,b}} = M_{\odot,b} - 2.5 \log_{10} S_b - 2.5 \log_{10} \Omega_1 - 5 \log_{10} D_A(z) \quad (\text{A.23})$$

The apparent magnitude of such square arcsecond patch will then be

$$\mu_b = M_b + 5 \log_{10} D_L - 5 = M_{\odot,b} - 2.5 \log_{10} S_b - 2.5 \log_{10} \Omega_1 - 5 + 5 \log_{10} (D_L/D_A) \quad (\text{A.24})$$

where $D_L = D_A(1+z)^2$ is luminosity distance. Substituting this and the value of Ω_1 into the equation above gives expression for surface brightness in the same band b in magnitudes per arcsecond square:

$$\mu_b = 21.5721 + M_{\odot,b} - 2.5 \log_{10} S_b + 10 \log_{10}(1+z), \quad (\text{A.25})$$

where the last term represents the $(1+z)^4$ cosmological dimming. To use this equation we need to know Sun's absolute magnitude in different bands. You can find a useful compilation of these [here](#).

A.6 Comoving volume

The *comoving volume* V_C is the volume measure in which number densities of non-evolving objects locked into Hubble flow are constant with redshift. It is the proper volume times three factors of the relative scale factor now to then, or $(1+z)^3$.

Given that the derivative of comoving distance with redshift is $1/E(z)$, the angular diameter distance converts a solid angle $d\Omega$ into a proper area, and two factors of $(1+z)$ convert a proper area into a comoving area, the comoving volume element in solid angle $d\Omega$ and redshift interval dz is

$$dV_C = D_H \frac{(1+z)^2 d_A^2}{E(z)} d\Omega dz \quad (\text{A.26})$$

The integral of the comoving volume element from the present to redshift z gives the total comoving volume, all-sky, out to redshift z

$$V_C = \begin{cases} \left(\frac{4\pi D_H^3}{2\Omega_k}\right) \left[\frac{d_M}{D_H} \sqrt{1 + \Omega_k \frac{d_M^2}{D_H^2}} - \frac{1}{\sqrt{|\Omega_k|}} \operatorname{arcsinh} \left(\sqrt{|\Omega_k|} \frac{d_M}{D_H} \right) \right] & \text{for } \Omega_k > 0 \\ \frac{4\pi}{3} d_M^3 & \text{for } \Omega_k = 0 \\ \left(\frac{4\pi D_H^3}{2\Omega_k}\right) \left[\frac{d_M}{D_H} \sqrt{1 + \Omega_k \frac{d_M^2}{D_H^2}} - \frac{1}{\sqrt{|\Omega_k|}} \operatorname{arcsin} \left(\sqrt{|\Omega_k|} \frac{d_M}{D_H} \right) \right] & \text{for } \Omega_k < 0 \end{cases} \quad (\text{A.27})$$

Carroll, Press, and Turner (1992). The comoving volume element and its integral are both used frequently in predicting number counts or luminosity densities.

A.7 Lookback time

The *lookback time* t_L to an object is the difference between the age t_0 of the Universe now (at observation) and the age t_e of the Universe at the time the photons were emitted (according to the object). It is used to predict properties of high-redshift objects with evolutionary models, such as passive stellar evolution for galaxies. Recall that $E(z)$ is the time derivative of the logarithm of the scale factor $a(t)$; the scale factor is proportional to $(1+z)$, so the product $(1+z)E(z)$ is proportional to the derivative of z with respect to the lookback time, or

$$t_L = t_H \int_0^z \frac{dz'}{(1+z')E(z')} \quad (\text{A.28})$$

Peebles (e.g., 1993, pp 313-315) and Kolb and Turner (1990, pp 52-56) give some analytic solutions to this equation, but they are concerned with the age $t(z)$, so they integrate from z to ∞ .

B. Supplementary material

B.1 Equilibrium equations

For a collisional fluid, the assumption of hydrostatic equilibrium results in the familiar hydrostatic equilibrium equation (HE hereafter) equation, in which the pressure gradient $\nabla p(\mathbf{x})$ at a point \mathbf{x} is balanced by the gradient of the local gravitational potential $\nabla\phi(\mathbf{x})$: $\nabla\phi(\mathbf{x}) = -\nabla p(\mathbf{x})/\rho_g(\mathbf{x})$, where $\rho_g(\mathbf{x})$ is the gas density. Under the further assumption of spherical symmetry, the HE equation can be written in the familiar form as $\rho_g^{-1}dp/dr = -GM(< r)/r^2$, where $M(< r)$ is the mass contained within the radius r . Assuming the equation of state of ideal gas, $p = \rho_g k_B T / \mu m_p$ where μ is the mean molecular weight and m_p is the proton mass, mass within r can be expressed in terms of the density and temperature profiles, $\rho_g(r)$ and $T(r)$, as

$$M_{\text{HE}}(< r) = -\frac{rk_B T(r)}{G\mu m_p} \left[\frac{d \ln \rho_g(r)}{d \ln r} + \frac{d \ln T(r)}{d \ln r} \right]. \quad (\text{B.1})$$

This equation governs distribution of plasma in stars or in galaxy groups and clusters.

Note that expressed in this way the hydrostatic equilibrium equation explicitly states that profiles of mass, fluid density, and its temperature are not independent in equilibrium but are related in a way prescribed by the equation. Another thing to note is that we can write relation between $M_{\text{HE}}(< r)$ and $T(r)$ for any r . The coefficient between $M(< r)$ and $rT(r)$ is set by the slopes of the gas density and temperature profiles.

For a collisionless system of particles, such as stars in a galaxy, the condition of equilibrium is given by the Jeans equation similar to the above HSE equation (e.g., Binney and Tremaine 2008). For a non-rotating spherically symmetric system, this equation can be written as

$$M_J(< r) = -\frac{r\sigma_r^2}{G} \left[\frac{d \ln n(r)}{d \ln r} + \frac{d \ln \sigma_r(r)^2}{d \ln r} + 2\beta(r) \right], \quad (\text{B.2})$$

where $n(r)$ is the radial number density profile of particles in the system, σ_r is rms velocity of particles in the radial direction, while $\beta = 1 - \frac{\sigma_t^2}{2\sigma_r^2}$ is the orbit anisotropy parameter defined in terms of the radial (σ_r) and tangential (σ_t) velocity dispersion components ($\beta = 0$ for isotropic velocity field). Note that particle velocity dispersion σ_r^2 in eq. B.2 plays the role of temperature in eq. B.1, while $n(r)$ is equivalent to $\rho_g(r)$. Equation B.2 tells us how radial distribution of stars and their velocities in spheroidal galaxies, in which rotation can be neglected, is related to the underlying mass distribution. This equation is also often used to

describe the equilibrium of cluster galaxies. Although, in principle, galaxies in groups and clusters are not strictly collisionless, interactions between galaxies are relatively rare and the Jeans equation should be quite accurate.

Equation of this kind (albeit more complicated) can be also derived for systems with significant rotation (see Nelson et al. 2014). Integration of the HSE and Jeans equations above over the entire volume of a system will give the global virial equilibrium relations. The HSE and global virial equations differ in details, but the scaling relation between M , σ (or T) and r is the same because it is set by dimensions of these quantities:

$$\frac{GM(< r)}{r} \propto \sigma_r^2 \quad (\text{B.3})$$

The role of equilibrium is to set the dimensionless factor in this scaling relation, which is usually a weak function of radius.

B.2 Power spectrum definition

Statistical properties of a uniform and isotropic Gaussian field can be fully characterized by its power spectrum, $P(k)$, which depends only on the modulus k of the wavevector, but not on its direction. The power spectrum is defined via the variance of the density field:

$$\langle \delta^2 \rangle = \frac{1}{(2\pi)^3} \int d^3 k P(k) = \frac{1}{2\pi^2} \int P(k) k^2 dk = \frac{1}{2\pi^2} \int P(k) k^3 d \ln k = \int \Delta^2(k) d \ln k, \quad (\text{B.4})$$

where $\Delta^2(k)$ is dimensionless power per logarithmic decade of k :

$$\Delta^2(k) = \frac{4\pi k^3}{(2\pi)^3} P(k) = \frac{1}{2\pi^2} k^3 P(k). \quad (\text{B.5})$$

Note that $\langle \delta^2 \rangle$ is dimensionless, so that $k^3 P(k)$ is dimensionless, i.e., $P(k)$ has units of volume.

Alternatively, one can define the power spectrum $P(k)$ as the average of the square amplitude of the Fourier components of perturbations with wavevector amplitude k (only the amplitude matters because we assume isotropic universe):

$$P(k) \equiv \langle \delta_{\mathbf{k}} \delta_{\mathbf{k}}^* \rangle = \delta_{\mathbf{k}} \delta_{-\mathbf{k}} \quad (\text{B.6})$$

$$\begin{aligned} &= \frac{1}{V} \iint \langle \delta(\mathbf{r}_1) \delta^*(\mathbf{r}_2) \rangle e^{-i\mathbf{k}\mathbf{r}_1} e^{i\mathbf{k}\mathbf{r}_2} d^3 \mathbf{r}_1 d^3 \mathbf{r}_2 \\ &= \int \xi(r_{12}) e^{-i\mathbf{k}\mathbf{r}_{12}} d^3 \mathbf{r}_{12} = \int \xi(r) e^{-i\mathbf{k}\mathbf{r}} d^3 \mathbf{r} \\ &= \int_0^\infty \xi(r) \left(\int_0^{2\pi} \int_0^\pi e^{-ikr \cos \theta} \sin \theta d\theta d\phi \right) r^2 dr \\ &= 2\pi \int_0^\infty \xi(r) \left(\int_1^{-1} e^{-ikru} [-du] \right) r^2 dr = \frac{4\pi}{k} \int_0^\infty \sin(kr) \xi(r) r dr. \end{aligned} \quad (\text{B.7})$$

where average is computed over some shell $[k, k + dk]$ in the Fourier space, and is normalized by the volume of the shell. Note that the units of $P(k)$ is volume (comoving), although expression for $P(k)$ does not explicitly depend on volume – i.e., measurement of $P(k)$ does not depend on specific choice of volume.

A related quantity is the variance of the density contrast field *smoothed* on some scale R : $\delta_R(\mathbf{x}) \equiv \int \delta(\mathbf{x} - \mathbf{r}) W(\mathbf{r}, R) d^3 r$, where

$$\langle \delta_R^2 \rangle \equiv \sigma^2(R) = \frac{1}{(2\pi)^3} \int P(k) |\tilde{W}(\mathbf{k}, R)|^2 d^3 k, \quad (\text{B.8})$$

where $\tilde{W}(\mathbf{k}, R)$ is the Fourier transform of the window (filter) function $W(\mathbf{r}, R)$, such that $\delta_R(\mathbf{k}) = \delta(\mathbf{k})\tilde{W}(\mathbf{k}, R)$. The most commonly used window functions are the top-hat in real space and Gaussian, which in the Fourier space have the following equivalents:

$$\tilde{W}_{\text{TH}}(x) = \frac{3(\sin x - x \cos x)}{x^3}, \quad \text{where } x = kR, \quad (\text{B.9})$$

$$\tilde{W}_{\text{G}}(x) = e^{-\frac{x^2}{2}}, \quad \text{where } x = kR. \quad (\text{B.10})$$

For the cases, when one is interested in only a narrow range of k the power spectrum can be approximated by the power-law form, $P(k) \propto k^n$, and the variance is $\sigma^2(R) \propto R^{-(n+3)}$.

At a sufficiently high redshift z , for the spherical top-hat window function mass and radius are interchangeable according to the relation $M = 4\pi/3\rho_m(z)R^3$. We can think about the density field smoothed on the scale R or the corresponding mass scale M . The characteristic amplitude of peaks in the δ_R (or δ_M) field smoothed on scale R (or mass scale M) is given by $\sigma(R) \equiv \sigma(M)$. The smoothed Gaussian density field is, of course, also Gaussian with the probability distribution function (PDF) given by

$$p(\delta_M) = \frac{1}{\sqrt{2\pi}\sigma(M)} \exp\left[-\frac{\delta_M^2}{2\sigma^2(M)}\right]. \quad (\text{B.11})$$

B.3 Gaussian peaks

The statistics of the peaks in the Gaussian random field is well characterized (e.g., Bardeen et al. 1986).

Profiles of isolated peaks in a Gaussian field conditioned to collapse by a given time can be computed using conditional pdf $p(\delta(< r)|\delta_{\text{TH}}, \delta'_{\text{TH}})$, in which probability of a given δ is conditioned by the value of δ_{TH} for $r = R_{\text{TH}}$ and by the derivative $\delta'_{\text{TH}} = d\delta/dr|_{r=R_{\text{TH}}}$. The first condition is to ensure that we have a peak that will actually collapse by a given epoch. The derivative condition is needed to ensure that we have an isolated peak. For example, $\delta' > 0$ would indicate that the peak is local and is a part of a larger peak that would collapse at an earlier epoch.

The conditional probability of a conditional variable $p(x|y)$ is also Gaussian, but with the mean and variance given by (proof is discussed [here](#)):

$$\mu_{x|y} = \langle x \rangle + \langle xy \rangle (y - \langle y \rangle) \langle yy \rangle^{-1} \quad (\text{B.12})$$

$$\sigma_{x|y}^2 = \langle xx \rangle - \langle xy \rangle \langle yy \rangle^{-1} \langle yx \rangle. \quad (\text{B.13})$$

B.4 Linear perturbation evolution equation

Overall evolution of perturbations from then onwards can be understood in the context of the familiar damped harmonic oscillator equation:

$$\ddot{x} + \frac{b}{m}\dot{x} + \frac{k}{m}x = 0. \quad (\text{B.14})$$

The above equation has [solutions](#) in the form of

$$x(t) \propto \exp(-bt/2m) \exp(\pm i\omega t), \quad (\text{B.15})$$

where $\omega = \sqrt{b^2 - 4mk}/2m$. When $\omega > 0$ and b/m is small these are just oscillations of constant amplitude; if b/m is of order $1/\omega$ the oscillations have decreasing amplitude, while if $\gamma \gg 1/\omega$ there will simply be a decay of amplitude without oscillations.

If ω is complex then there will be no oscillations, but rather two growing and decaying modes in the solution or just a decaying mode.

These features of the solutions of the damped harmonic oscillator should be kept in mind when considering evolution of density perturbations in the universe, which are governed by such equation. This equation can be derived in the Newtonian limit by combining continuity equation (mass conservation) with the Euler equation (momentum conservation), the Poisson equation ($\nabla^2 \phi = 4\pi G\rho$), and equation of state which specifies the pressure forces and sound speed and linearizing the derived equation by considering density perturbations to be very small $\delta \equiv \rho/\bar{\rho} - 1 \ll 1$ (see, e.g., S 4.1 in the MvdBW book):

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = \left[4\pi G\bar{\rho}\delta - \frac{c_s^2}{a^2}\nabla^2\delta\right] - \frac{2}{3}\frac{\bar{T}}{a^2}\nabla^2S. \quad (\text{B.16})$$

or, in the Fourier space (where spatial gradients become $\nabla \rightarrow ik$ and $\nabla^2 \rightarrow -k^2$):

$$\ddot{\delta}_k + 2\frac{\dot{a}}{a}\dot{\delta}_k = \left[4\pi G\bar{\rho} - \frac{c_s^2}{a^2}k^2\right]\delta - \frac{2}{3}\frac{\bar{T}}{a^2}k^2S_k. \quad (\text{B.17})$$

Here c_s is the sound speed of plasma, $\bar{\rho}$ is the mean density of the universe at epoch a and S is plasma entropy. Although Newtonian limit does not strictly apply to some of the situations we will consider, such as evolution of super-horizon perturbations and early stages of evolution, it will give qualitatively correct results in all cases. Moreover, the form of the Newtonian limit equations is mathematically identical to the equations derived using GR (see note below).

The equations above show that expansion of the universe acts as a damping term ($2\delta\dot{a}/a$) in the evolution of perturbation, while gravity and entropy perturbations source fluctuation growth. Gravity is reduced by effective pressure of the fluid. If we make the standard assumption that perturbations are adiabatic (fluctuations in all components are proportional to each other), $S_k = 0$, the above equations are exact analogues of the damped oscillator equation. We will assume adiabatic perturbations below. This is consistent with data, although non-adiabatic, "isocurvature" fluctuations are also being explored in the literature.

Let's briefly consider some important solutions to this equation in several regimes.

The following equation describes 1) evolution of super-horizon perturbations during inflation, and 2) evolution of weakly interacting dark matter after inflation both on superhorizon and sub-horizon scales.

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = 4\pi G\bar{\rho}\delta, \quad (\text{B.18})$$

During inflation, the average density is reduced by exponential expansion by a factor of $\sim \exp(-3N_{\text{e-f}}) \sim 10^{-80} \div 10^{-92}$ - i.e., to negligible levels, which allows us to neglect the term on the right hand side. In this stage, $\dot{a}/a = H \approx \text{const}$ and the equation has solution $\delta \approx c_1 + c_2 \exp(-2Ht)$, where c_i are constants. Thus, amplitude of superhorizon perturbations of the growing mode is preserved at this stage.

Radiation domination era. Once, the universe is reheated and is dominated by energy, expansion is considerably slower, $a(t) = t^{1/2}$ which corresponds to $\dot{a}/a = 1/(2t)$, and the right hand side of the perturbation equation cannot be neglected (and needs to account for the relativistic component). At the *super-horizon scales* all components evolve together because pressure forces cannot affect the baryon-photon plasma. Perturbation equation in this case has the form:

$$\ddot{\delta} + 2H\dot{\delta} = 4H^2\delta$$

or

$$\ddot{\delta} + \frac{1}{t}\dot{\delta} = \delta/t^2.$$

Seeking solution in the form $\delta \propto t^n$ gives $n(n-1) + n = 1$ and solution $n = \pm 1$. The growing mode in this case is thus $\delta \propto t \propto a^2$. Note that evolution of density perturbations in this regime depends on the choice of gauge.

At the sub-horizon scales, the baryon-photon plasma will oscillate without collapsing ($\omega^2 > 0$) because pressure forces are extremely large (even right before recombination the Jeans scale is of order of the Hubble radius). These are the so-called *baryonic acoustic oscillations* (BAO).

Non-baryonic dark matter, however, will not oscillate because it is not affected by pressure forces. Evolution of its perturbations is governed by the equation:

$$\ddot{\delta}_{\text{dm}} + \frac{1}{t} \dot{\delta}_{\text{dm}} = 4\pi G \bar{\rho}_{\text{dm}} \delta = \frac{3}{2} H^2 \Omega_m \delta_{\text{dm}} = \frac{3}{8t^2} \Omega_{\text{dm}} \delta_{\text{dm}}.$$

Note that the mean density in the r.h.s. of the perturbation equation is actually the mean dark matter density, $\bar{\rho}_{\text{dm}} = \Omega_{\text{dm}} \bar{\rho}$). Deep in the radiation domination regime $\Omega_{\text{dm}} \ll 1$ and the equation $\ddot{\delta}_{\text{dm}} + \frac{1}{t} \dot{\delta}_{\text{dm}} \approx 0$ has solution

$$\delta(t) \approx s_1 + s_2 \ln(t).$$

That is the amplitude of perturbations is constant to zeroth order and can grow only logarithmically overall.

Matter domination era. During the matter domination era expansion factor grows as $a(t) = t^{2/3}$ so that $H(t) = \dot{a}/a = 2/(3t)$ and the perturbation equation for *dark matter* perturbations is ($\bar{\rho} = 3H^2(t)/8\pi G$):

$$\ddot{\delta} + 2 \frac{\dot{a}}{a} \dot{\delta} = 4\pi G \bar{\rho} \delta = \frac{3}{2} \left(\frac{\dot{a}}{a} \right)^2 \delta,$$

or

$$\ddot{\delta} + \frac{4}{3t} \dot{\delta} = \frac{2}{3t^2} \delta,$$

searching for the solution in the form $\delta \propto t^n$ gives $n(n-1) + 4n/3 - 2/3 = 0$, which has solutions $n = 2/3$ and $n = -1$, corresponding to the growing and decaying modes:

$$\delta_+ \propto t^{2/3} \propto a, \quad (\text{B.19})$$

$$\delta_- \propto t^{-1} \propto a^{-3/2}. \quad (\text{B.20})$$

This solution also applies to the total matter perturbations at later epochs after recombination when baryons fully catch up with the dark matter perturbations in their growth.

Note that in this regime as density perturbations grow as $\propto a$, so that potential perturbations $\delta_{\Phi,k} \propto k^{-2} \bar{\rho} \Delta \propto a^2 a^{-3} a = \text{const}$ do not grow. Potential only starts to change at late epochs in $\Omega_\Lambda > 0$ cosmologies.

The full solution through a_{eq} is (Heath 1977; and sections 10-13 in Peebles 1980):

$$\delta(a) \propto D_+(a) = \frac{5\Omega_m}{2} E(a) \int_0^a \frac{da'}{[a'E(a')]^3}, \quad (\text{B.21})$$

where $E(a)$ is the normalized expansion rate, a' is a dummy integration variable. If we consider only matter, cosmological constant and radiation as components driving the expansion of the universe, the above equation has solution $\delta(a) = \delta_i D_+(a)/D_+(a_i)$, where the linear growth factor $D_+(a)$ is given by:

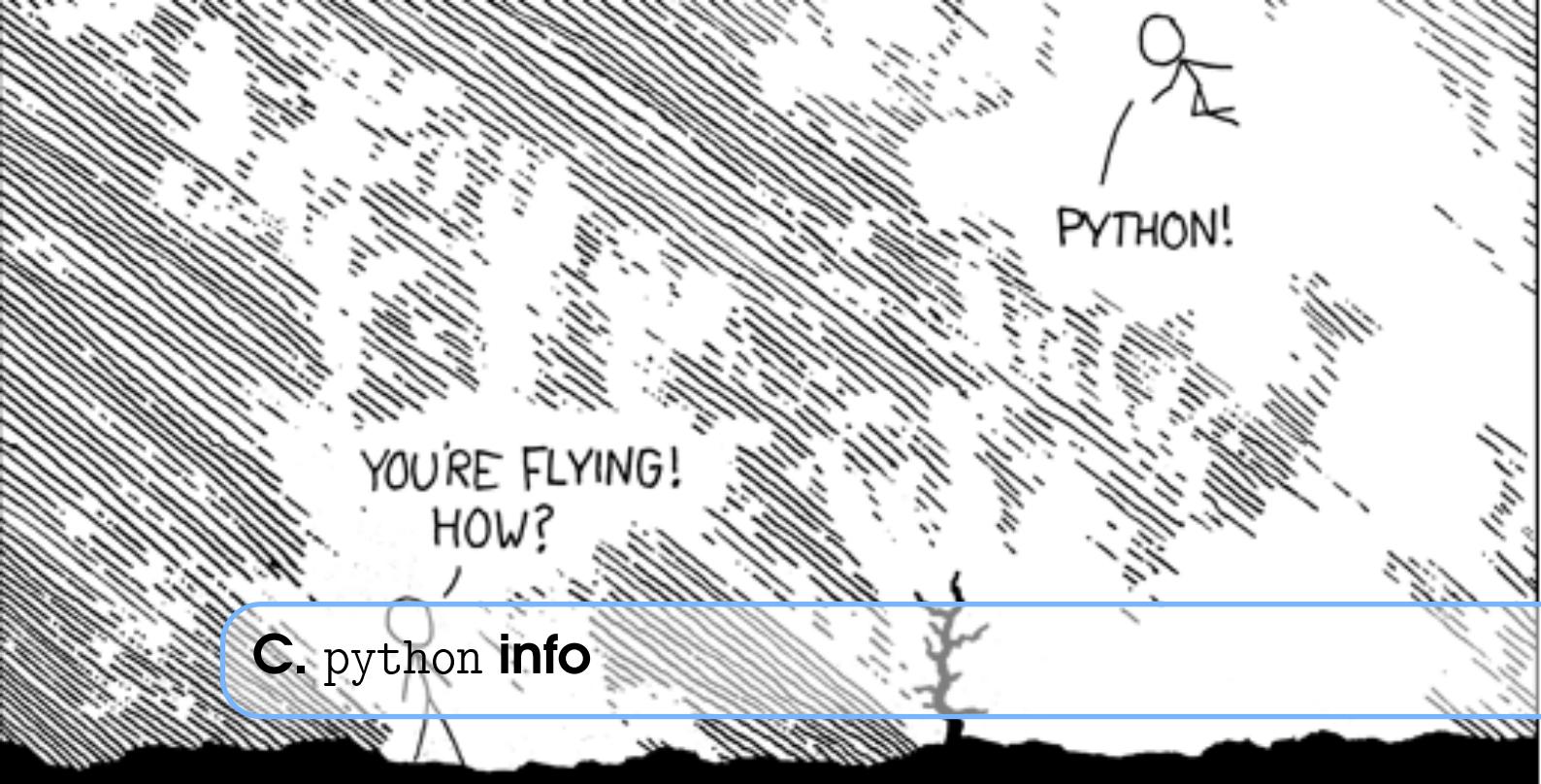
$$D_+(a) = D_{+,m} + \frac{2}{3} a_{\text{eq}} + \frac{a_{\text{eq}}}{2 \ln(2) - 3} \left[2\sqrt{1+x} + \left(\frac{2}{3} + x \right) \ln \frac{\sqrt{1+x} - 1}{\sqrt{1+x} + 1} \right],$$

where $a_{\text{eq}} = \Omega_r/\Omega_m$ is the expansion factor of the matter-radiation equality and $x = a/a_{\text{eq}}$. $D_{+,m}$ is the growth factor in the matter-dominated regime neglecting radiation, when $E(a)$ can be approximated by

$$E(a) \equiv \frac{H(a)}{H_0} = [\Omega_m a^{-3} + (1 - \Omega_m - \Omega_\Lambda) a^{-2} + \Omega_\Lambda]^{1/2},$$

and $D_{+,m}(a)$ can be obtained either by direct integration of the above integral or via approximate (but accurate to better than a couple per cent) formula given by eq. 29 in Carroll, Press, and Turner (1992). Expressions for the growth factor in the uniform dark energy models are somewhat different and can be found in Percival (2005).

Remark on the validity of the Newtonian approximation. Recall that perturbation evolution equation above was derived in the Newtonian limit. One may question the validity of the Newtonian approximation on large scales. However, by coincidence the above equation is identical to the corresponding fully relativistic linear perturbation growth equation in the synchronous gauge or, more generally, in the comoving “total matter gauge.” The equations governing evolution of perturbations in the synchronous gauge are presented, for example, in Ma and Bertschinger (1995). We can derive equation identical to Equation above by (1) substituting their Equation 21(a) into 21(c), (2) considering the case of universe with energy density dominated by non-relativistic matter ($\delta T_0^0 = \bar{\rho} \delta$, $\delta T_i^i = 0$), for which $dh/d\tau = -2\delta$ (their Equation (42)), and (3) by making appropriate transformations from conformal time τ to physical time t using $d\tau = dt/a(\tau)$. This means that linear evolution of perturbations for non-relativistic matter component is treated correctly in the Newtonian approximation of cosmological codes at any scale, provided one interprets results of simulations in the appropriate gauge (see, e.g., Wands and Slosar 2009, Chisari and Zaldarriaga 2011). This is actually why the Newtonian approximation is still ok to use in N -body simulations when Gigaparsec scale volumes are simulated.



C. python info

C.1 Ipython notebooks

I will be using IPython Notebooks to supplement whatever I show in class. These can be viewed, edited, and produced with IDE editors of the main python distros (Enthought or Anaconda).¹ Also, I recommend the PyCharm IDE environment² for code development, which can be obtained free for academic users.

If you don't have a python distribution installed on your machine, please do so asap. The rationale for using Notebooks can be found [here](#) - it simplifies work flow and closely connects plots that we will be discussing with the code that produces them. You can use these codes as a starting point for your own experimentation with data or calculations.

To run examples presented here, you will need to install python (Enthought or Anaconda distributions) with standard libraries (matplotlib, numpy, scipy, etc). In addition, make sure you have pyfits installed for reading FITS files. The FITS file with the SDSS data will be distributed (also I will distribute SQL script that was used to query SDSS DR8 to produce it).

C.2 colossus

For cosmological functions, such as distances, we will use the `colossus` package written by Benedikt Diemer. We will also use it when we explore properties of dark matter halo profiles. You can install it by

```
pip install https://bitbucket.org/bdiemer/colossus/get/tip.tar.gz
or
easy_install https://bitbucket.org/bdiemer/colossus/get/tip.tar.gz
```

C.3 AstroML

We will also use some functions and examples from the AstroML python library, especially when we will come to more sophisticated analyses, such as clustering. This library was developed to support the book

¹Note that you can request a free Enthought academic license which will allow you to get distribution with many add-ons and libraries

²<https://www.jetbrains.com/pycharm/>

”Statistics, Data Mining, and Machine Learning in Astronomy” (Ivezić et al., 2013). The book itself is not needed for this course, although I will draw on it in parts of the course. It is very good though and I highly recommend to get and study it (it is not available in our library yet, unfortunately, but many of your fellow graduate students already have it). In any case, all of the figures from the book and the python code used to produce them is available online here.



Context

(optional) Task Name

DR8

My Tasks | My Query

Samples Recent

```

1 |SELECT
2 |    S.ra, G.dec, S.mjd, S.plate, S.fiberID,
3 |    S.z, S.zErr, S.rChi2, S.velDisp, S.velDispErr,
4 |    G.extinction_r, G.extinction_g, G.petroMag_r, G.psfMag_r, G.psfMagErr_r,
5 |    G.modelMag_u, G.modelMagErr_u, G.modelMag_g, G.modelMagErr_g,
6 |    G.modelMag_r, G.modelMagErr_r, G.modelMag_i, G.modelMagErr_i,
7 |    G.modelMag_z, G.modelMagErr_z, G.cmodelMag_u, G.cmodelMagErr_u, G.cmodelMag_g, G.cmodelMagErr_g,
8 |    G.cmodelMag_r, G.cmodelMagErr_r, G.cmodelMag_i, G.cmodelMagErr_i,
9 |    G.petroR50_r, G.petroR90_r, G.petroR50Err_r, G.petroR90Err_r, G.fracdeV_r, G.expRad_r, G.expRadErr_r,
10 |   G.deVRad_r, G.deVRadErr_r, G.devMag_r, G.devMagErr_r, G.expMag_r, G.expMagErr_r, G.lnLDeV_r, G.lnLExp_r,
11 |   G.devAB_r, G.devABErr_r, G.expAB_r, G.expABErr_r,
12 |   GSL.h_alpha_flux, GSL.h_alpha_flux_err, GSL.oiii_5007_flux, GSL.oiii_5007_flux_err,
13 |   GSX.d4000, GSX.d4000_err, GSE.bptClass,
14 |   GSE.lgm_tot_p50, GSE.sfr_tot_p50, G.objID, GSI.specObjID
15 |INTO myDR8 SELECT * FROM db.NearbyObjQ(S.ra, S.dec, 0.01) N, Galaxy G,
16 |      GalSpecInfo GSI, GalSpecLine GSL, GalSpecIndx GSX, GalSpecExtra GSE
17 |WHERE N.objID = G.objID
18 |

```

D. Data access info

D.1 SDSS DR8 main galaxy sample

For many exercises here and subsequent lectures we will be using SDSS data. The binary FITS file used in the explorations below can be downloaded here (it is 165 Mb). It was produced at the SDSS CasJobs server where time-intensive SQL queries can be submitted. The FITS file used below is large because it includes a number of properties that will be useful in our explorations and because it selects almost all low- z galaxies from the SDSS (called the main galaxy sample, to differentiate from the quasar and LRG samples).

The SQL script used to produce the FITS file below can be found [here](#). Description of various entries for SDSS objects classified as GALAXY in DR8 can be found [here](#), for STAR objects see [here](#), while QSOs are [here](#). If you have not queried SDSS data base yet, I encourage you to use this example, to create your own queries for particular properties.

D.1.1 Petrosian magnitudes and sizes

In §1.4.1 we discussed the Petrosian (1976) definition of the galaxy magnitude and size and the specific implementation of this definition in the SDSS (see [here](#) for more details) computes the following function as a function of angular radius, R , from galaxy center:

$$\eta(R) \equiv \frac{\int_{0.8R}^{1.25R} dR' 2\pi R' \Sigma(R') / [\pi(1.25^2 - 0.8^2)R^2]}{\int_0^R 2\pi R' \Sigma(R') dR' / (\pi R^2)} \quad (\text{D.1})$$

where $\Sigma(r)$ is the surface brightness profile. The *Petrosian radius*, R_P , is then defined by the SDSS pipeline as the radius where $\eta(R_P) = 0.2$. Galaxy flux is then measured within some multiple of R_P :

$$F_P \equiv \int_0^{N_P R_P} 2\pi R' I(R') dR' \quad (\text{D.2})$$

The aperture $2R_P$ used in the SDSS measurements.

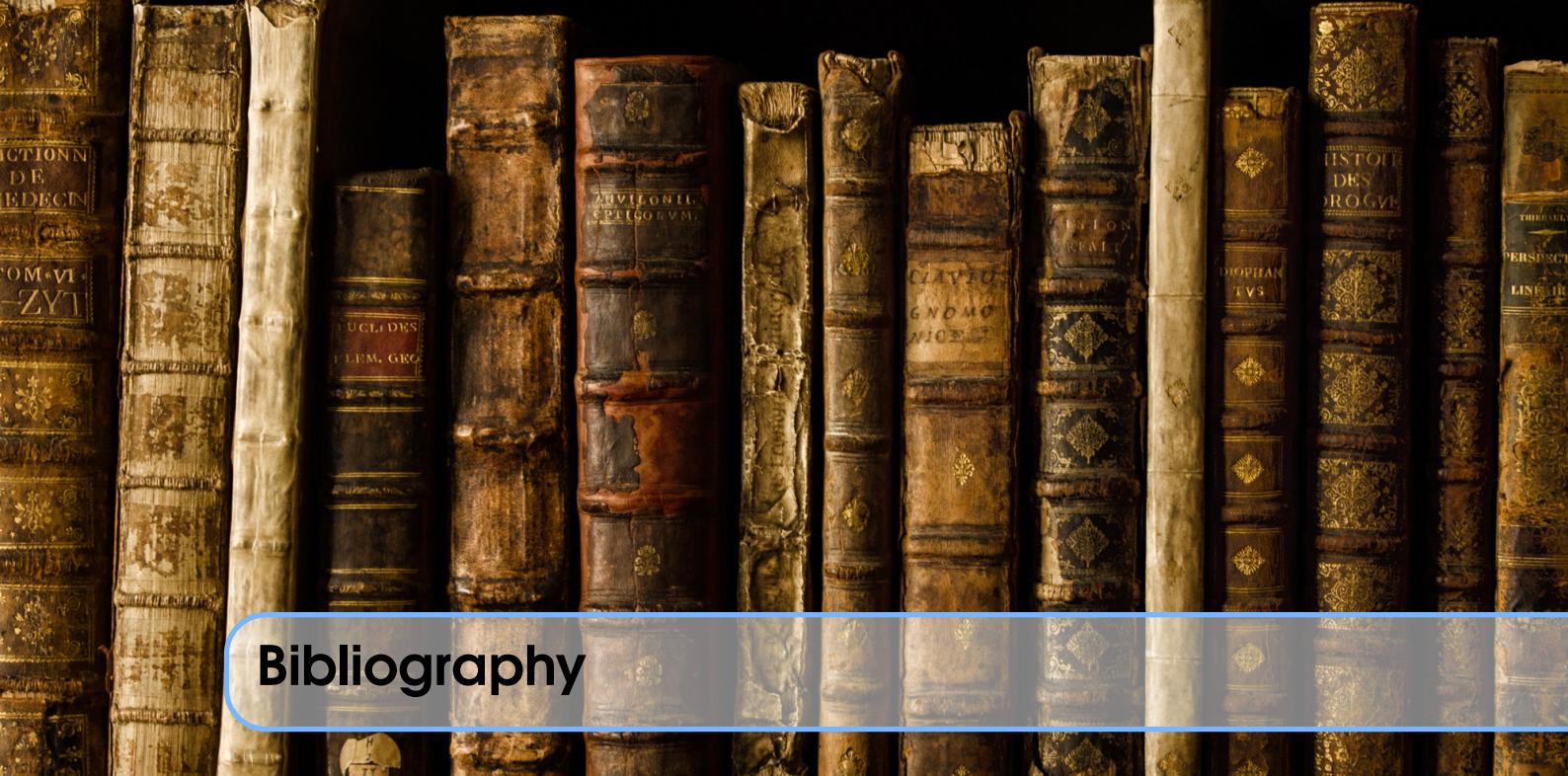
The choices for η and N_P are heuristic. It is argued to be large enough to contain nearly all of the flux for many galaxies (in particular late type galaxies described by the exponential profile), but small enough that the sky noise is sub-dominant in F_P . In this case, even substantial errors in R_P cause only small errors in the Petrosian flux (typical statistical errors near the spectroscopic flux limit of $r \sim 17.7$ are $< 5\%$). The main

draw of the Petrosian's definition, however, is that the fraction of recovered light is robust and depend only weakly on the galaxy axis ratio or size variation due to worse seeing or greater distance (e.g., Blanton et al., 2001).

The Petrosian radius in each band is the parameter `petroRad` in the database with the subscript corresponding to particular filter (e.g., for *r*-band, `petroRad_r`) and the Petrosian magnitude in each band (calculated using only `petroRad` for the *r* band) is the parameter `petroMag` (e.g., for *r*-band, `petroMag_r`).

SDSS main galaxy sample also provides radii enclosing 50% and 90% of the total light of the Petrosian magnitude (e.g., `petroR50_r` and `petroR90_r` for the *r* band).

D.1.2 `cmodel magnitudes`



Bibliography

- Adhikari, S., N. Dalal, and R. T. Chamberlain (2014). “Splashback in accreting dark matter halos”. In: *JCAP* 11, 019, page 019. DOI: 10.1088/1475-7516/2014/11/019. arXiv: 1409.4482.
- Applegate, D. E. et al. (2014). “Weighing the Giants - III. Methods and measurements of accurate galaxy cluster weak-lensing masses”. In: *MNRAS* 439, pages 48–72. DOI: 10.1093/mnras/stt2129. arXiv: 1208.0605.
- Applegate, D. E. et al. (2016). “Cosmology and astrophysics from relaxed galaxy clusters - IV. Robustly calibrating hydrostatic masses with weak lensing”. In: *MNRAS* 457, pages 1522–1534. DOI: 10.1093/mnras/stw005. arXiv: 1509.02162.
- Audit, E., R. Teyssier, and J.-M. Alimi (1997). “Non-linear dynamics and mass function of cosmic structures. I. Analytical results.” In: *A&A* 325, pages 439–449. eprint: astro-ph/9704023.
- Babcock, H. W. (1939). “The rotation of the Andromeda Nebula”. In: *Lick Observatory Bulletin* 19, pages 41–51. DOI: 10.5479/ADS/bib/1939LicOB.19.41B.
- Bailin, J. and W. E. Harris (2008). “Inclination-Independent Galaxy Classification”. In: *ApJ* 681, 225–231, pages 225–231. DOI: 10.1086/588091. arXiv: 0803.1274.
- Baldry, I. K. et al. (2012). “Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function at $z < 0.06$ ”. In: *MNRAS* 421, pages 621–634. DOI: 10.1111/j.1365-2966.2012.20340.x. arXiv: 1111.5707.
- Bardeen, J. M., P. J. Steinhardt, and M. S. Turner (1983). “Spontaneous creation of almost scale-free density perturbations in an inflationary universe”. In: *Phys. Rev. D* 28, pages 679–693. DOI: 10.1103/PhysRevD.28.679.
- Bardeen, J. M. et al. (1986). “The statistics of peaks of Gaussian random fields”. In: *ApJ* 304, pages 15–61. DOI: 10.1086/164143.
- Bell, E. F. et al. (2003). “The Optical and Near-Infrared Properties of Galaxies. I. Luminosity and Stellar Mass Functions”. In: *ApJS* 149, pages 289–312. DOI: 10.1086/378847. eprint: astro-ph/0302543.
- Bernardi, M. et al. (2010). “Galaxy luminosities, stellar masses, sizes, velocity dispersions as a function of morphological type”. In: *MNRAS* 404, pages 2087–2122. DOI: 10.1111/j.1365-2966.2010.16425.x. arXiv: 0910.1093.

- Bernardi, M. et al. (2013). “The massive end of the luminosity and stellar mass functions: dependence on the fit to the light profile”. In: *MNRAS* 436, pages 697–704. DOI: 10.1093/mnras/stt1607. arXiv: 1304.7778 [astro-ph.CO].
- Bertschinger, E. (1985). “Self-similar secondary infall and accretion in an Einstein-de Sitter universe”. In: *ApJS* 58, pages 39–65. DOI: 10.1086/191028.
- (1998). “Simulations of Structure Formation in the Universe”. In: *ARA&A* 36, pages 599–654. DOI: 10.1146/annurev.astro.36.1.599.
- Bhattacharjee, P., S. Chaudhury, and S. Kundu (2014). “Rotation Curve of the Milky Way out to ~200 kpc”. In: *ApJ* 785, 63, page 63. DOI: 10.1088/0004-637X/785/1/63. arXiv: 1310.2659.
- Binney, J. and S. Tremaine (2008). *Galactic Dynamics: Second Edition*. Princeton University Press.
- Bland-Hawthorn, J. and O. Gerhard (2016). “The Galaxy in Context: Structural, Kinematic and Integrated Properties”. In: *ARA&A, in press (arXiv/1602.07702)*. arXiv: 1602.07702.
- Blanton, M. R. and J. Moustakas (2009). “Physical Properties and Environments of Nearby Galaxies”. In: *ARA&A* 47, pages 159–210. DOI: 10.1146/annurev-astro-082708-101734. arXiv: 0908.3017.
- Blanton, M. R. and S. Roweis (2007). “K-Corrections and Filter Transformations in the Ultraviolet, Optical, and Near-Infrared”. In: *AJ* 133, pages 734–754. DOI: 10.1086/510127. eprint: astro-ph/0606170.
- Blanton, M. R. et al. (2001). “The Luminosity Function of Galaxies in SDSS Commissioning Data”. In: *AJ* 121, pages 2358–2380. DOI: 10.1086/320405. eprint: astro-ph/0012085.
- Blanton, M. R. et al. (2003). “The Galaxy Luminosity Function and Luminosity Density at Redshift $z = 0.1$ ”. In: *ApJ* 592, pages 819–838. DOI: 10.1086/375776. eprint: astro-ph/0210215.
- Blanton, M. R. et al. (2005). “The Properties and Luminosity Function of Extremely Low Luminosity Galaxies”. In: *ApJ* 631, pages 208–230. DOI: 10.1086/431416. eprint: astro-ph/0410164.
- Blumenthal, G. R., H. Pagels, and J. R. Primack (1982). “Galaxy formation by dissipationless particles heavier than neutrinos”. In: *Nature* 299, page 37. DOI: 10.1038/299037a0.
- Blumenthal, G. R. et al. (1984). “Formation of galaxies and large-scale structure with cold dark matter”. In: *Nature* 311, pages 517–525. DOI: 10.1038/311517a0.
- Bond, J. R., L. Kofman, and D. Pogosyan (1996). “How filaments of galaxies are woven into the cosmic web”. In: *Nature* 380, pages 603–606. DOI: 10.1038/380603a0. eprint: astro-ph/9512141.
- Bond, J. R. and S. T. Myers (1996). “The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms”. In: *ApJS* 103, page 1. DOI: 10.1086/192267.
- Bond, J. R., A. S. Szalay, and M. S. Turner (1982). “Formation of galaxies in a gravitino-dominated universe”. In: *Physical Review Letters* 48, pages 1636–1639. DOI: 10.1103/PhysRevLett.48.1636.
- Borgani, S. and A. Kravtsov (2011). “Cosmological Simulations of Galaxy Clusters”. In: *Advanced Science Letters* 4, pages 204–227. DOI: 10.1166/asl.2011.1209. arXiv: 0906.4370 [astro-ph.CO].
- Bothun, G., C. Impey, and S. McGaugh (1997). “Low-Surface-Brightness Galaxies: Hidden Galaxies Revealed”. In: *PASP* 109, pages 745–758. DOI: 10.1086/133941.
- Bouwens, R. J. et al. (2015). “UV Luminosity Functions at Redshifts $z=4$ to $z=10$: 10,000 Galaxies from HST Legacy Fields”. In: *ApJ* 803, 34, page 34. DOI: 10.1088/0004-637X/803/1/34. arXiv: 1403.4295.
- Bovy, J. and H.-W. Rix (2013). “A Direct Dynamical Measurement of the Milky Way’s Disk Surface Density Profile, Disk Scale Length, and Dark Matter Profile at $4 \text{ kpc} < R < 9 \text{ kpc}$ ”. In: *ApJ* 779, 115, page 115. DOI: 10.1088/0004-637X/779/2/115. arXiv: 1309.0809 [astro-ph.GA].
- Bovy, J. and S. Tremaine (2012). “On the Local Dark Matter Density”. In: *ApJ* 756, 89, page 89. DOI: 10.1088/0004-637X/756/1/89. arXiv: 1205.4033 [astro-ph.GA].
- Bryan, G. L. and M. L. Norman (1998). “Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons”. In: *ApJ* 495, pages 80–99. DOI: 10.1086/305262. eprint: astro-ph/9710107.
- Bullock, J. S. et al. (2001). “Profiles of dark haloes: evolution, scatter and environment”. In: *MNRAS* 321, pages 559–575. DOI: 10.1046/j.1365-8711.2001.04068.x. eprint: astro-ph/9908159.
- Busha, M. T. et al. (2011). “The Mass Distribution and Assembly of the Milky Way from the Properties of the Magellanic Clouds”. In: *ApJ* 743, 40, page 40. DOI: 10.1088/0004-637X/743/1/40. arXiv: 1011.2203.

- Buta, R. J. (2011). “Galaxy Morphology”. In: *in Planets, Stars, and Stellar Systems, Vol. 6, Series Editor T. D. Oswalt, Volume editor W. C. Keel, Springer Reference, 2011 (arXiv/1102.0550)*. arXiv: 1102.0550 [astro-ph.CO].
- Cardone, V. F., E. Piedipalumbo, and C. Tortora (2005). “Spherical galaxy models with power-law logarithmic slope”. In: *MNRAS* 358, pages 1325–1336. DOI: 10.1111/j.1365-2966.2005.08834.x. eprint: astro-ph/0501151.
- Carroll, S. M., W. H. Press, and E. L. Turner (1992). “The cosmological constant”. In: *ARA&A* 30, pages 499–542. DOI: 10.1146/annurev.aa.30.090192.002435.
- Cavaliere, A. G., H. Gursky, and W. H. Tucker (1971). “Extragalactic X-ray Sources and Associations of Galaxies”. In: *Nature* 231, pages 437–438. DOI: 10.1038/231437a0.
- Chilingarian, I. V., A.-L. Melchior, and I. Y. Zolotukhin (2010). “Analytical approximations of K-corrections in optical and near-infrared bands”. In: *MNRAS* 405, pages 1409–1420. DOI: 10.1111/j.1365-2966.2010.16506.x. arXiv: 1002.2360 [astro-ph.IM].
- Chilingarian, I. V. and I. Y. Zolotukhin (2012). “A universal ultraviolet-optical colour-colour-magnitude relation of galaxies”. In: *MNRAS* 419, pages 1727–1739. DOI: 10.1111/j.1365-2966.2011.19837.x. arXiv: 1102.1159.
- Chisari, N. E. and M. Zaldarriaga (2011). “Connection between Newtonian simulations and general relativity”. In: *Phys. Rev. D* 83.12, 123505, page 123505. DOI: 10.1103/PhysRevD.83.123505. arXiv: 1101.3555 [astro-ph.CO].
- Christodoulou, D. M., I. Shlosman, and J. E. Tohline (1995). “A new criterion for bar-forming instability in rapidly rotating gaseous and stellar systems. 1: Axisymmetric form”. In: *ApJ* 443, pages 551–562. DOI: 10.1086/175547. eprint: astro-ph/9502073.
- Cole, S. and C. Lacey (1996). “The structure of dark matter haloes in hierarchical clustering models”. In: *MNRAS* 281, page 716. DOI: 10.1093/mnras/281.2.716. eprint: astro-ph/9510147.
- Conroy, C. (2013). “Modeling the Panchromatic Spectral Energy Distributions of Galaxies”. In: *ARA&A* 51, pages 393–455. DOI: 10.1146/annurev-astro-082812-141017. arXiv: 1301.7095.
- Correa, C. A. et al. (2015). “The accretion history of dark matter haloes - II. The connections with the mass power spectrum and the density profile”. In: *MNRAS* 450, pages 1521–1537. DOI: 10.1093/mnras/stv697. arXiv: 1501.04382.
- Courteau, S. et al. (2014). “Galaxy masses”. In: *Reviews of Modern Physics* 86, pages 47–119. DOI: 10.1103/RevModPhys.86.47. arXiv: 1309.3276.
- Courtin, J. et al. (2011). “Imprints of dark energy on cosmic structure formation - II. Non-universality of the halo mass function”. In: *MNRAS* 410, pages 1911–1931. DOI: 10.1111/j.1365-2966.2010.17573.x. arXiv: 1001.3425.
- Cuesta, A. J. et al. (2008). “The virialized mass of dark matter haloes”. In: *MNRAS* 389, pages 385–397. DOI: 10.1111/j.1365-2966.2008.13590.x. arXiv: 0710.5520.
- da Costa, L. N. et al. (1994). “A complete southern sky redshift survey”. In: *ApJ* 424, pages L1–L4. DOI: 10.1086/187260.
- Dalal, N., Y. Lithwick, and M. Kuhlen (2010). “The Origin of Dark Matter Halo Profiles”. In: *arXiv/1010.2539*. arXiv: 1010.2539 [astro-ph.CO].
- Dalal, N. et al. (2008). “Halo Assembly Bias in Hierarchical Structure Formation”. In: *ApJ* 687, 12-21, pages 12–21. DOI: 10.1086/591512. arXiv: 0803.3453.
- de Haan, T. et al. (2016). “Cosmological Constraints from Galaxy Clusters in the 2500 square-degree SPT-SZ Survey”. In: *ApJ submitted (arXiv/1603.06522)*. arXiv: 1603.06522.
- de Lapparent, V., M. J. Geller, and J. P. Huchra (1986). “A slice of the universe”. In: *ApJ* 302, pages L1–L5. DOI: 10.1086/184625.
- de Vaucouleurs, G. (1953a). “Evidence for a local supergalaxy”. In: *AJ* 58, page 30. DOI: 10.1086/106805.
- (1953b). “On the distribution of mass and luminosity in elliptical galaxies”. In: *MNRAS* 113, page 134. DOI: 10.1093/mnras/113.2.134.

- de Vaucouleurs, G. (1970). "The Case for a Hierarchical Cosmology". In: *Science* 167, pages 1203–1213. DOI: 10.1126/science.167.3922.1203.
- Diaz, J. D. et al. (2014). "Balancing mass and momentum in the Local Group". In: *MNRAS* 443, pages 1688–1703. DOI: 10.1093/mnras/stu1210. arXiv: 1405.3662.
- Diemand, J., M. Kuhlen, and P. Madau (2007). "Formation and Evolution of Galaxy Dark Matter Halos and Their Substructure". In: *ApJ* 667, pages 859–877. DOI: 10.1086/520573. eprint: astro-ph/0703337.
- Diemand, J., B. Moore, and J. Stadel (2005). "Earth-mass dark-matter haloes as the first structures in the early Universe". In: *Nature* 433, pages 389–391. DOI: 10.1038/nature03270. eprint: astro-ph/0501589.
- Diemand, J. et al. (2008). "Clumps and streams in the local dark matter distribution". In: *Nature* 454, pages 735–738. DOI: 10.1038/nature07153. arXiv: 0805.1244.
- Diemer, B. and A. V. Kravtsov (2014). "Dependence of the Outer Density Profiles of Halos on Their Mass Accretion Rate". In: *ApJ* 789, 1, page 1. DOI: 10.1088/0004-637X/789/1/1. arXiv: 1401.1216.
- (2015). "A Universal Model for Halo Concentrations". In: *ApJ* 799, 108, page 108. DOI: 10.1088/0004-637X/799/1/108. arXiv: 1407.4730.
- Doroshkevich, A. G. (1970). "Spatial structure of perturbations and origin of galactic rotation in fluctuation theory". In: *Astrophysics* 6, pages 320–330. DOI: 10.1007/BF01001625.
- (1973). "Development of Rotation in Galaxies in the Nonlinear Theory of Gravitational Instability." In: *Soviet Ast.* 16, page 986.
- Driver, S. (2004). "Beyond the Galaxy Luminosity Function". In: *PASA* 21, pages 344–351. DOI: 10.1071/AS04035. eprint: astro-ph/0409461.
- Driver, S. P. et al. (2005). "The Millennium Galaxy Catalogue: the space density and surface-brightness distribution(s) of galaxies". In: *MNRAS* 360, pages 81–103. DOI: 10.1111/j.1365-2966.2005.08990.x. eprint: astro-ph/0503228.
- Drlica-Wagner, A. et al. (2015). "Eight Ultra-faint Galaxy Candidates Discovered in Year Two of the Dark Energy Survey". In: *ApJ* 813, 109, page 109. DOI: 10.1088/0004-637X/813/2/109. arXiv: 1508.03622.
- Dubinski, J. and R. G. Carlberg (1991). "The structure of cold dark matter halos". In: *ApJ* 378, pages 496–503. DOI: 10.1086/170451.
- Efstathiou, G., G. Lake, and J. Negroponte (1982). "The stability and masses of disc galaxies". In: *MNRAS* 199, pages 1069–1088. DOI: 10.1093/mnras/199.4.1069.
- Einasto, J. (1965). "On the Construction of a Composite Model for the Galaxy and on the Determination of the System of Galactic Parameters". In: *Trudy Astrofizicheskogo Instituta Alma-Ata* 5, pages 87–100.
- (2009). "Dark Matter". In: *arXiv/0901.0632*. arXiv: 0901.0632 [astro-ph.CO].
- Eisenstein, D. J. and A. Loeb (1995). "An analytical model for the triaxial collapse of cosmological perturbations". In: *ApJ* 439, pages 520–541. DOI: 10.1086/175193. eprint: astro-ph/9405012.
- Faber, S. M. et al. (1997). "The Centers of Early-Type Galaxies with HST. IV. Central Parameter Relations." In: *AJ* 114, page 1771. DOI: 10.1086/118606. eprint: astro-ph/9610055.
- Fakhouri, O., C.-P. Ma, and M. Boylan-Kolchin (2010). "The merger rates and mass assembly histories of dark matter haloes in the two Millennium simulations". In: *MNRAS* 406, pages 2267–2278. DOI: 10.1111/j.1365-2966.2010.16859.x. arXiv: 1001.2304 [astro-ph.CO].
- Fall, S. M. and G. Efstathiou (1980). "Formation and rotation of disc galaxies with haloes". In: *MNRAS* 193, pages 189–206. DOI: 10.1093/mnras/193.2.189.
- Fang, J. J. et al. (2013). "A Link between Star Formation Quenching and Inner Stellar Mass Density in Sloan Digital Sky Survey Central Galaxies". In: *ApJ* 776, 63, page 63. DOI: 10.1088/0004-637X/776/1/63. arXiv: 1308.5224.
- Felten, J. E. (1977). "Study of the luminosity function for field galaxies". In: *AJ* 82, pages 861–878. DOI: 10.1086/112140.
- Fillmore, J. A. and P. Goldreich (1984). "Self-similar gravitational collapse in an expanding universe". In: *ApJ* 281, pages 1–8. DOI: 10.1086/162070.

- Forman, W. et al. (1972). "Observations of the Extended X-Ray Sources in the Perseus and Coma Clusters from UHURU". In: *ApJ* 178, pages 309–316. DOI: 10.1086/151791.
- Freedman, W. L. and B. F. Madore (2010). "The Hubble Constant". In: *ARA&A* 48, pages 673–710. DOI: 10.1146/annurev-astro-082708-101829. arXiv: 1004.1856 [astro-ph.CO].
- Freeman, K. C. (1970). "On the Disks of Spiral and so Galaxies". In: *ApJ* 160, page 811. DOI: 10.1086/150474.
- Frieman, J. A., M. S. Turner, and D. Huterer (2008). "Dark Energy and the Accelerating Universe". In: *ARA&A* 46, pages 385–432. DOI: 10.1146/annurev.astro.46.060407.145243. arXiv: 0803.0982.
- Fukugita, M. et al. (1996). "The Sloan Digital Sky Survey Photometric System". In: *AJ* 111, page 1748. DOI: 10.1086/117915.
- Gao, L. et al. (2008). "The redshift dependence of the structure of massive Λ cold dark matter haloes". In: *MNRAS* 387, pages 536–544. DOI: 10.1111/j.1365-2966.2008.13277.x. arXiv: 0711.0746.
- Gnedin, N. Y. (2000). "Effect of Reionization on Structure Formation in the Universe". In: *ApJ* 542, pages 535–541. DOI: 10.1086/317042. eprint: astro-ph/0002151.
- Gnedin, N. Y. and N. Hollon (2012). "Cooling and Heating Functions of Photoionized Gas". In: *ApJS* 202, 13, page 13. DOI: 10.1088/0067-0049/202/2/13. arXiv: 1201.5116.
- Gnedin, N. Y. and A. V. Kravtsov (2011). "Environmental Dependence of the Kennicutt-Schmidt Relation in Galaxies". In: *ApJ* 728, 88, page 88. DOI: 10.1088/0004-637X/728/2/88. arXiv: 1004.0003 [astro-ph.CO].
- González, R. E., A. V. Kravtsov, and N. Y. Gnedin (2014). "On the Mass of the Local Group". In: *ApJ* 793, 91, page 91. DOI: 10.1088/0004-637X/793/2/91. arXiv: 1312.2587.
- Gott III, J. R. et al. (2005). "A Map of the Universe". In: *ApJ* 624, pages 463–484. DOI: 10.1086/428890. eprint: astro-ph/0310571.
- Graham, A. et al. (1996). "Brightest Cluster Galaxy Profile Shapes". In: *ApJ* 465, page 534. DOI: 10.1086/177440. eprint: astro-ph/9603006.
- Graham, A. W. and S. P. Driver (2005). "A Concise Reference to (Projected) Sérsic Quantities, Including Concentration, Profile Slopes, Petrosian Indices, and Kron Magnitudes". In: *PASA* 22, pages 118–127. DOI: 10.1071/AS05001. eprint: astro-ph/0503176.
- Graham, A. W. et al. (2005). "Total Galaxy Magnitudes and Effective Radii from Petrosian Magnitudes and Radii". In: *AJ* 130, pages 1535–1544. DOI: 10.1086/444475. eprint: astro-ph/0504287.
- Graham, A. W. et al. (2006). "Empirical Models for Dark Matter Halos. II. Inner Profile Slopes, Dynamical Profiles, and ρ/σ^3 ". In: *AJ* 132, pages 2701–2710. DOI: 10.1086/508990. eprint: astro-ph/0608613.
- Gregory, S. A. and L. A. Thompson (1978). "The Coma/A1367 supercluster and its environs". In: *ApJ* 222, pages 784–799. DOI: 10.1086/156198.
- Gross, M. A. K. et al. (1998). "Cold dark matter variant cosmological models - I. Simulations and preliminary comparisons". In: *MNRAS* 301, pages 81–94. DOI: 10.1046/j.1365-8711.1998.01998.x. eprint: astro-ph/9712142.
- Groth, E. J. and P. J. E. Peebles (1977). "Statistical analysis of catalogs of extragalactic objects. VII - Two- and three-point correlation functions for the high-resolution Shane-Wirtanen catalog of galaxies". In: *ApJ* 217, pages 385–405. DOI: 10.1086/155588.
- Gunn, J. E. (1977). "Massive galactic halos. I - Formation and evolution". In: *ApJ* 218, pages 592–598. DOI: 10.1086/155715.
- Gunn, J. E. and J. R. Gott III (1972). "On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution". In: *ApJ* 176, page 1. DOI: 10.1086/151605.
- Gursky, H. et al. (1971). "A Strong X-Ray Source in the Coma Cluster Observed by UHURU". In: *ApJ* 167, page L81. DOI: 10.1086/180765.
- Guth, A. H. (1981). "Inflationary universe: A possible solution to the horizon and flatness problems". In: *Phys. Rev. D* 23, pages 347–356. DOI: 10.1103/PhysRevD.23.347.
- Guth, A. H. and S.-Y. Pi (1982). "Fluctuations in the new inflationary universe". In: *Physical Review Letters* 49, pages 1110–1113. DOI: 10.1103/PhysRevLett.49.1110.

- Harrison, E. R. (1970). “Fluctuations at the Threshold of Classical Cosmology”. In: *Phys. Rev. D* 1, pages 2726–2730. DOI: 10.1103/PhysRevD.1.2726.
- Hart, R. and R. Berendzen (1971). “Hubble’s Classification of Non-Galactic Nebulae, 1922-1926”. In: *Journal for the History of Astronomy* 2, page 109.
- Heath, D. J. (1977). “The growth of density perturbations in zero pressure Friedmann-Lemaitre universes”. In: *MNRAS* 179, pages 351–358. DOI: 10.1093/mnras/179.3.351.
- Hernquist, L. (1990). “An analytical model for spherical galaxies and bulges”. In: *ApJ* 356, pages 359–364. DOI: 10.1086/168845.
- Hessman, F. V. (2015). “The difficulty of measuring the local dark matter density”. In: *A&A* 579, A123, A123. DOI: 10.1051/0004-6361/201526022. arXiv: 1506.00384.
- Hoeft, M. et al. (2006). “Dwarf galaxies in voids: suppressing star formation with photoheating”. In: *MNRAS* 371, pages 401–414. DOI: 10.1111/j.1365-2966.2006.10678.x. eprint: astro-ph/0501304.
- Hogg, D. W. (1999). “Distance measures in cosmology”. In: arXiv/9905116. eprint: astro-ph/9905116.
- Hogg, D. W. et al. (2002). “The K correction”. In: *ArXiv Astrophysics e-prints*. eprint: astro-ph/0210394.
- Holmberg, E. (1958). “A photographic photometry of extragalactic nebulae.” In: *Meddelanden fran Lunds Astronomiska Observatorium Serie II* 136, page 1.
- Hu, W. and A. V. Kravtsov (2003). “Sample Variance Considerations for Cluster Surveys”. In: *ApJ* 584, pages 702–715. DOI: 10.1086/345846. eprint: astro-ph/0203169.
- Huang, S. et al. (2013). “The Carnegie-Irvine Galaxy Survey. III. The Three-component Structure of Nearby Elliptical Galaxies”. In: *ApJ* 766, 47, page 47. DOI: 10.1088/0004-637X/766/1/47. arXiv: 1212.2639 [astro-ph.CO].
- Hubble, E. (1926a). “Non-Galactic Nebulae (Abstract)”. In: *PASP* 38, pages 258–260. DOI: 10.1086/123603.
- (1929). “A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae”. In: *Proceedings of the National Academy of Science* 15, pages 168–173. DOI: 10.1073/pnas.15.3.168.
- (1934). “The Realm of the Nebulae”. In: *The Scientific Monthly* 39, pages 193–202.
- Hubble, E. P. (1926b). “Extragalactic nebulae.” In: *ApJ* 64, page 321. DOI: 10.1086/143018.
- (1930). “Distribution of luminosity in elliptical nebulae.” In: *ApJ* 71. DOI: 10.1086/143250.
- Huertas-Company, M. et al. (2011). “Revisiting the Hubble sequence in the SDSS DR7 spectroscopic sample: a publicly available Bayesian automated classification”. In: *A&A* 525, A157, page 157. DOI: 10.1051/0004-6361/201015735. arXiv: 1010.3018.
- Huertas-Company, M. et al. (2015). “A Catalog of Visual-like Morphologies in the 5 CANDELS Fields Using Deep Learning”. In: *ApJS* 221, 8, page 8. DOI: 10.1088/0067-0049/221/1/8. arXiv: 1509.05429.
- Huss, A., B. Jain, and M. Steinmetz (1999). “How Universal Are the Density Profiles of Dark Halos?” In: *ApJ* 517, pages 64–69. DOI: 10.1086/307161. eprint: astro-ph/9803117.
- Ivezić, Ž. et al. (2013). *Statistics, Data Mining, and Machine Learning in Astronomy*.
- Jacoby, G. H. et al. (1992). “A critical review of selected techniques for measuring extragalactic distances”. In: *PASP* 104, pages 599–662. DOI: 10.1086/133035.
- Jeans, J. H. (1919). *Problems of cosmogony and stellar dynamics*.
- (1922). “The motions of stars in a Kapteyn universe”. In: *MNRAS* 82, pages 122–132. DOI: 10.1093/mnras/82.3.122.
- Joeveer, M. and J. Einasto (1978). “Has the universe the cell structure”. In: *Large Scale Structures in the Universe*. Edited by M. S. Longair and J. Einasto. Volume 79. IAU Symposium, pages 241–250.
- Kahn, F. D. and L. Woltjer (1959). “Intergalactic Matter and the Galaxy.” In: *ApJ* 130, page 705. DOI: 10.1086/146762.
- Kaiser, N. (1984). “On the spatial correlations of Abell clusters”. In: *ApJ* 284, pages L9–L12. DOI: 10.1086/184341.
- (1986). “Evolution and clustering of rich clusters”. In: *MNRAS* 222, pages 323–345. DOI: 10.1093/mnras/222.2.323.
- Kapteyn, J. C. (1922). “First Attempt at a Theory of the Arrangement and Motion of the Sidereal System”. In: *ApJ* 55, page 302. DOI: 10.1086/142670.

- Karachentsev, I. D., D. I. Makarov, and E. I. Kaisina (2013). “Updated Nearby Galaxy Catalog”. In: *AJ* 145, 101, page 101. DOI: 10.1088/0004-6256/145/4/101. arXiv: 1303.5328 [astro-ph.CO].
- Katz, N. (1991). “Dissipationless collapse in an expanding universe”. In: *ApJ* 368, pages 325–336. DOI: 10.1086/169696.
- Kazantzidis, S., A. R. Zentner, and A. V. Kravtsov (2006). “The Robustness of Dark Matter Density Profiles in Dissipationless Mergers”. In: *ApJ* 641, pages 647–664. DOI: 10.1086/500579. eprint: astro-ph/0510583.
- Kolb, E. W. and M. S. Turner (1990). *The early universe*.
- Komatsu, E. et al. (2011). “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation”. In: *ApJS* 192, 18, page 18. DOI: 10.1088/0067-0049/192/2/18. arXiv: 1001.4538 [astro-ph.CO].
- Koposov, S. E. et al. (2015). “Beasts of the Southern Wild : Discovery of nine Ultra Faint satellites in the vicinity of the Magellanic Clouds.” In: *ApJ* 805, 130, page 130. DOI: 10.1088/0004-637X/805/2/130. arXiv: 1503.02079.
- Kormendy, J. (1977). “Brightness distributions in compact and normal galaxies. II - Structure parameters of the spheroidal component”. In: *ApJ* 218, pages 333–346. DOI: 10.1086/155687.
- Kravtsov, A., A. Vikhlinin, and A. Meshcheryakov (2014). “Stellar mass – halo mass relation and star formation efficiency in high-mass halos”. In: *arXiv/1401.7329*. arXiv: 1401.7329 [astro-ph.CO].
- Kravtsov, A. V. (2013). “The Size-Virial Radius Relation of Galaxies”. In: *ApJ* 764, L31, page L31. DOI: 10.1088/2041-8205/764/2/L31. arXiv: 1212.2980 [astro-ph.CO].
- Krumholz, M. R. and A. Dekel (2012). “Metallicity-dependent Quenching of Star Formation at High Redshift in Small Galaxies”. In: *ApJ* 753, 16, page 16. DOI: 10.1088/0004-637X/753/1/16. arXiv: 1106.0301.
- Kuminski, E. and L. Shamir (2016). “Computer-generated visual morphology catalog of ~3,000,000 SDSS galaxies”. In: *arXiv/1602.06854*. arXiv: 1602.06854.
- Lahav, O. et al. (1991). “Dynamical effects of the cosmological constant”. In: *MNRAS* 251, pages 128–136. DOI: 10.1093/mnras/251.1.128.
- Lake, S. E. and E. L. Wright (2016). “K-corrections: an Examination of their Contribution to the Uncertainty of Luminosity Measurements”. In: *ApJ submitted (arXiv/1603.07299)*. arXiv: 1603.07299.
- Land, K. et al. (2008). “Galaxy Zoo: the large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey”. In: *MNRAS* 388, pages 1686–1692. DOI: 10.1111/j.1365-2966.2008.13490.x. arXiv: 0803.3247.
- Lau, E. T. et al. (2015). “Mass Accretion and its Effects on the Self-similarity of Gas Profiles in the Outskirts of Galaxy Clusters”. In: *ApJ* 806, 68, page 68. DOI: 10.1088/0004-637X/806/1/68. arXiv: 1411.5361.
- Lee, J. and S. F. Shandarin (1998). “The Cosmological Mass Distribution Function in the Zeldovich Approximation”. In: *ApJ* 500, pages 14–27. DOI: 10.1086/305710. eprint: astro-ph/9709200.
- Li, C. and S. D. M. White (2009). “The distribution of stellar mass in the low-redshift Universe”. In: *MNRAS* 398, pages 2177–2187. DOI: 10.1111/j.1365-2966.2009.15268.x. arXiv: 0901.0706.
- Li, Y.-S. and S. D. M. White (2008). “Masses for the Local Group and the Milky Way”. In: *MNRAS* 384, pages 1459–1468. DOI: 10.1111/j.1365-2966.2007.12748.x. arXiv: 0710.3740.
- Libeskind, N. I. et al. (2015). “Planes of satellite galaxies and the cosmic web”. In: *MNRAS* 452, pages 1052–1059. DOI: 10.1093/mnras/stv1302. arXiv: 1503.05915.
- Lima Neto, G. B., D. Gerbal, and I. Márquez (1999). “The specific entropy of elliptical galaxies: an explanation for profile-shape distance indicators?” In: *MNRAS* 309, pages 481–495. DOI: 10.1046/j.1365-8711.1999.02849.x. eprint: astro-ph/9905048.
- Limber, D. N. (1959). “Effects of Intracluster Gas and Duct upon the Virial Theorem.” In: *ApJ* 130, page 414. DOI: 10.1086/146733.
- Lineweaver, C. H. (2005). “Inflation and the Cosmic Microwave Background”. In: *The New Cosmology*. Edited by M. Colless, pages 31–65. DOI: 10.1142/9789812702357_0002. eprint: astro-ph/0305179.
- Lithwick, Y. and N. Dalal (2011). “Self-similar Solutions of Triaxial Dark Matter Halos”. In: *ApJ* 734, 100, page 100. DOI: 10.1088/0004-637X/734/2/100. arXiv: 1010.3723 [astro-ph.CO].

- Ma, C.-P. and E. Bertschinger (1995). “Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges”. In: *ApJ* 455, page 7. DOI: 10.1086/176550. eprint: astro-ph/9506072.
- MacMillan, J. D., L. M. Widrow, and R. N. Henriksen (2006). “On Universal Halos and the Radial Orbit Instability”. In: *ApJ* 653, pages 43–52. DOI: 10.1086/508602. eprint: astro-ph/0604418.
- Malmquist, K. G. (1922). “On some relations in stellar statistics”. In: *Meddelanden fran Lunds Astronomiska Observatorium Serie I* 100, pages 1–52.
- Mamon, G. A. and E. L. Łokas (2005). “Dark matter in elliptical galaxies - I. Is the total mass density profile of the NFW form or even steeper?” In: *MNRAS* 362, pages 95–109. DOI: 10.1111/j.1365-2966.2005.09225.x. eprint: astro-ph/0405466.
- McBride, J., O. Fakhouri, and C.-P. Ma (2009). “Mass accretion rates and histories of dark matter haloes”. In: *MNRAS* 398, pages 1858–1868. DOI: 10.1111/j.1365-2966.2009.15329.x. arXiv: 0902.3659.
- McConnachie, A. W. (2012). “The Observed Properties of Dwarf Galaxies in and around the Local Group”. In: *AJ* 144, 4, page 4. DOI: 10.1088/0004-6256/144/1/4. arXiv: 1204.1562.
- McMillan, P. J. (2011). “Mass models of the Milky Way”. In: *MNRAS* 414, pages 2446–2457. DOI: 10.1111/j.1365-2966.2011.18564.x. arXiv: 1102.4340.
- Meekins, J. F. et al. (1971). “Physical Sciences: X-rays from the Coma Cluster of Galaxies”. In: *Nature* 231, pages 107–108. DOI: 10.1038/231107a0.
- Meert, A., V. Vikram, and M. Bernardi (2015). “A catalogue of 2D photometric decompositions in the SDSS-DR7 spectroscopic main galaxy sample: preferred models and systematics”. In: *MNRAS* 446, pages 3943–3974. DOI: 10.1093/mnras/stu2333. arXiv: 1406.4179.
- Melchiorri, F. et al. (1981). “Fluctuations in the microwave background at intermediate angular scales”. In: *ApJ* 250, pages L1–L4. DOI: 10.1086/183662.
- Merritt, D. et al. (2006). “Empirical Models for Dark Matter Halos. I. Nonparametric Construction of Density Profiles and Comparison with Parametric Models”. In: *AJ* 132, pages 2685–2700. DOI: 10.1086/508988. eprint: astro-ph/0509417.
- Mo, H., F. C. van den Bosch, and S. White (2010). *Galaxy Formation and Evolution*.
- Mo, H. J., S. Mao, and S. D. M. White (1998). “The formation of galactic discs”. In: *MNRAS* 295, pages 319–336. DOI: 10.1046/j.1365-8711.1998.01227.x. eprint: astro-ph/9707093.
- Monaco, P. (1995). “The Mass Function of Cosmic Structures with Nonspherical Collapse”. In: *ApJ* 447, page 23. DOI: 10.1086/175853. eprint: astro-ph/9406029.
- Moni Bidin, C. et al. (2012). “Kinematical and Chemical Vertical Structure of the Galactic Thick Disk. II. A Lack of Dark Matter in the Solar Neighborhood”. In: *ApJ* 751, 30, page 30. DOI: 10.1088/0004-637X/751/1/30. arXiv: 1204.3924.
- Moni Bidin, C. et al. (2015). “On the local dark matter density”. In: *A&A* 573, A91, A91. DOI: 10.1051/0004-6361/201424675.
- Montero-Dorta, A. D. and F. Prada (2009). “The SDSS DR6 luminosity functions of galaxies”. In: *MNRAS* 399, pages 1106–1118. DOI: 10.1111/j.1365-2966.2009.15197.x. arXiv: 0806.4930.
- Moore, B. et al. (1999). “Cold collapse and the core catastrophe”. In: *MNRAS* 310, pages 1147–1152. DOI: 10.1046/j.1365-8711.1999.03039.x. eprint: astro-ph/9903164.
- More, S., B. Diemer, and A. V. Kravtsov (2015). “The Splashback Radius as a Physical Halo Boundary and the Growth of Halo Mass”. In: *ApJ* 810, 36, page 36. DOI: 10.1088/0004-637X/810/1/36. arXiv: 1504.05591.
- More, S. et al. (2016). “Detection of the Splashback Radius and Halo Assembly bias of Massive Galaxy Clusters”. In: *Phys. Rev. D* in press (arXiv/1601.06063). arXiv: 1601.06063.
- Navarro, J. F., C. S. Frenk, and S. D. M. White (1995). “Simulations of X-ray clusters”. In: *MNRAS* 275, pages 720–740. DOI: 10.1093/mnras/275.3.720. eprint: astro-ph/9408069.
- (1996). “The Structure of Cold Dark Matter Halos”. In: *ApJ* 462, page 563. DOI: 10.1086/177173. eprint: astro-ph/9508025.
- (1997). “A Universal Density Profile from Hierarchical Clustering”. In: *ApJ* 490, pages 493–508. eprint: astro-ph/9611107.

- Navarro, J. F. et al. (2004). “The inner structure of Λ CDM haloes - III. Universality and asymptotic slopes”. In: *MNRAS* 349, pages 1039–1051. DOI: 10.1111/j.1365-2966.2004.07586.x. eprint: astro-ph/0311231.
- Nelson, K. et al. (2014). “Weighing Galaxy Clusters with Gas. II. On the Origin of Hydrostatic Mass Bias in Λ CDM Galaxy Clusters”. In: *ApJ* 782, 107, page 107. DOI: 10.1088/0004-637X/782/2/107. arXiv: 1308.6589 [astro-ph.CO].
- Okamoto, T., L. Gao, and T. Theuns (2008). “Mass loss of galaxies due to an ultraviolet background”. In: *MNRAS* 390, pages 920–928. DOI: 10.1111/j.1365-2966.2008.13830.x. arXiv: 0806.0378.
- Oke, J. B. and J. E. Gunn (1983). “Secondary standard stars for absolute spectrophotometry”. In: *ApJ* 266, pages 713–717. DOI: 10.1086/160817.
- Oke, J. B. and A. Sandage (1968). “Energy Distributions, K Corrections, and the Stebbins-Whitford Effect for Giant Elliptical Galaxies”. In: *ApJ* 154, page 21. DOI: 10.1086/149737.
- Oort, J. H. (1932). “The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems”. In: *BAIN* 6, page 249.
- Ostriker, J. P. and P. J. E. Peebles (1973). “A Numerical Study of the Stability of Flattened Galaxies: or, can Cold Galaxies Survive?” In: *ApJ* 186, pages 467–480. DOI: 10.1086/152513.
- Patej, A. and A. Loeb (2015). “A Simple Physical Model for the Gas Distribution in Galaxy Clusters”. In: *ApJ* 798, L20, page L20. DOI: 10.1088/2041-8205/798/1/L20. arXiv: 1411.2971.
- Peñarrubia, J. et al. (2014). “A dynamical model of the local cosmic expansion”. In: *MNRAS* 443, pages 2204–2222. DOI: 10.1093/mnras/stu879. arXiv: 1405.0306.
- Peebles, P. J. E. (1974). “The Nature of the Distribution of Galaxies”. In: *A&A* 32, page 197.
- (1980). *The large-scale structure of the universe*. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.
- (1982). “Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations”. In: *ApJ* 263, pages L1–L5. DOI: 10.1086/183911.
- (1993). *Principles of Physical Cosmology*.
- Peebles, P. J. E. and M. G. Hauser (1974). “Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs”. In: *ApJS* 28, page 19. DOI: 10.1086/190308.
- Percival, W. J. (2005). “Cosmological structure formation in a homogeneous dark energy background”. In: *A&A* 443, pages 819–830. DOI: 10.1051/0004-6361:20053637. eprint: astro-ph/0508156.
- Persic, M., P. Salucci, and F. Stel (1996). “The universal rotation curve of spiral galaxies - I. The dark matter connection”. In: *MNRAS* 281, pages 27–47. DOI: 10.1093/mnras/281.1.27. eprint: astro-ph/9506004.
- Petrosian, V. (1976). “Surface brightness and evolution of galaxies”. In: *ApJ* 209, pages L1–L5. DOI: 10.1086/182253.
- Pitts, E. and R. J. Tayler (1997). “Absorption-line cloud models: distribution of column densities and velocity profiles”. In: *MNRAS* 288, pages 457–469.
- Planck Collaboration (2016). “Planck 2015 results. XIII. Cosmological parameters”. In: *A&A* 594, A13, A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589.
- Planck Collaboration et al. (2015). “Planck 2015 results. XIII. Cosmological parameters”. In: *å submitted (arXiv/1502.01589)*. arXiv: 1502.01589.
- Pohlen, M. and I. Trujillo (2006). “The structure of galactic disks. Studying late-type spiral galaxies using SDSS”. In: *A&A* 454, pages 759–772. DOI: 10.1051/0004-6361:20064883. eprint: astro-ph/0603682.
- Polyachenko, V. L. and I. G. Shukhman (1981). “General Models of Collisionless Spherically Symmetric Stellar Systems - a Stability Analysis”. In: *Soviet Ast.* 25, page 533.
- Press, W. H. and P. Schechter (1974). “Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation”. In: *ApJ* 187, pages 425–438. DOI: 10.1086/152650.
- Proctor, R. A. (1869). “Distribution of the Nebulae”. In: *MNRAS* 29, page 337.

- Raychaudhury, S. (1989). "The distribution of galaxies in the direction of the 'Great Attractor'". In: *Nature* 342, pages 251–255. DOI: 10.1038/342251a0.
- Richings, A. J., J. Schaye, and B. D. Oppenheimer (2014). "Non-equilibrium chemistry and cooling in the diffuse interstellar medium - II. Shielded gas". In: *MNRAS* 442, pages 2780–2796. DOI: 10.1093/mnras/stu1046. arXiv: 1403.6155.
- Roberts, M. S. and A. H. Rots (1973). "Comparison of Rotation Curves of Different Galaxy Types". In: *A&A* 26, pages 483–485.
- Robertson, B. E. et al. (2009). "Collapse Barriers and Halo Abundance: Testing the Excursion Set Ansatz". In: *ApJ* 696, pages 636–652. DOI: 10.1088/0004-637X/696/1/636. arXiv: 0812.3148.
- Roediger, J. C. and S. Courteau (2015). "On the uncertainties of stellar mass estimates via colour measurements". In: *MNRAS* 452, pages 3209–3225. DOI: 10.1093/mnras/stv1499. arXiv: 1507.03016.
- Rubin, V. C. and W. K. Ford Jr. (1970). "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". In: *ApJ* 159, page 379. DOI: 10.1086/150317.
- Rubin, V. C., W. K. J. Ford, and N. Thonnard (1980). "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/". In: *ApJ* 238, pages 471–487. DOI: 10.1086/158003.
- Rubin, V. C., N. Thonnard, and W. K. Ford Jr. (1978). "Extended rotation curves of high-luminosity spiral galaxies. IV - Systematic dynamical properties, SA through SC". In: *ApJ* 225, pages L107–L111. DOI: 10.1086/182804.
- Ryden, B. S. and J. E. Gunn (1987). "Galaxy formation by gravitational collapse". In: *ApJ* 318, pages 15–31. DOI: 10.1086/165349.
- Sandage, A. (2005). "The Classification of Galaxies: Early History and Ongoing Developments". In: *ARA&A* 43, pages 581–624. DOI: 10.1146/annurev.astro.43.112904.104839.
- Schechter, P. (1976). "An analytic expression for the luminosity function for galaxies." In: *ApJ* 203, pages 297–306. DOI: 10.1086/154079.
- Schmidt, M. (1968). "Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources". In: *ApJ* 151, page 393. DOI: 10.1086/149446.
- Seigar, M. S., A. W. Graham, and H. Jerjen (2007). "Intracluster light and the extended stellar envelopes of cD galaxies: an analytical description". In: *MNRAS* 378, pages 1575–1588. DOI: 10.1111/j.1365-2966.2007.11899.x. eprint: astro-ph/0612229.
- Sersic, J. L. (1968). *Atlas de galaxias australes*.
- Shapley, H. (1930). "Note on a Remote Cloud of Galaxies in Centaurus". In: *Harvard College Observatory Bulletin* 874, pages 9–12.
- Sharma, P. et al. (2012). "Thermal instability and the feedback regulation of hot haloes in clusters, groups and galaxies". In: *MNRAS* 420, pages 3174–3194. DOI: 10.1111/j.1365-2966.2011.20246.x. arXiv: 1106.4816.
- Sheth, R. K., H. J. Mo, and G. Tormen (2001). "Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes". In: *MNRAS* 323, pages 1–12. DOI: 10.1046/j.1365-8711.2001.04006.x. eprint: astro-ph/9907024.
- Sheth, R. K. and G. Tormen (2002). "An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier". In: *MNRAS* 329, pages 61–75. DOI: 10.1046/j.1365-8711.2002.04950.x. eprint: astro-ph/0105113.
- Smith, S. (1936). "The Mass of the Virgo Cluster". In: *ApJ* 83, page 23. DOI: 10.1086/143697.
- Smolčić, V. et al. (2006). "The rest-frame optical colours of 99000 Sloan Digital Sky Survey galaxies". In: *MNRAS* 371, pages 121–137. DOI: 10.1111/j.1365-2966.2006.10662.x.
- Sohn, S. T., J. Anderson, and R. P. van der Marel (2012). "The M31 Velocity Vector. I. Hubble Space Telescope Proper-motion Measurements". In: *ApJ* 753, 7, page 7. DOI: 10.1088/0004-637X/753/1/7. arXiv: 1205.6863 [astro-ph.GA].
- Sparke, L. S. and J. S. Gallagher III (2007). *Galaxies in the Universe*.

- Starobinskii, A. A. (1983). “The Perturbation Spectrum Evolving from a Nonsingular Initially De-Sitter Cosmology and the Microwave Background Anisotropy”. In: *Soviet Astronomy Letters* 9, pages 302–304.
- Starobinsky, A. A. (1980). “A new type of isotropic cosmological models without singularity”. In: *Physics Letters B* 91, pages 99–102. DOI: 10.1016/0370-2693(80)90670-X.
- Strateva, I. et al. (2001). “Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data”. In: *AJ* 122, pages 1861–1874. DOI: 10.1086/323301. eprint: astro-ph/0107201.
- Szomoru, D. et al. (2013). “The Stellar Mass Structure of Massive Galaxies from $z = 0$ to $z = 2.5$: Surface Density Profiles and Half-mass Radii”. In: *ApJ* 763, 73, page 73. DOI: 10.1088/0004-637X/763/2/73. arXiv: 1208.4363.
- Tasitsiomi, A. et al. (2004). “Density Profiles of Λ CDM Clusters”. In: *ApJ* 607, pages 125–139. DOI: 10.1086/383219. eprint: astro-ph/0311062.
- Thompson, L. A. and S. A. Gregory (2011). “An Historical View: The Discovery of Voids in the Galaxy Distribution”. In: *arXiv/1109.1268*. arXiv: 1109.1268.
- Valluri, M. et al. (2007). “On Relaxation Processes in Collisionless Mergers”. In: *ApJ* 658, pages 731–747. DOI: 10.1086/511298. eprint: astro-ph/0609612.
- van Albada, G. B. (1960). “Formation and evolution of clusters of galaxies (Errata: 15 330)”. In: *BAIN* 15, page 165.
- van Dokkum, P. G. et al. (2015). “Forty-seven Milky Way-sized, Extremely Diffuse Galaxies in the Coma Cluster”. In: *ApJ* 798, L45, page L45. DOI: 10.1088/2041-8205/798/2/L45. arXiv: 1410.8141.
- Vitvitska, M. et al. (2002). “The Origin of Angular Momentum in Dark Matter Halos”. In: *ApJ* 581, pages 799–809. DOI: 10.1086/344361. eprint: astro-ph/0105349.
- Wands, D. and A. Slosar (2009). “Scale-dependent bias from primordial non-Gaussianity in general relativity”. In: *Phys. Rev. D* 79.12, 123507, page 123507. DOI: 10.1103/PhysRevD.79.123507. arXiv: 0902.1084 [astro-ph.CO].
- Wechsler, R. H. et al. (2002). “Concentrations of Dark Halos from Their Assembly Histories”. In: *ApJ* 568, pages 52–70. DOI: 10.1086/338765. eprint: astro-ph/0108151.
- Weinberg, D. H., L. Hernquist, and N. Katz (1997). “Photoionization, Numerical Resolution, and Galaxy Formation”. In: *ApJ* 477, pages 8–20. eprint: astro-ph/9604175.
- Widrow, L. M. et al. (2014). “Bending and breathing modes of the Galactic disc”. In: *MNRAS* 440, pages 1971–1981. DOI: 10.1093/mnras/stu396. arXiv: 1404.4069.
- Wiersma, R. P. C., J. Schaye, and B. D. Smith (2009). “The effect of photoionization on the cooling rates of enriched, astrophysical plasmas”. In: *MNRAS* 393, pages 99–107. DOI: 10.1111/j.1365-2966.2008.14191.x. arXiv: 0807.3748.
- Yoachim, P. and J. J. Dalcanton (2006). “Structural Parameters of Thin and Thick Disks in Edge-on Disk Galaxies”. In: *AJ* 131, pages 226–249. DOI: 10.1086/497970. eprint: astro-ph/0508460.
- Zeldovich, Y. B. (1972). “A hypothesis, unifying the structure and the entropy of the Universe”. In: *MNRAS* 160, 1P. DOI: 10.1093/mnras/160.1.1P.
- Zhao, D. H. et al. (2003). “The growth and structure of dark matter haloes”. In: *MNRAS* 339, pages 12–24. DOI: 10.1046/j.1365-8711.2003.06135.x. eprint: astro-ph/0204108.
- Zhao, D. H. et al. (2009). “Accurate Universal Models for the Mass Accretion Histories and Concentrations of Dark Matter Halos”. In: *ApJ* 707, pages 354–369. DOI: 10.1088/0004-637X/707/1/354. arXiv: 0811.0828.
- Zibetti, S., S. Charlot, and H.-W. Rix (2009). “Resolved stellar mass maps of galaxies - I. Method and implications for global mass estimates”. In: *MNRAS* 400, pages 1181–1198. DOI: 10.1111/j.1365-2966.2009.15528.x. arXiv: 0904.4252 [astro-ph.CO].
- Zwicky, F. (1937). “On the Masses of Nebulae and of Clusters of Nebulae”. In: *ApJ* 86, page 217. DOI: 10.1086/143864.