

# **AST4320: Cosmology and Extragalactic Astrophysics.**

Mark Dijkstra (Institute of Theoretical Astrophysics, University of Oslo)

(Dated: November 30, 2016)

Lecture notes for AST4320.

## Contents

<b>I. Preface</b>	3
<b>II. Lecture 1: Introduction &amp; Background</b>	4
A. A Brief Summary of Cosmology	4
B. Outline of this Course	7
<b>III. Simplified Picture of Growth of Structure: Application of Kirchhoff Theorem</b>	8
<b>IV. Growth of Small Perturbations in a Expanding &amp; Static Media</b>	9
A. Basic Equations	10
B. The Unperturbed Equations	10
C. The Perturbed Equations	10
D. Switch to Comoving Coordinates	11
E. The ‘Final’ Equation: A Differential Equation which determines $\delta(t)$	12
F. Static Medium & The Jeans Length	13
G. Linear Perturbation Theory in Expanding Media	14
H. Implications for Growth of Structure in an Expanding Universe	16
<b>V. Spherical Top-Hat Model</b>	18
<b>VI. Virialization</b>	21
A. The Virial Theorem	21
B. Virialization	22
C. Useful Quantities Related to Virialization	23
D. Virialization beyond the Top-Hat Model	24
E. Virialization of Baryons: A First Taste of Gastrophysics	24
F. Cooling of Virialized Gas	25
G. Gas Cooling & Galaxy Formation	27
<b>VII. Gaussian Random Fields</b>	29
A. Introduction	29
B. Single-Variate Gaussian Random Fields	30
C. Multi-Variate Gaussian Random Fields	30
D. Multi-Variate Gaussian Random Fields in Fourier Space	31
E. $\sigma(M)$ : Time evolution & Relation to $P(k)$	31
<b>VIII. Statistical Descriptions of Observations of Galaxies</b>	33
A. The Galaxy Luminosity Function (LF)	33
B. Clustering	34
<b>IX. The Press Schechter Formalism</b>	34
A. The Basic Press-Schechter Formalism	34
B. The Origin of the Fudge Factor of 2	38
<b>X. Galaxy Clustering in the Press-Schechter Formalism</b>	40
A. The Two-Point Function of Collapsed Objects and ‘Bias’ in the PS Formalism	40
<b>XI. Cosmological Uses of Galaxy Clustering</b>	43
A. Acoustic Scale and Baryon Acoustic Oscillations	43
B. Two 2D-Two-Point Function I: The Alcock-Paczynski Test	44
C. Two 2D-Two-Point Function II & 2D Power Spectrum: Redshift Space Distortions	45
<b>XII. What Shapes the Matter Power Spectrum <math>P(k)</math></b>	46
A. Intermezzo: The Horizon.	47
B. The Transfer Function $T(k)$ .	48
<b>XIII. The Matter Power Spectrum <math>P(k)</math></b>	50
A. The Transfer Function $T(k)$	50
B. The Power Spectrum $P(k)$ .	50
C. Why $n = 1$ corresponds to Scale Invariance.	51
D. Supplemental Reading Material I: Pre-Midterm	53
<b>XIV. Structure of Dark Matter Halos &amp; Cusp-Core Problem</b>	54
A. The Isothermal Density Profile	54
B. Simulated Dark Matter Density Profiles: Navarro-Frenk-White (NFW) Profiles	55
C. Observational Constraints on Dark Matter Density Profiles: Rotation Curves	57
<b>XV. Structure of Dark Matter Halos &amp; Missing Satellite Problem</b>	60
A. Feedback to the Rescue: Photoionization Feedback	60

B. Feedback to the Rescue: (A simple model for) Supernova Feedback	62
C. Has the MSP morphed into the ‘Too Big to Fail’ Problem?	64
<b>XVI. Is the Small-Scale Crisis in CDM related to the Nature of Dark Matter?</b>	65
A. Cold vs Warm Dark Matter	65
B. Constraints of Other Properties of Dark Matter	67
C. Candidate Dark Matter Particles	68
D. Searches & Claimed Detections for Dark Matter	68
<b>XVII. Ly<math>\alpha</math> Forest</b>	69
A. The Mean Number Density of Hydrogen Nuclei/Atoms in the Universe	69
B. The Ly $\alpha$ Forest	70
C. Optical Depth of the Photoionized IGM to Photons with $\nu > \nu_0$	72
<b>XVIII. Ly<math>\alpha</math> Forest II</b>	74
A. Derivation of Opacity of the IGM in Ionized Gas	74
B. The Photoionization Rate, The Volume Emissivity of Ionizing Photons, Sources of Reionization	76
C. Ly $\alpha$ Forest as a Cosmological Probe	77
<b>XIX. Galaxy Formation in More Detail</b>	79
A. Cold vs Hot Model Accretion in Spherical Collapse	79
B. Non-Spherical Gas Accretion and Cold Flows	81
C. Galaxy Formation: Why Disks?	81
D. Galaxy Formation: Why Ellipticals?	82
<b>XX. Star Formation</b>	82
A. How do Giant Molecular Clouds (GMCs) Form?	83
B. How do Stars form from GMCs?	85
<b>XXI. Black Holes &amp; their Formation</b>	87
A. Origin of Stellar Mass Black Holes	87
B. Quantitative Derivation of the Chandrasekhar Mass: Not Needed for Exam!	88
C. Geometrically Thin Accretion Disks: Not Needed for Exam!	90
D. Temperature Structure in Geometrically Thin Accretion Disks.	91
E. Spectra Emerging from Accretion Disks: Not Needed for Exam!	92
F. The Innermost Stable Circular Orbit (ISCO)	92
G. Radiative Efficiency $\eta$	93
H. Derivation of Eddington Luminosity	93
I. Black Hole Mass Growth in the Eddington Limit	94
<b>XXII. Useful Reading</b>	94
<b>References</b>	95

## I. PREFACE

The main purpose of this course is to address one of the most fundamental questions known to mankind: *What is the origin of our Earth, our Sun, and all stars and galaxies?* More specifically, some questions we will address are:

- How do stars, galaxies and black holes form?
- Why do stars & galaxies have the masses/properties they have?
- How can we test our leading theories?
- What are the main challenges our current theories are facing?
- How do we expect to overcome these challenges?
- Who is Elton John?

There is a mid-term exam midway through this course, which divides the class up into two parts:

1. In the first part *before* the midterm, we discuss the basics of how structures grow within our Universe, i.e. how a small fluctuation in a cosmological density field evolves with time and how this may lead to the formation of a galaxy. This simple picture of galaxy formation offers insight into basic properties of galaxies. With this simple picture of structure formation in place, we connect the statistical properties of a cosmological (Gaussian random) density field, to observed statistical properties of galaxies, and again find that this picture works remarkably

well. We also discuss how we can use the observed statistical properties of galaxies to learn about the details of the cosmological density field (in spite of uncertainties associated with our understanding of galaxy formation).

2. In the second part, *after* the midterm, we start looking at galaxy formation etc in more detail. We look gas assembles inside collapsed dark matter halos, how stars form from this gas, how these stars evolve, their radiation. We look at the origin of black holes, both stellar & super massive black holes. We also test discuss how we can test the structure formation paradigm with observations of the Ly $\alpha$  forest, which provides a novel & powerful way of directly observing fluctuations in the cosmological density field.

Throughout these lectures, I will use CGS units, as is common in the astronomy literature. Table I provides an overview of numerical values of physical constants and other parameters that you will encounter frequently within this course.

TABLE I Symbol Dictionary

Symbol	Definition
$k_B$	Boltzmann constant: $k_B = 1.38 \times 10^{-16}$ erg K $^{-1}$
$h_P$	Planck constant: $h_P = 6.67 \times 10^{-27}$ erg s
$\hbar$	reduced Planck constant: $\hbar = \frac{h_P}{2\pi}$
$m_p$	proton mass: $m_p = 1.66 \times 10^{-24}$ g
$m_e$	electron mass: $m_e = 9.1 \times 10^{-28}$ g
$q (= e)$	electron charge: $q = e = 4.8 \times 10^{-10}$ esu
$c$	speed of light: $c = 2.9979 \times 10^{10}$ cm s $^{-1}$
$G$	gravitational constant: $G = 6.67 \times 10^{-8}$ cm $^3$ g $^{-1}$ s $^{-2}$
pc	parsec: 1 pc = $3.08 \times 10^{18}$ cm
Mpc	megaparsec: Mpc = $3.08 \times 10^{24}$ cm

Finally, in this course you will be given assignment sheets, which are intended to help you understand the material covered in these notes. Note that occasionally, a question is printed in bold-face. Sometimes, these questions do not make it into the assignments. I would still recommend you try and answer these questions, as it may help you prepare for the mid-term.

## II. LECTURE 1: INTRODUCTION & BACKGROUND

An excellent introductory text can be found in Chapter 1 of ‘*The First Galaxies in the Universe*’ by A. Loeb & S. Furlanetto. The first chapter of the book is freely available at <http://press.princeton.edu/chapters/s9914.pdf>. In this chapter, we will (re)derive & (re)introduce some key physical concepts in cosmology that are important throughout the lecture series.

### A. A Brief Summary of Cosmology

Consider a spherical shell of radius  $R$ , which is expanding with velocity  $v$ . We further assume that the gas is distributed in a spherically symmetric configuration around  $r = 0$ . This gas is characterized by its density  $\rho(r)$ , and its kinematics is determined only by the total gas mass enclosed<sup>1</sup> within  $R$ , which equals  $M = 4\pi R^3 \rho/3$ . The total energy per unit mass (the *specific* energy) of the shell equals

$$E_{\text{tot}}/M_{\text{shell}} = \frac{1}{2}v^2 - \frac{GM}{r} \equiv \frac{1}{2}v^2(1 - \Omega), \quad \Omega \equiv \frac{\rho}{\rho_{\text{crit}}}, \quad \rho_{\text{crit}} = \frac{3H^2}{8\pi G}, \quad H \equiv \frac{v}{R}. \quad (1)$$

---

<sup>1</sup> Gauss Law

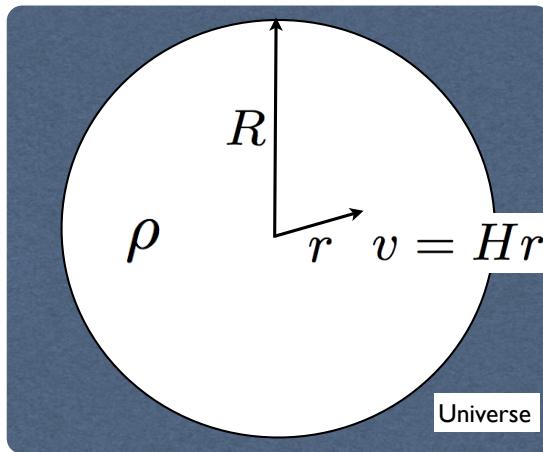


FIG. 1 Adopted geometry which we used to introduce the ‘critical density’,  $\rho_{\text{crit}}$ , using classical Newtonian mechanics only.

Eq 1 has three interesting regimes:

1.  $\Omega > 1$ , i.e.  $\rho > \rho_{\text{crit}}$ : the total specific energy is negative. This means that the negative gravitational potential energy exceeds the kinetic energy, and the shell falls back to  $r = 0$ .
2.  $\Omega = 1$ , i.e.  $\rho = \rho_{\text{crit}}$ : the specific energy is zero. This means that the shell comes to rest at  $r \rightarrow \infty$ .
3.  $\Omega < 1$ , i.e.  $\rho < \rho_{\text{crit}}$ : the specific energy is positive. This means that the shell continues to expand forever. The enclosed density within  $R$  thus represents a key parameter in determining the fate of the gaseous shell.

The previous discussion is based on classical Newtonian mechanics. The expansion of the Universe should formally be treated in general relativity (GR). However, in GR we get the same result: the overall expansion rate of the Universe and its evolution with time depend critically on the mean (matter) density in the Universe,  $\bar{\rho}$ , and whether or not it exceeds the critical density  $\rho_{\text{crit}} = \frac{3H^2}{8\pi G} \sim 1.88 \times 10^{-29} h^2 \text{ g cm}^{-3}$  for a Hubble parameter  $H = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ . An additional aspect of GR that distinguishes itself from the Newtonian calculation, is that the overall expansion history of the Universe is coupled to its overall geometry: the case  $\Omega = 1$  now corresponds to a ‘flat Universe’, which is characterized by a ‘flat’ geometry which refers to ‘standard’ Euclidian geometry that we are taught in high-school. For  $\Omega > 1$  the Universe is ‘closed’. An example of a closed 2-D plane is the surface of a sphere. For  $\Omega < 1$  the Universe is ‘open’. For some visual illustrations of such geometries, I refer you to the lecture notes. For a more detailed discussion on open, closed and flat Universes, I refer you to the cosmology course (or any text book on cosmology)

Throughout these lectures, we often distinguish between ‘comoving’ and ‘proper’ coordinates. Comoving coordinates/distances ‘comove’ with the expansion of the Universe, while proper coordinates/distances correspond to actual physical coordinates/distances. The relation between the two is

$$\mathbf{r}_p \equiv a(t)\mathbf{r}_c, \quad (2)$$

where  $\mathbf{r}_p$  denotes a proper distance/coordinate,  $\mathbf{r}_c$  denotes a comoving distance/coordinate, and  $a(t)$  denotes the ‘scale factor’ which quantifies the overall expansion of the Universe. It is common to normalize the scale factor  $a(t_0) = 1$  to 1 to the present-day Universe. This implies that  $a(t) < 1$  at earlier times  $t < t_0$ . There are various quantities that relate to the scale factor including the Hubble parameter  $H(t)$  and redshift  $z$ :

$$H \equiv \frac{\dot{a}}{a}, \quad \dot{a} \equiv \frac{da}{dt}, \quad a \equiv \frac{1}{1+z}. \quad (3)$$

The overall expansion history of the Universe is set by the overall material content of the Universe (as mentioned above). From GR we can derive the Friedmann-Robertson-Walker equations, which we can use to write the time

evolution of the Hubble parameter as:

$$\frac{H(t)}{H_0} = \sqrt{\Omega_m a^{-3} + \Omega_\Lambda + \Omega_r a^{-4}}, \quad (4)$$

where  $\Omega_m \equiv \rho/\rho_{\text{crit}}$  denotes the density parameter for the total material density in the Universe,  $\Omega_{\text{rad}}$  indicates the contribution from photons. Even though these are mass-less, in GR they affect the overall expansion history of the Universe. It is left as an exercise to show that:

$$\begin{aligned} a(t) &\propto t^{-2/3}, & \text{matter domination} \\ a(t) &\propto t^{-1/2}, & \text{radiation domination} \\ a(t) &\propto \exp[H_{\text{vac}}t], & H_{\text{vac}} \equiv k, \quad \text{vacuum } (\Omega_\Lambda) \text{ domination.} \end{aligned} \quad (5)$$

A key concept which we will talk about frequently throughout these lectures is ‘cosmological redshifting’. As a photon travels a proper distance  $ds$ , a (proper) time  $dt = ds/c$  elapses. During this time, the Universe as a whole stretches an infinitesimal amount, which stretches the wavelength of the photon. In other words, the frequency of the photon of frequency  $\nu$  becomes smaller by an amount  $d\nu$ . We can calculate  $d\nu$  by pretending that the cosmological redshift is due to a Doppler shift imparted by the expansion ‘velocity’,  $dv = Hds$ , of the Universe over  $ds$ , i.e.

$$\boxed{\frac{d\nu}{\nu} = -\frac{Hds}{c}}. \quad (6)$$

This is an important result: consider photons emitted by some distant source (i.e. a galaxy) at some frequency  $\nu$ . Then, due to the expansion of the Universe, the photon changes its frequency continuously as it is propagating, and we can for example replace a line-integral  $\int ds \rightarrow \int d\nu$ . As we will discuss in later lectures, this change of variables is very useful in extra-galactic astrophysics. We can also substitute  $ds = cdt$  and  $H = \dot{a}/a$  into Eq 6 to obtain

$$\frac{d\nu}{\nu} = -\frac{\dot{a}dt}{a} = -\frac{da}{a} \Rightarrow \nu \propto a^{-1}, \quad \lambda \propto a. \quad (7)$$

These last two equations are also important in extragalactic astrophysics: the wavelengths of a photon increases with the scale factor  $\lambda \propto a = (1+z)^{-1}$ . Non-relativistic particles like protons, neutrons, electrons, heavier atomic nuclei, also have a characteristic wavelength associated with them, namely the so-called ‘De-Broglie’ wavelength,  $\lambda_{\text{DB}} = h_p/p$ . This wavelength also stretches along with the expansion of the Universe. For non-relativistic particles, the energy depends on scale factor as

$$E_{\text{nr}} \propto p^2 \propto \lambda_{\text{DB}}^{-2} \propto a^{-2}. \quad (8)$$

Another way of writing this is by casting the (mean) energy of a thermal non-relativistic particle as  $\langle E_{\text{nr}} \rangle = k_b T$ , where  $T$  denotes the gas temperature. In other words, the temperature of ‘ordinary’ gas that is adiabatically cooling in the expanding Universe decreases as

$$T_{\text{gas}} \propto a^{-2} = (1+z)^2, \quad (9)$$

while for photons the photon frequency, and this energy scales as  $E_\gamma \propto a^{-1} = (1+z)$ . Note that this scaling does not apply to photons alone, it applies to any relativistic particle. This is because for relativistic particles we have  $E_r = pc \propto \lambda_{\text{DB}}^{-1} \propto a^{-1}$ . For a black-body spectrum of photons, such as the Cosmic Microwave Background (CMB, more on this below), the frequency distribution of photons is set entirely by a single parameter, the temperature  $T_\gamma$ . It is left as an exercise to show that

$$T_\gamma^{\text{bb}} \propto a^{-1} = (1+z). \quad (10)$$

The fact that adiabatic cooling affects the CMB temperature more slowly than that of adiabatically cooling gas inside it, has some interesting implications for the detectability of gas in the Universe before any stars, or galaxies were able to form.

The finite speed of light allows us to see the temporal evolution of our Universe in action. We see objects at redshift  $z$  at a time when the scale factor of the Universe was  $a = 1/(1+z)$ . Our most distant observations are those of the CMB. The surface of last scattering of the CMB has a redshift of  $z \sim 1090$ , and it characterized by a blackbody spectrum.

Some other useful quantities include the ‘luminosity distance’ to a source at  $z_{\text{em}}$ ,  $d_L(z_{\text{em}})$ , which is defined as

$$\text{flux} = \frac{\text{luminosity}}{4\pi d_L^2(z_{\text{em}})}, \quad (11)$$

where ‘flux’ denotes the observed flux of an astronomical object, and the ‘luminosity’ denotes its total radiative output.

Similarly, we can define the ‘angular diameter distance’,  $d_A(z_{\text{em}})$  as:

$$\text{size} = \text{angle} \times d_A(z_{\text{em}}) \quad (12)$$

where ‘size’ denotes the physical/proper size of an astronomical object, while ‘angle’ denotes the angle it subtends on the sky. The luminosity distance and angular diameter distance can be computed as

$$d_L(z_{\text{em}}) = R_{\text{em}}(1+z_{\text{em}}), \quad d_A(z_{\text{em}}) = R_{\text{em}}/(1+z_{\text{em}}), \quad (13)$$

where

$$R_{\text{em}} = \int_{t_{\text{em}}}^{t_{\text{obs}}} \frac{cdt}{a(t)} = \frac{c}{H_0} \int_0^{z_{\text{em}}} \frac{dz}{\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}}. \quad (14)$$

*‘Most of our information on astronomical objects are derived from the radiation we receive from it, or by the absorption it causes in the light of background sources* (source: Mo, Van den Bosch & White book). This information is encoded within the objects Spectral Energy Distribution (SED)

$$f_\lambda d\lambda, \quad f_\nu d\nu \quad \nu \pm d\nu/2 \quad (15)$$

where  $f_\nu d\nu$  gives the total flux within the frequency range  $\nu \pm d\nu/2$ . Astronomers commonly use ‘AB-magnitudes’ as a measure of flux density, which are defined as

$$m_{\text{AB}} = -2.5 \log f_\nu - 48.60. \quad (16)$$

The ‘absolute magnitude’  $M_X$  is defined as the magnitude an astronomical object would have, if you were to place it at 10pc away from us. Furthermore, ‘X’ usually denotes what part of the electromagnetic spectrum we are interested in, i.e. X=UV refers to the ultraviolet part of the spectrum, while X=R refers to the ‘red’ part of the spectrum. In the lecture slides, we gave some examples of numerical values of absolute magnitudes.

## B. Outline of this Course

The outline of these lectures is as follows. We infer from the intensity/temperature fluctuations in the Cosmic Microwave Background that the density of baryons was extremely homogeneous, with density fluctuations of order  $\delta\rho/\rho \sim 10^{-5}$ . We would like to address how we expect these tiny density fluctuations to evolve with cosmic time, and how we can connect these fluctuations with ‘fully’ grown stars, and galaxies that we observe around us today. More specifically:

- how do density fluctuations evolve as a function of time. How does this depend on the size/mass of the fluctuation? (Lectures 2+3)
- How is growth of a density fluctuation affected by gas pressure? (Lectures 2+3)
- How does the overall expansion affect the ‘growth’ of structure? (Lectures 2+3)
- Is there enough time to ‘grow’ galaxies? (Lectures 2+3)

- What happens to a pressureless fluid that collapses under the influence of gravity? (Lecture 4, 5)
- Statistical properties of cosmological density field, and statistical descriptions of galaxies. (Lecture 6-10)
- Given that we can measure the statistical properties of density fluctuations in the very young Universe, can we predict statistical properties of galaxies from this? (Lecture 6-10)
- What does ordinary gas do differently when it ‘collapsed’ to very high densities? Lecture (5,12)
- How can we test our theoretical framework for all these processes? What are observational challenges to our current understanding? (Lectures 12-16)
- How do galaxies form? (Lecture 16)
- How do stars form inside galaxies? (Lecture 17)
- How do black holes form from stars? Other formation mechanisms? (Lectures 18/19)

As you can see, in these lectures, we take a chronological approach in which we ‘descend’ into increasingly advanced stages of structure formation, from the early ‘linear’ evolution of density fluctuations, all the way to the collapse of a compact star supported by degenerate pressure into a black hole.

### III. SIMPLIFIED PICTURE OF GROWTH OF STRUCTURE: APPLICATION OF KIRCHHOFF THEOREM

Kirchhoff theorem states that for a spherically symmetric distribution of matter, the gravitational force at any given distance  $r$  from the center of this distribution is determined entirely by the mass at  $< r$ , and we can ignore the matter outside of  $r$ . We can apply this theorem to a density field  $\rho(\mathbf{x})$  to get some basic insight into growth of structure in the Universe. Consider a density field that is expanding.

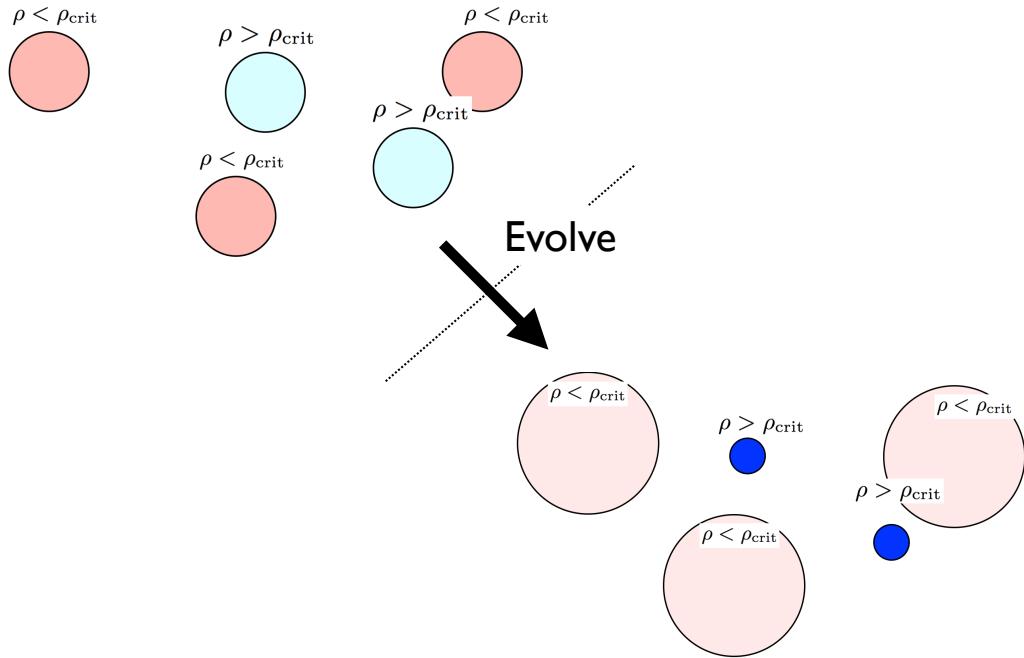


FIG. 2 Growth of structure as an application of Kirchhoff’s theorem.

Figure 2 shows a schematic representation of this density field in which there exist three spherical patches in which  $\rho < \rho_{\text{crit}}$  (i.e.  $\Omega < 1$ ), and two patches with  $\rho > \rho_{\text{crit}}$  (i.e.  $\Omega > 1$ ). Kirchhoff theorem now states that these five patches

do not care about their environments, and that they evolve as mini-Universes, each with their own density parameter. From the discussion in § II.A we know how these patches evolve with time: the red *underdense* patches are expected to expand faster than the Universe as a whole, while the blue *overdense* patches expand slower (and eventually even collapse). Over time, we expect underdense regions to become more underdense, while overdense regions become more overdense. This illustrates that gravity acts to amplify density contrast that exists in a matter density field.

(Think about why assuming these patches to be spherical is reasonable within the standard cosmological framework. Also discuss why Kirchhoff theorem should formally not apply to the mass distribution sketched above.)

#### IV. GROWTH OF SMALL PERTURBATIONS IN A EXPANDING & STATIC MEDIA

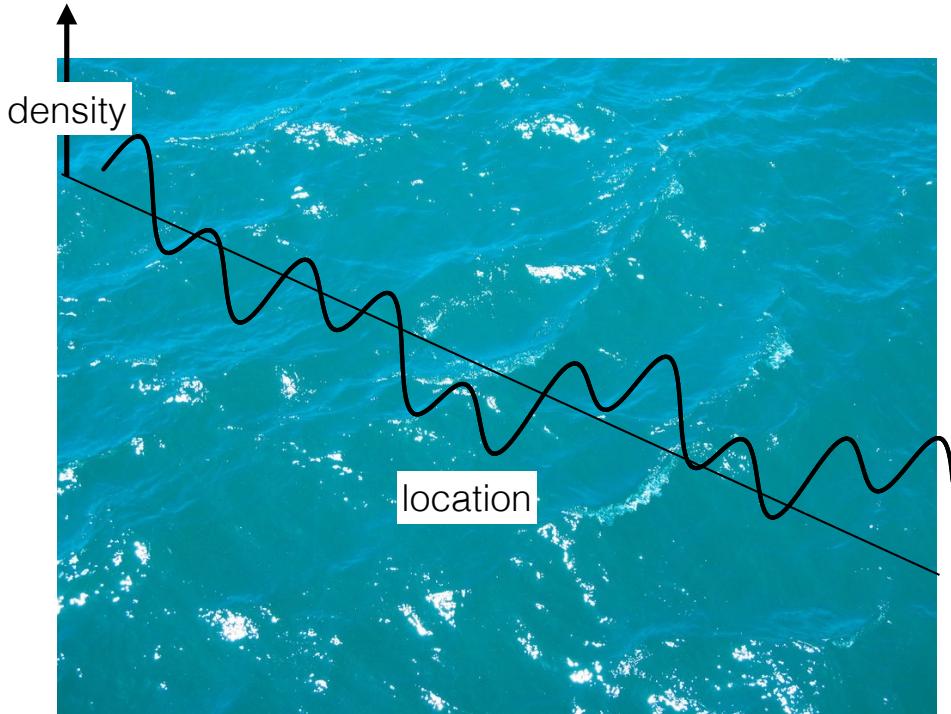


FIG. 3 Ripples on the surface of water in the pool represent tiny perturbations in the pool depth  $L(\mathbf{x}, t)$  at some arbitrary position  $\mathbf{x}$ . We can express  $L(\mathbf{x}, t) \equiv L + \delta L(\mathbf{x}, t)$ , where  $|\delta L(\mathbf{x}, t)|/L \ll 1$ , and  $L$  denotes the average depth. Solving the equations of motion for  $\delta L(\mathbf{x}, t)$  is generally easier when  $|\delta L(\mathbf{x}, t)|/L \ll 1$  (see text).

Structure formation in our Universe always represents a competition between gravity and processes that want to resist gravity (gas pressure, radiation pressure,...). The goal of this lecture is to derive an equation describing the time evolution of the density in a small perturbation in expanding & static media. This discussion closely follows Chapter 11.2 in Longair, but in these notes I present intermediate steps not given in this book.

Consider the surface of water in a pool. The average depth of the pool is  $L$ . There exist ripples on the surface of the water, which can be described by small perturbations from the average. That is, the depth of the pool at some arbitrary location,  $L(\mathbf{x})$ , equals  $L + \delta L(\mathbf{x})$ , where  $|\delta L(\mathbf{x})|/L \ll 1$ . This example represents a lower dimensional analogue of the early Universe, in which the 3-D gas density  $\rho(\mathbf{x})$  can be described accurately by  $\rho(\mathbf{x}) = \bar{\rho} + \delta\rho(\mathbf{x})$ . Observations of the CMB indicate that  $\delta\rho(\mathbf{x})/\bar{\rho} \sim 10^{-5} \ll 1$ . The reasons that we decompose  $\rho(\mathbf{x})$  into  $\bar{\rho} + \delta\rho(\mathbf{x})$  are that (i) we can simply assume that  $\bar{\rho}$  obeys the relevant equations (see below) without specifying its functional form, and (ii) solve for  $\delta\rho(\mathbf{x})$ , which is easier to solve for than  $\rho(\mathbf{x})$ , because we know that  $\delta\rho(\mathbf{x})/\bar{\rho} \ll 1$ , which allows us to simplify the relevant equations of motion. This approach is called *linear perturbation theory*. It will hopefully become more apparent what it is, when we start actually solving the equations.

## A. Basic Equations

We denote the density at a position  $\mathbf{x}$  and time  $t$  with  $\rho(\mathbf{x}, t)$ , the velocity of the an ‘ideal’ fluid with  $\mathbf{v}(\mathbf{x}, t)$ , the pressure of the fluid with  $p(\mathbf{x}, t)$ , and the gravitational potential with  $\phi(\mathbf{x}, t)$ . Throughout, I will drop writing the dependence on both  $\mathbf{x}$  and  $t$ . Pressure and density provide us with a complete description for an ideal fluid.

The three equations which describe the dynamics of this ideal fluid in a gravitational field are:

$$\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{v} = 0; \quad \text{Continuity equation} \quad (17)$$

The continuity equation ensures that flows of matter conserve mass. Furthermore,  $\frac{d}{dt} \equiv (\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla)$  denotes the ‘total’ time derivative which gives the time evolution of a quantity in a frame that is moving along with the fluid<sup>2</sup>.

$$\frac{d\mathbf{v}}{dt} = -\frac{1}{\rho} \nabla p - \nabla \phi; \quad \text{equation of motion/Euler equation} \quad (18)$$

The Euler equation is a more general way of writing Newton’s second law ( $F = ma$ ) in the presence of a gravitational potential. Especially the Euler equation illustrates why it is natural to write the equation using a total time derivative: this equation simply says that acceleration equals the sum of all forces (per unit mass).

$$\nabla^2 \phi = 4\pi G \rho \quad \text{Poisson equation.} \quad (19)$$

The Poisson equatin describes how matter sources the gravitational potential.

## B. The Unperturbed Equations

As we mentioned earlier, we will solve for the time-evolution of a density fluctuation  $\delta\rho(\mathbf{x}) = \bar{\rho} - \rho(\mathbf{x})$ , assuming that the ‘unperturbed’ fluid obeys all three fluid equations. We define the quantities of the ‘unperturbed’ fluid to be  $\bar{\rho}$ ,  $\mathbf{v}_0$ ,  $p_0$ , and  $\phi_0$ . These quantities must obey the continuity, Euler & Poisson equations, i.e.

$$\frac{d\bar{\rho}}{dt} + \bar{\rho} \nabla \cdot \mathbf{v}_0 = 0; \quad \text{Continuity equation} \quad (20)$$

$$\frac{d\mathbf{v}_0}{dt} = -\frac{1}{\bar{\rho}} \nabla p_0 - \nabla \phi_0; \quad \text{equation of motion/Euler equation} \quad (21)$$

$$\nabla^2 \phi_0 = 4\pi G \bar{\rho} \quad \text{Poisson equation.} \quad (22)$$

It is interesting to point out that in a static medium, these solutions do not exist beyond the solution  $\bar{\rho} = p_0 = \mathbf{v}_0 = 0$  (Google ‘Jeans swindle’). We will ignore this for now. We are most interested in the solution in an expanding medium.

## C. The Perturbed Equations

We define the quantities of the ‘perturbed’ fluid to be  $\rho \equiv \bar{\rho} + \delta\rho$ ,  $\mathbf{v} \equiv \mathbf{v}_0 + \delta\mathbf{v}$ ,  $p \equiv p_0 + \delta p$ , and  $\phi \equiv \phi_0 + \delta\phi$ . These quantities must also obey the continuity, Euler & Poisson equations. **The goal of the next few sections is to obtain an equation that describes the time evolution of  $\delta\rho$ .** We are going to just plug everything in, recalling that  $(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla) \equiv \frac{d}{dt}$ . In the lecture, we focussed on the continuity equation for the perturbed fluid, which reads

<sup>2</sup> In the book ‘Physical Cosmology’ by J. Peacock (Chapter 2.1) an interesting example of the relevance of the total derivative is given; ‘The changes experienced by an observed moving with the fluid are inevitably a mixture of temporal and spatial changes. If I start to feel rain as I cycle towards my destination, it might be a good idea to cycle harder in the hope of arriving before the downpour really starts, but it could also be that it is raining near my destination, and I should stop and wait for the local shower to finish. ...’

$$\frac{\partial \rho}{\partial t} + (\mathbf{v} \cdot \nabla) \rho + \rho \nabla \cdot \mathbf{v} = 0 \Rightarrow \quad (23)$$

$$\frac{(\partial \bar{\rho} + \delta \rho)}{\partial t} + ([\mathbf{v}_0 + \delta \mathbf{v}] \cdot \nabla)(\bar{\rho} + \delta \rho) + (\bar{\rho} + \delta \rho) \nabla \cdot (\mathbf{v}_0 + \delta \mathbf{v}) = 0 \quad (24)$$

We are going to explicitly write out all terms in this equation. I have color-coded some terms to simplify the analysis of this equation:

$$\begin{aligned} \frac{\partial \bar{\rho}}{\partial t} + \frac{\partial \delta \rho}{\partial t} + & \underset{I}{(\mathbf{v}_0 \cdot \nabla) \bar{\rho}} + \underset{II}{(\mathbf{v}_0 \cdot \nabla) \delta \rho} + \underset{III}{(\delta \mathbf{v} \cdot \nabla) \bar{\rho}} + \underset{IV}{(\delta \mathbf{v} \cdot \nabla) \delta \rho} + \\ & + \bar{\rho} \nabla \cdot \mathbf{v}_0 + \bar{\rho} \nabla \cdot \delta \mathbf{v} + \delta \rho \nabla \cdot \mathbf{v}_0 + \delta \rho \nabla \cdot \delta \mathbf{v} = 0 \end{aligned} \quad (25)$$

The *red terms* correspond to the unperturbed continuity equation (Eq 20), and so their sum is 0. The *blue terms* contain products of two arbitrarily small quantities, and can be ignored. Finally, the *green term* contains spatial derivatives of the unperturbed density field  $\bar{\rho}$ . We assume this unperturbed density field to be homogeneous, and we can therefore ignore its spatial derivative. Note that sum of the terms with the ‘I’ and ‘II’ under them simply correspond to the total derivative of  $\delta \rho$ . We are thus left with

$$\frac{d \delta \rho}{dt} + \bar{\rho} \nabla \cdot \delta \mathbf{v} + \delta \rho \nabla \cdot \mathbf{v}_0 = 0. \quad (26)$$

We can simplify this further if we define the over density  $\delta \equiv \delta \rho / \bar{\rho}$ . Substituting this into the equation above gives

$$\begin{aligned} \frac{d \bar{\rho} \delta}{dt} + \bar{\rho} \nabla \cdot \delta \mathbf{v} + \bar{\rho} \delta \nabla \cdot \mathbf{v}_0 &= \\ \delta \frac{d \bar{\rho}}{dt} + \bar{\rho} \frac{d \delta}{dt} + \bar{\rho} \nabla \cdot \delta \mathbf{v} + \bar{\rho} \delta \nabla \cdot \mathbf{v}_0 &= 0. \end{aligned} \quad (27)$$

Recall from the unperturbed continuity equation (Eq 20) that  $\frac{d \bar{\rho}}{dt} = -\bar{\rho} \nabla \cdot \mathbf{v}_0$ . The first and fourth term then cancel out, and we are left with

$$\frac{d \delta}{dt} = -\nabla \cdot \delta \mathbf{v}. \quad (28)$$

In **Exercise 1** you are asked to go through a similar analysis for the Euler and Poisson equations. The three equations we are left with are:

$$\begin{aligned} \frac{d \delta}{dt} &= -\nabla \cdot \delta \mathbf{v} && \text{Continuity equation} \\ \frac{d \delta \mathbf{v}}{dt} + (\delta \mathbf{v} \cdot \nabla) \mathbf{v}_0 &= -\frac{1}{\bar{\rho}} \nabla \delta p - \nabla \delta \phi; && \text{equation of motion/Euler equation} \\ \nabla^2 \delta \phi &= 4\pi G \delta \rho && \text{Poisson equation.} \end{aligned} \quad (29)$$

Note that these equations for the perturbed quantities closely resemble the original equations.

#### D. Switch to Comoving Coordinates

We next switch to co-moving coordinates  $\mathbf{r}_c$  which expand along with the Universe (see Eq 2), which relate to physical/proper coordinates  $\mathbf{r}_p$  via  $\mathbf{r}_p \equiv a(t)\mathbf{r}_c$ . Spatial derivatives in comoving coordinates are denoted with  $\nabla_c \equiv$

$a(t)\nabla$ . The velocity of the perturbed fluid element at  $\mathbf{r}_p$  is  $\mathbf{v} = \frac{\partial \mathbf{r}_p}{\partial t}$ .

$$\mathbf{v} = \frac{\partial \mathbf{r}_p}{\partial t} = \frac{\partial a(t)\mathbf{r}_c}{\partial t} = a(t) \frac{\partial \mathbf{r}_c}{\partial t} + \mathbf{r}_c \frac{\partial a}{\partial t}. \quad (30)$$

Recall  $\mathbf{v} = \delta\mathbf{v} + \mathbf{v}_0$ .

We can identify the second terms on the right hand side as the velocity arising from the expansion of the Universe. The first term on the other hand, denotes additional 'peculiar' velocity. We have  $\delta v = a(t) \frac{\partial \mathbf{r}_c}{\partial t} \equiv a(t)\mathbf{u}$ . If we substitute this into the perturbed continuity equation (first line in Eq 29), we get

$$\frac{d\delta}{dt} = -\nabla \cdot a\mathbf{u} = -\nabla_c \cdot \mathbf{u}$$

Continuity equation

(31)

We now turn our attention to the perturbed Euler equation (second line in Eq 29):

$$\begin{aligned} \frac{da\mathbf{u}}{dt} + (a\mathbf{u} \cdot \nabla)\mathbf{v}_0 &= -\frac{1}{\bar{\rho}}\nabla\delta p - \nabla\delta\phi \Rightarrow \\ a\frac{d\mathbf{u}}{dt} + \mathbf{u}\frac{da}{dt} + (a\mathbf{u} \cdot \nabla)\dot{a}\mathbf{r}_0 &= -\frac{1}{a\bar{\rho}}\nabla_c\delta p - \frac{\nabla_c\delta\phi}{a}. \end{aligned} \quad (32)$$

On the first line, we substituted  $\delta v = a(t)\mathbf{u}$ . In the second line, we replaced the proper/physical spatial derivatives with comoving derivatives **and** we replace  $\mathbf{v}_0 = \mathbf{r}_c \frac{da}{dt} \equiv \mathbf{r}_c \dot{a}$  (Eq 30).

The third term on the right hand side equals

$$(a\mathbf{u} \cdot \nabla)\dot{a}\mathbf{r}_c = \dot{a}\mathbf{u} \quad (33)$$

**One way to convince yourself may be by explicitly writing out each component of both the L.H.S and R.H.S.** With this simplification,

$$a\frac{d\mathbf{u}}{dt} + 2\dot{a}\mathbf{u} = -\frac{1}{a\bar{\rho}}\nabla_c\delta p - \frac{\nabla_c\delta\phi}{a}. \quad (34)$$

Now we take the comoving spatial derivative on both sides of this equation:

$$a\frac{d\nabla_c \cdot \mathbf{u}}{dt} + 2\dot{a}\nabla_c \cdot \mathbf{u} = -\frac{1}{a\bar{\rho}}\nabla_c^2\delta p - \frac{\nabla_c^2\delta\phi}{a}. \quad (35)$$

Note that we used that  $(\nabla_c \frac{d}{dt}) = (\frac{d}{dt} \nabla_c)$  (convince yourself). Next use Eq 31 which states that  $\nabla_c \cdot \mathbf{u} = -\frac{d\delta}{dt}$ , and we are left with

$$\begin{aligned} -a\frac{d^2\delta}{dt^2} - 2\dot{a}\frac{d\delta}{dt} &= -\frac{1}{a\bar{\rho}}\nabla_c^2\delta p - \frac{\nabla_c^2\delta\phi}{a} \Rightarrow \\ \frac{d^2\delta}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta}{dt} &= \frac{1}{a^2\bar{\rho}}\nabla_c^2\delta p + \frac{\nabla_c^2\delta\phi}{a^2}. \end{aligned} \quad (36)$$

#### E. The 'Final' Equation: A Differential Equation which determines $\delta(t)$

We get to our final solution in 3 steps

1. Get rid of the  $\delta p$  term by introducing the sound speed,  $c_s^2 \equiv \partial p / \partial \rho$ , which allows us to recast  $\delta p = c_s^2 \delta \rho = c_s^2 \bar{\rho} \delta$ .
2. Use the perturbed Poisson equation to replace  $\nabla_c^2 \delta \phi / a^2 = \nabla^2 \delta \phi = 4\pi G \delta \rho$ .

These two steps give us

$$\begin{aligned} \frac{d^2\delta}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta}{dt} &= \frac{c_s^2}{a^2\bar{\rho}}\nabla_c^2\delta\rho + 4\pi G\delta\rho \Rightarrow \\ \frac{d^2\delta}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta}{dt} &= \frac{c_s^2}{a^2}\nabla_c^2\delta + 4\pi G\bar{\rho}\delta. \end{aligned} \quad (37)$$

**3.** We seek solutions of the form  $\delta \propto \exp(i\mathbf{k}_c \cdot \mathbf{r}_c)$ , where  $\mathbf{k}_c \equiv a\mathbf{k}$  denotes the 'comoving wavenumber'<sup>3</sup>. This choice represents a general way of finding solutions for second order differential equations. While our solutions depend on  $\mathbf{k}_c$ , they do apply more generally if we consider perturbation of the form  $\delta \propto \exp(i\mathbf{k}_c \cdot \mathbf{r}_c)$  to represent only a single Fourier model of a more general density perturbation. Another way to look at this way of viewing the choice  $\delta \propto \exp(i\mathbf{k}_c \cdot \mathbf{r}_c)$  is that we consider the time evolution of a perturbation of 'size'  $R_c \sim 1/|\mathbf{k}_c|$ . Note that  $\mathbf{k}_c \cdot \mathbf{r}_c = a\mathbf{k} \cdot \mathbf{r}_c = \mathbf{k} \cdot \mathbf{r}_p$ , and  $\nabla_c^2/a^2 = \nabla^2$ . We therefore finally get:

$$\boxed{\frac{d^2\delta}{dt^2} + 2\frac{\dot{a}(t)}{a(t)}\frac{d\delta}{dt} = \delta(4\pi G\bar{\rho} - k^2 c_s^2)}, \quad (38)$$

where  $k \equiv 2\pi/\lambda$  denotes the wavenumber of the perturbation. We now have a second order differential equation for  $\delta(t)$ , which we can solve once we have specified the time-evolution of the scale factor  $a(t)$  (i.e. if we specify how the Universe expands in time), and the mean matter density  $\bar{\rho}$ . We will consider some examples in the next few sections.

Before proceeding, it is interesting to take a step back and try to think about why we used perturbation theory s to predict the time evolution of small perturbations: it is actually *not* that the equations for the perturbed quantities were easier. It is also *not* that we were able to replace  $\delta p = c_s^2 \delta \rho$  (we could have replaced  $P = c_s^2 \rho$ ). One of the main things is that perturbation theory allows us to really isolate the perturbation from the background solution, which adds to accuracy. For example, we could have taken a spatial derivative on both sides of the original Euler equation. On the right hand side, you can see that the  $\nabla^2\phi$ -term can be replaced with  $4\pi G\rho$  (as in the perturbation case). However, we are also left with a term of the form  $-\nabla_\rho^2 \nabla c_s^2 \rho$ . To simplify this term, we would need to do a perturbation type analysis. This is because  $\delta\rho/\rho \ll 1$ , which implies that a term of the form  $\nabla\rho$  is very close to zero, but not exactly. It is precisely the non-zero part which describes the evolution of the perturbation. So in the end, we would need to expand  $\rho$  around its equilibrium value  $\bar{\rho}$ , which comes down to a similarly tedious manipulation as we discussed above.

## F. Static Medium & The Jeans Length

First we focus on the static case  $\dot{a} = 0$ . We are then left with

$$\frac{d^2\delta}{dt^2} = A\delta, \quad A = (4\pi G\bar{\rho} - k^2 c_s^2) \quad (39)$$

We will try to find a solution of the form  $\delta = C \exp(\lambda t)$ . Then,  $\frac{d\delta}{dt} = \lambda\delta$  and  $\frac{d^2\delta}{dt^2} = \lambda^2\delta$ . We therefore have  $\lambda = \pm\sqrt{A}$ . Solutions with  $A < 0$  correspond to imaginary values of  $\lambda$ , and thus correspond to oscillatory solutions (i.e.  $\exp i\sqrt{A}t = \cos \sqrt{A}t + i \sin \sqrt{A}t$ ). On the other hand, solutions with  $A > 0$  correspond real values of  $\lambda$ . In this case

$$\delta(t) = C_1 \exp(\sqrt{A}t) + C_2 \exp(-\sqrt{A}t), \quad A \geq 0 \quad (40)$$

where the first/second solution shows a solution that is exponentially growing/decaying. This second solution is not physical: gravity will not drive apart overdense regions. The condition  $A \geq 0$  corresponds to  $4\pi G\bar{\rho} - k^2 c_s^2 > 0$ , i.e.

<sup>3</sup> The 'comoving' wavenumber of a photon that is emitted corresponds to the 'rest-frame' wavelength. For example, an photon with  $E = 1$  keV will always have a rest-frame energy (wavenumber corresponding to)  $E = 1$  keV. The proper energy/wavenumber will change as the background Universe expands

$k^2 c_s^2 < 4\pi G \bar{\rho}$ . We substitute  $k = 2\pi/\lambda$  to get  $4\pi^2 c_s^2 / \lambda^2 < 4\pi G \bar{\rho}$ , or

$$\lambda > c_s \sqrt{\frac{\pi}{G \bar{\rho}}} \equiv \lambda_J \quad (41)$$

, where  $\lambda_J$  denotes the ‘Jeans length’. Modes in the perturbation larger than  $\lambda_J$  will grow as  $\exp(\sqrt{A}t) \equiv \exp(t/t_{\text{coll}})$ , where we defined the *collapse time* as  $t_{\text{coll}} = \frac{1}{\sqrt{4\pi G \bar{\rho}}}$ . Modes shorter than the Jeans length do not collapse, instead the density oscillates around  $\bar{\rho}$  (see Fig 4).

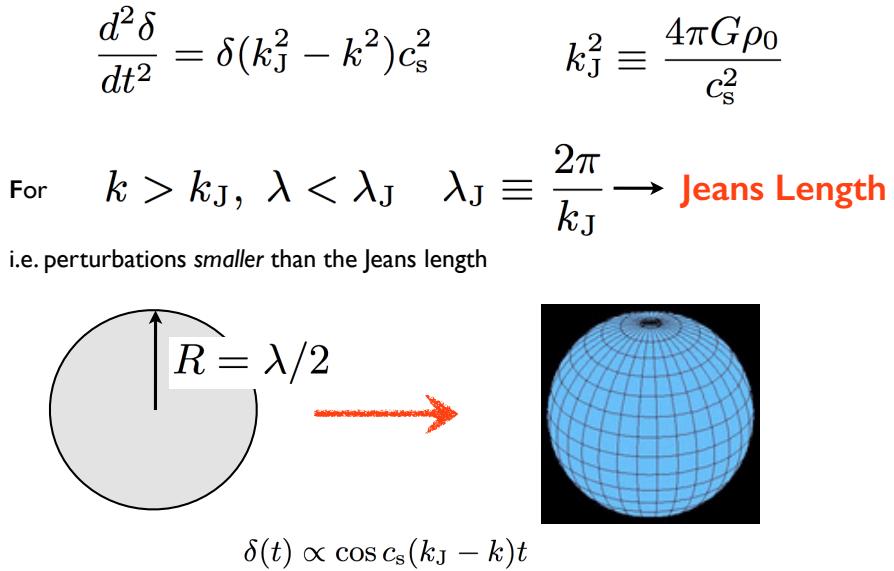


FIG. 4 This Figure illustrates how perturbations *smaller* than the Jeans length do not collapse under the influence of gravity. Instead, pressure forces counteract collapse and causes the density of the perturbation to oscillate in time.

We can similarly define a Jeans-mass as

$$M_J = \frac{4\pi}{3} \left( \frac{\lambda_J}{2} \right)^3 \bar{\rho} = \frac{\pi^{5/2}}{6G^{3/2} \bar{\rho}^{1/2}} \left( \frac{kT}{\mu m_p} \right)^{3/2}. \quad (42)$$

In the last step I used that  $p = nk_B T$  and  $\rho = nm_p \mu$ , in which  $n$  denotes the number density of particles,  $\mu$  their mean molecular weight<sup>4</sup>

Under these assumptions,  $c_s^2 = \frac{k_B T}{\mu m_p}$ . Substituting some numbers gives

$$M_J = 50 M_\odot \left( \frac{n}{10^3 \text{ cm}^{-3}} \right)^{-1/2} \left( \frac{T}{10 \text{ K}} \right)^{3/2}. \quad (43)$$

I have substituted some numbers for temperature and density which are typically encountered in so-called Giant Molecular Clouds, which is where stars are thought to form (we will discuss this in more detail in later lectures).

---

<sup>4</sup> See the piazza forum for an extended discussion of  $\mu$ .

## G. Linear Perturbation Theory in Expanding Media

In the previous section, we focussed on a static Universe/medium, i.e.  $\dot{a} = 0$  in Eq 38. Here, we focus on an expanding medium. Because the right hand-side of Eq 38 is unaffected, we again expect a transition from pressure-supported to non-pressure supported at the Jeans scale. For simplicity, we will assume that the perturbation of interest is much larger than the Jeans length, and that we can ignore the pressure term, and we will solve

$$\boxed{\frac{d^2\delta}{dt^2} + 2\frac{\dot{a}(t)}{a(t)}\frac{d\delta}{dt} = \delta 4\pi G\bar{\rho}}. \quad (44)$$

Because now  $\dot{a} \neq 0$ , we have a function of  $t$  in front of the term  $\frac{d\delta}{dt}$ . Recall from **Lecture 1** that the time evolution of the scale factor  $a(t)$  in a flat Universe is given by (see Eq 3 and Eq 4)

$$\left(\frac{\dot{a}}{a}\right)^2 \equiv H^2 = H_0^2 \left[ \frac{\Omega_m}{a^3} + \frac{\Omega_r}{a^4} + \Omega_\Lambda \right]. \quad (45)$$

It is common to analyse the solution for a Universe with  $\Omega_m = 1$  and  $\Omega_\Lambda = 0.0$  because (i) a simple functional form exists for  $a(t)$  (as we see below), and (ii) at high redshift this provides an excellent description of flat Universes with non-zero  $\Omega_\Lambda$  (because the relative importance of the vacuum to the Universal energy density becomes increasingly small). In this case,

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 a^{-3}. \quad (46)$$

We can solve this differential equation to get

$$a(t) = \left( \frac{3}{2} H_0 t + \frac{3}{2} C_1 \right)^{2/3}, \quad (47)$$

where  $C_1$  is an integration constant. We know that  $a(t=0) = 0$ , i.e. as  $t \rightarrow 0$   $a(t) \rightarrow 0$ , which implies that  $C_1 = 0$ . Also note that it is common to define that  $a(t_0) = 1$ , where  $t_0$  is the current age of the Universe. This immediately relates the age of the Universe to the Hubble parameter as

$$t_0 = \frac{2}{3H_0}. \quad (48)$$

Note that if we substitute the numerical value for the Hubble parameter of  $H_0 = 70$  km/s/Mpc, then we find  $t_0 = 9.3$  Gyr, which is younger than the age we would get for  $\Lambda$ -CDM (think about why). The expression for the Hubble parameter now becomes remarkably simple:

$$\frac{\dot{a}}{a} = \frac{2}{3t}. \quad (49)$$

Therefore the differential equation simplifies to

$$\frac{d^2\delta}{dt^2} + \frac{4}{3t} \frac{d\delta}{dt} = \delta 4\pi G\bar{\rho}. \quad (50)$$

For an expanding Universe, we can see that unperturbed quantities as the ‘mean quantities of the Universe, which we denote with a bar on top, i.e.  $\bar{\rho}$  denotes the mean, i.e. unperturbed, density. Then,  $\bar{\rho}_0$  denotes the unperturbed density at  $t = t_0$  (where  $t_0$  denotes the present time).

$$\frac{d^2\delta}{dt^2} + \frac{4}{3t} \frac{d\delta}{dt} = \delta 4\pi G\bar{\rho}. \quad (51)$$

The unperturbed density  $\bar{\rho}$  evolves with time because of the Expansion of the Universe, i.e.  $\bar{\rho} = \bar{\rho}_0 a^{-3}$ . Using Eq 5

we can write

$$4\pi G \bar{\rho} = 4\pi G \frac{\bar{\rho}_0}{\left(\frac{3}{2}H_0 t\right)^2} = 4\pi G \frac{\frac{3H_0^2}{8\pi G}}{\left(\frac{3}{2}H_0 t\right)^2} = \frac{2}{3t^2}, \quad (52)$$

where we used that  $\Omega_m = 1 \equiv \frac{\bar{\rho}_0}{\rho_{\text{crit},0}}$  and  $\rho_{\text{crit},0} = \frac{3H_0^2}{8\pi G}$  (see **Lecture 1**). All this simplifies the differential equation enormously:

$$\frac{d^2\delta}{dt^2} + \frac{4}{3t} \frac{d\delta}{dt} = \frac{2}{3t^2}. \quad (53)$$

We look for solutions of the form  $\delta = Ct^n$ :  $\frac{d\delta}{dt} = Cnt^{n-1}$ ,  $\frac{d^2\delta}{dt^2} = Cn(n-1)t^{n-2}$ .

$$Cn(n-1)t^{n-2} + \frac{4}{3}Cnnt^{n-2} = Ct^n \left( \frac{2}{3t^2} \right). \Rightarrow$$

$n(n-1) + \frac{4}{3}n = \frac{2}{3}.$

(54)

This equation has two solutions:  $n = 2/3$  and  $n = -1$ . The decaying solution with  $n = -1$  is not physical. So we are left with the solution  $\delta = Ct^{2/3} \propto a$ . This leads us to a very important conclusion: in a Universe that is undergoing expansion the density contrast of a plain wave perturbation much larger than the Jeans length increases as

$$\delta = \frac{\delta\rho}{\rho} \propto a \propto (1+z)^{-1} \quad (55)$$

This growth rate that occurs an an ‘algebraic’ (power-law) rate should be contrasted with the exponential growth of the static medium (see Eq 40). In the **exercise** of new week you will do a similar analysis in a flat Universe with a cosmological constant. Once the cosmological constant starts accelerating the overall dynamics of the Universe, not surprisingly, the overall the growth of the over density is suppressed.

## H. Implications for Growth of Structure in an Expanding Universe

Now that we understand that the density of perturbations grows under the influence of gravity as  $\delta \propto a \propto (1+z)^{-1}$ , we can make some seriously interesting statements of structure formation in the Universe. Specifically, we know from observations of the Cosmic Microwave Background that temperature fluctuations are of order  $\delta T/T \sim 10^{-5}$ . The plasma physics of a mixture of photons and ionized primordial gas is well understood, and shows that temperature fluctuations in the CMB are sourced by density fluctuations. We therefore expect that  $\delta \equiv \delta\rho/\rho \sim 10^{-5}$  when the CMB photons were first released, which occurred at  $z \sim 10^3$  (or  $a \sim 10^{-3}$ ). Since  $\delta \propto a$ , this implies that density is only expected to grow up to  $\delta \sim 10^{-2}$  at the present day (see Fig 5)!

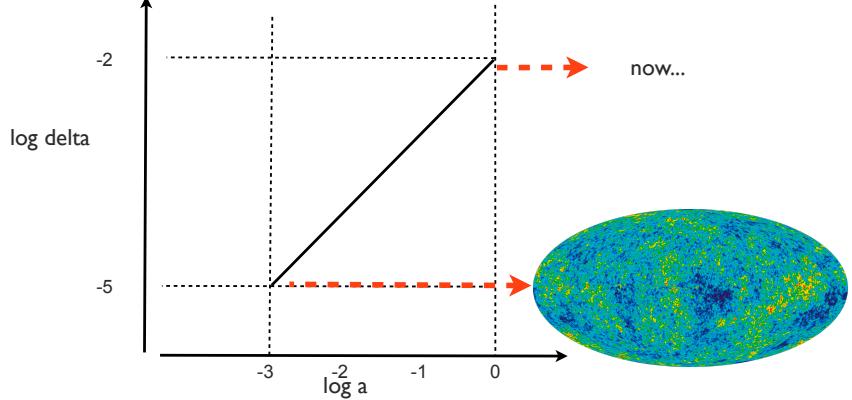
This is problematic. The number density of hydrogen nuclei (i.e. protons) averaged over the entire Universe is about  $10^{-7} \text{ cm}^{-3}$ . In contrast, the mean density of the interstellar medium is about  $1 \text{ cm}^{-3}$ . Galaxies are much more overdense than we can possibly account for with our simplified perturbation theory. This problem does not reflect the simplified nature of our analysis, and can also not be cured by changing the cosmology **Discuss why**. Instead, this problem is related to the Jeans length that we discuss earlier: the CMB photons were released when the Universe recombined. Prior to recombination, gas in the Universe was ionized. The free electrons efficiently scattered electromagnetic radiation, which tightly coupled the photons to the baryons. The combined photon-baryon fluid is a relativistic fluid, for which the sound speed is  $c_s = c/\sqrt{3}$ , which  $c$  denotes the speed of light. The Jeans mass prior to recombination is therefore tremendous:

$$M_J = \frac{4\pi}{3} \left( \frac{\lambda_J}{2} \right)^3 \bar{\rho} = \frac{\pi^{5/2}}{6} \frac{c_s^3}{G^{3/2} \bar{\rho}^{1/2}} \approx ([1+z]/10^3)^{-3/2} 10^{18} M_\odot, \quad z > z_{\text{rec}}. \quad (56)$$

The tremendous Jeans mass implies that no baryonic structures of astrophysical importance (galaxies, galaxy clusters) were allowed to start collapsing prior to  $z_{\text{rec}}$ . Instead, for any baryonic perturbation that existed prior to recombination, the radiation pressure would prevent them from collapsing. Instead, any perturbations would oscillate

$$\delta(t) \propto t^{2/3} \propto a(t) = \frac{1}{1+z}$$

Temperature fluctuations are sourced by density fluctuations, with  $\Delta\rho/\rho \sim \Delta T/T \sim 10^{-5}$



Perturbations can have grown only to an overdensity of about  $\sim 0.01$

FIG. 5

around the equilibrium value of about  $\delta\rho/\rho \sim 10^{-5}$ . **Unclear why this particular value?**. This is illustrated in Fig 6 which shows the pre-recombination evolution of any baryonic perturbation.

So how does this help us understand structure formation in the Universe? The previous section highlights how pressure halts the collapse of any *baryonic* perturbation. We know from a suite of observations that the mass density of the Universe is dominated by pressureless/collisionless dark matter. Because dark matter has no pressure, there is no associated Jeans length/mass. Dark matter structures actually can grow prior to  $z_{\text{rec}}$ . Figure 6 also shows the time-evolution of the density in dark matter of a perturbation. This density grows  $\delta_{\text{DM}} \propto a$  prior to recombination<sup>5</sup>. What is important is that as the Universe recombines, the baryonic density field fluctuates at a level of  $\delta\rho/\bar{\rho} \sim 10^{-5}$ . However, the dark matter density field can display much larger fluctuations. The radiation pressure prevents the baryons from flowing into dark matter overdensities, until the Universe recombines: at recombination, the sound speed in the baryons drops by  $\sim 5$  orders of magnitude (from  $c_s = c/\sqrt{3}$  to  $c_s = \sqrt{\frac{k_B T}{\mu m_p}}$ ). As a result, baryonic structures with masses  $\gtrsim 10^3 M_\odot$  are suddenly permitted to collapse into the gravitational potential wells sources by the dark matter. As we will discuss in later lectures, collapse of gas into dark matter halos alone is not enough to allow for the formation of the first galaxies etc. It is however, an essential requirement. *If it were not for this dark matter, structure in our Universe could not have evolved sufficiently yet to form stars, galaxies, and all we see around us.*

## V. SPHERICAL TOP-HAT MODEL

Our previous analysis described the time-evolution of a density perturbation. Our analysis assumed that the perturbed quantities were much smaller than the unperturbed (equilibrium) quantities. That is, our linear perturbation analysis starts to break down when  $\delta\rho/\bar{\rho}$  becomes of order unity. Since known astronomical objects such as stars and galaxies are much more overdense than that, we need to go beyond linear perturbation theory. Given the complexity of linear perturbation theory, it may not be surprising that there generally are no analytic solutions that are valid for

---

<sup>5</sup> The precise time-evolution of dark matter density fluctuations also depends on their size. We will revisit this in detail in Lecture~15

## Structure Formation: the need for Dark Matter

The previous discussion provides motivation of the statement that we need dark matter - or generally a pressureless fluid of particles with finite mass - to explain structure formation in the Universe.

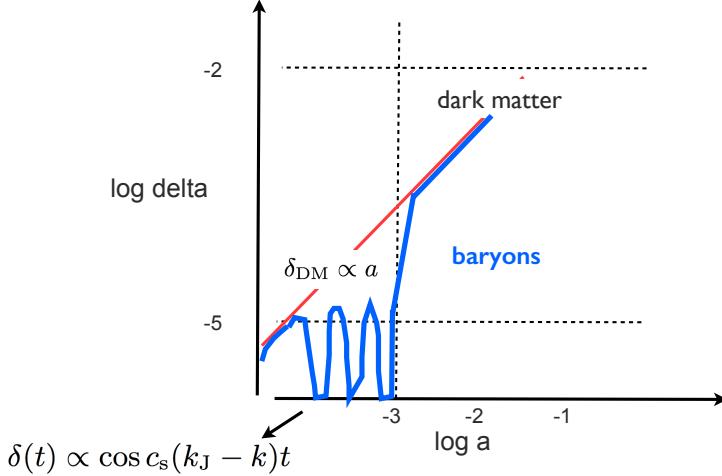


FIG. 6

all densities. The only case that we can solve completely analytically is the so-called ‘spherical top-hat’ model. In this model, the perturbation is a perfect sphere of radius  $R$ , and has a uniform density inside of it.

Before we start our full analysis, it is useful to repeat two previously encountered equations, which describe the background Einstein-de Sitter universe with  $(\Omega_\Lambda, \Omega_m) = (1.0, 0.0)$  include

$$a(t) = \left(\frac{3H_0 t}{2}\right)^{2/3}, \quad \bar{\rho}(t) = \frac{1}{6\pi G t^2}. \quad (57)$$

These quantities will be useful to compare the density evolution of the top-hat to.

The radius of the sphere evolves as

$$\ddot{R} = -\frac{GM}{R^2}, \quad (58)$$

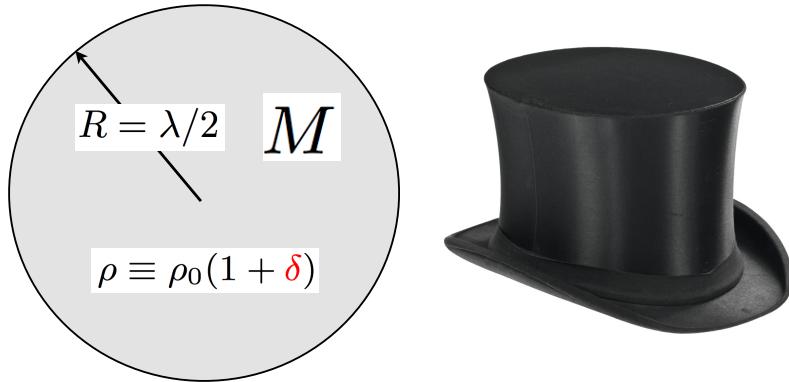
where  $M$  is the total mass inside the sphere, where  $M = \frac{4}{3}\pi R_i^3 \bar{\rho}_i$ , where  $\bar{\rho}_i$  denotes the average, or unperturbed, density of the Universe<sup>6</sup> at some early time  $t_i$  (when the density contrast between the perturbation and the Universe as a whole was still tiny). Solving this equation is not trivial. We first multiply both sides by  $\dot{R}$

$$\dot{R}\ddot{R} = -\dot{R}\frac{GM}{R^2}, \quad (59)$$

---

<sup>6</sup> This relation is not strictly obeyed. Technically, the perturbation is overdense relative to the background Universe (otherwise it would expand with the Universe, which for Einstein-de Sitter means that the perturbation becomes stationary when  $R \rightarrow \infty$ ), and  $M = \frac{4}{3}\pi R_i^3 \bar{\rho}_i(1 + \delta_i)$ . Because  $\delta_i \ll 1$  at  $t_i$  we can make this approximation.

## Non-linear Evolution: Spherical Top-Hat Model



The density inside the sphere is constant, and this model is referred to as the ‘spherical top-hat model’.

FIG. 7 The ‘top-hat’ model is the only model in which we can analytically solve for the time-evolution of a density perturbation. Also shown is an actual top-hat.

which can be written as

$$\frac{1}{2} \frac{d}{dt} (\dot{R})^2 = \frac{d}{dt} \frac{GM}{R}. \quad (60)$$

We therefore have

$$\frac{1}{2} \dot{R}^2 - \frac{GM}{R} = E, \quad (61)$$

where  $E$  is an integration constant. Note however, that the left hand side of this equation denotes the *total energy* (kinetic + potential) per unit mass for a mass element at  $R$  (just as in Fig 1). In **the assignment** you will be asked to show that the following *parametrised* solution indeed obeys Equation (6) above.

$$\begin{aligned} R &= A(1 - \cos \theta) \\ t &= B(\theta - \sin \theta) \\ A^3 &= GM B^2. \end{aligned} \quad (62)$$

Equation (7) shows:

- The sphere reaches its maximum at  $\theta = \pi$  and time  $t = \pi B$ .
- Sphere completely collapses at  $\theta = 2\pi$  and time  $t = 2\pi B$ .
- To first order in  $\theta$  we have  $R = A\theta^2/2$  ( $\cos \theta = 1 - \frac{\theta^2}{2!} + \dots$ ), and  $t = B\theta^3/6$  ( $\sin \theta = \theta - \frac{\theta^3}{3!} + \dots$ ). Thus, to first order we have

$$R(t) = \frac{A}{2} \left( \frac{6t}{B} \right)^{2/3} = \frac{(GM)^{1/3}}{2} (6t)^{2/3} = \left( \frac{GM36t^2}{8} \right)^{1/3}. \quad (63)$$

Substitute  $M = \frac{4}{3}\pi R_i^3 \bar{\rho}_i$ , with  $\bar{\rho}_i = \frac{1}{6\pi G R_i^2}$ , and we are left with

$$R(t) = R_i \left( \frac{t}{t_i} \right)^{2/3}. \quad (64)$$

So initially the sphere increases as  $R(t) \propto t^{2/3} \propto a$ , and simply expands along with the Universe.

- The density contrast  $\delta \equiv \delta\rho/\rho_m$  equals:

$$\begin{aligned} \delta_m &= \frac{\rho_{\text{sphere}}}{\rho_m} - 1 = 6\pi G t^2 \frac{M}{\frac{4}{3}\pi R^3} - 1 = \\ \frac{9GMt^2}{2R^3} - 1 &= \frac{9GM}{2} \frac{B^2(\theta - \sin\theta)^2}{A^3(1 - \cos\theta)^3} - 1 = \frac{9}{2} \frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3} - 1. \end{aligned} \quad (65)$$

The features discussed above are also apparent in the Fig 8 shown below.

### Non-Linear Density Evolution. Spherical Top-Hat.

The complete evolution of a ‘spherical top hat’ was described by parameterized solutions.

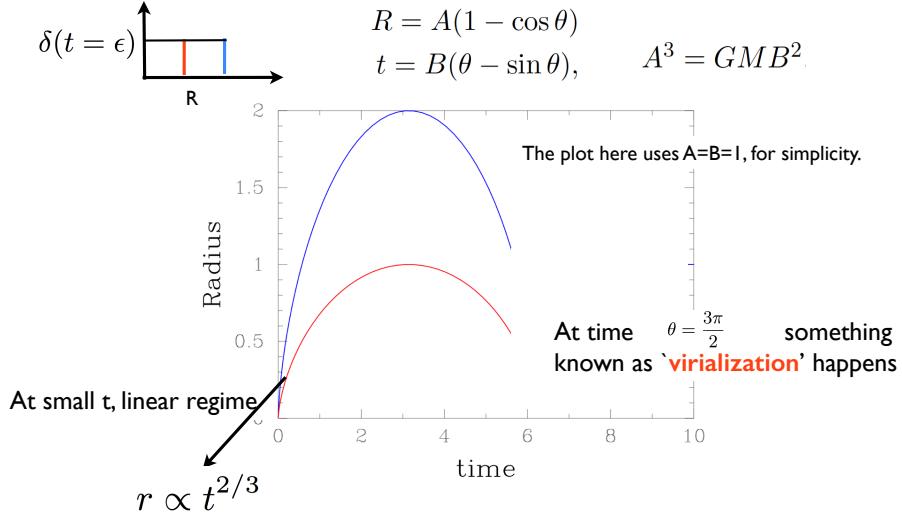


FIG. 8

Our next goal is to show that this non-linear evolution of the sphere connects perfectly to the solution we obtained previously from linear perturbation theory (given by Eq 55). To do this we Taylor expand the expression for  $\delta_m$  in  $\theta$ . First, we Taylor expand  $\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \dots$ ,  $\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + \dots$ . Substituting we get

$$\begin{aligned}
\delta_m &= \frac{9}{2} \frac{(\theta - \sin \theta)^2}{(1 - \cos \theta)^3} - 1 = \frac{9}{2} \frac{\left(\frac{\theta^3}{3!} - \frac{\theta^5}{5!} + \dots\right)^2}{\left(\frac{\theta^2}{2!} - \frac{\theta^4}{4!} + \dots\right)^3} - 1 = \\
&\quad \frac{9}{2} \frac{\left(\frac{\theta^3}{3!}\right)^2 \left(1 - \frac{\theta^2}{20} + \dots\right)^2}{\left(\frac{\theta^2}{2}\right)^3 \left(1 - \frac{\theta^2}{12} + \dots\right)^3} - 1 = \frac{\frac{9}{36}}{\frac{2}{8}} \frac{1 - \frac{2\theta^2}{20} + \dots}{1 - \frac{3\theta^2}{12} + \dots} - 1 = \\
&\quad \left(1 - \frac{2\theta^2}{20} + \dots\right) \left(1 + \frac{3\theta^2}{12} + \dots\right) - 1 = \frac{3}{20} \theta^2 + \dots = \frac{3}{20} \left(\frac{6t}{B}\right)^{2/3} + \dots
\end{aligned} \tag{66}$$

To first order, the density inside the sphere goes as  $\delta_m \propto t^{2/3} \propto a$ , exactly as we obtained in Eq 55.

We also know that collapse occurs at  $t = 2\pi B$ . In the linear regime this would correspond to:

$$\boxed{\delta_{\text{lin}} = 1.69}. \tag{67}$$

An object is said to collapse/virialize the moment, its linearly extrapolated density reaches  $\delta_{\text{lin}}$ . This number  $\delta_{\text{lin}} = 1.69$  will come in many applications. The sub-script ‘lin’ explicitly shows that this overdensity is obtained from linear theory.

Figure 8 also indicates that at  $\theta = \frac{3\pi}{2}$  ‘virialization’ happens. If all matter was on perfectly radial orbits, then the sphere would collapse indefinitely into a point. In reality, the orbits of individual particles (atoms, nuclei, and/or dark matter) are not perfectly radial. These particles will fall down to  $r = 0$  at a large velocity, miss  $r = 0$ , and then shoot outward at again a high velocity. This gives rise to complicated dynamics, the end result of which being that the sphere reaches an equilibrium state at a finite radius. The sphere is said to ‘virialize’. We discuss this process in more detail below, because virialization directly determines the extend of so-called dark matter halos (to be discussed in later lectures) which is where galaxies form and reside. Virialization therefore also indirectly sets the extend of galaxies.

## VI. VIRIALIZATION

Virialization represents the equilibrium end-state of a density perturbation that was allowed to collapse under the influence of gravity. In our Universe, the end-state of collapse of a dark matter structure is a ‘dark matter halo’. The dark matter halos form the gravitational potential wells in which baryons can collect to form galaxies, and thus play a central role in galaxy formation. First, we will derive the ‘virial theorem’, which provides a general relation between the total potential & kinetic energy of any self gravitating system. We then derive some basic properties of virialized dark matter halos, and then discuss how the evolution of baryons differs. This last discussion represents an important step toward understanding galaxy formation.

### A. The Virial Theorem

Consider a system of  $N$  particles. The acceleration of particle ‘i’ due to the gravitational pull of all other particles is given by

$$\ddot{\mathbf{r}}_i = - \sum_{j \neq i} \frac{Gm_j \mathbf{e}_{ij}}{|\mathbf{r}_i - \mathbf{r}_j|^2} = - \sum_{j \neq i} \frac{Gm_j (\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} \tag{68}$$

where we have used that  $\mathbf{e}_{ij} \equiv \frac{(\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|}$  denotes a vector of unit length in the direction of the separation between particles ‘i’ and ‘j’. Next, we take the scalar (‘dot’) product with  $m_i \mathbf{r}_i$  on both sides of the equality sign:

$$m_i \mathbf{r}_i \cdot \ddot{\mathbf{r}}_i = - \sum_{j \neq i} \frac{Gm_i m_j \mathbf{r}_i \cdot (\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} \tag{69}$$

In order to rewrite Eq 69 we first compute

$$\frac{d^2}{dt^2}(\mathbf{r}_i \cdot \mathbf{r}_i) = \frac{d}{dt}\left(\frac{d}{dt}\mathbf{r}_i \cdot \mathbf{r}_i\right) = \frac{d}{dt}(2\ddot{\mathbf{r}}_i \cdot \mathbf{r}_i) = 2\ddot{\mathbf{r}}_i \cdot \mathbf{r}_i + 2\dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i. \quad (70)$$

We therefore can recast Eq 69 as:

$$\frac{1}{2} \frac{d^2}{dt^2}(\mathbf{r}_i \cdot \mathbf{r}_i) - \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i = - \sum_{j \neq i} \frac{Gm_j \mathbf{r}_i \cdot (\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} \quad (71)$$

Now sum over all particles ‘i’

$$\frac{1}{2} \sum_i m_i \frac{d^2}{dt^2}(\mathbf{r}_i \cdot \mathbf{r}_i) - \sum_i m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i = - \sum_i \sum_{j \neq i} \underbrace{\frac{Gm_i m_j \mathbf{r}_i \cdot (\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3}}_{\equiv M_{ij}}. \quad (72)$$

The quantity under the right hand side that is being summed over ‘i’ and ‘j’ is an  $N \times N$  matrix  $\mathbf{M}$  whose diagonal elements ‘ $M_{ii}$ ’ (or ‘ $M_{jj}$ ’) are zero (because  $\mathbf{r}_i - \mathbf{r}_i = 0$ ).

We introduce a new matrix  $\mathbf{F}$  with elements

$$\begin{aligned} F_{ij} &\equiv M_{ij} + M_{ji} = \frac{Gm_i m_j \mathbf{r}_i \cdot (\mathbf{r}_i - \mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|^3} + \frac{Gm_j m_i \mathbf{r}_j \cdot (\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_j - \mathbf{r}_i|^3} = \\ &= \frac{Gm_i m_j}{|\mathbf{r}_j - \mathbf{r}_i|^3} (\mathbf{r}_i \cdot (\mathbf{r}_i - \mathbf{r}_j) + \mathbf{r}_j \cdot (\mathbf{r}_j - \mathbf{r}_i)) = \frac{Gm_i m_j}{|\mathbf{r}_j - \mathbf{r}_i|^3} (\mathbf{r}_i - \mathbf{r}_j)^2 = \frac{Gm_i m_j}{|\mathbf{r}_j - \mathbf{r}_i|} \end{aligned} \quad (73)$$

When we sum over all elements of  $\mathbf{F}$  we have

$$\sum_i \sum_{j \neq i} F_{ij} = \sum_i \sum_{j \neq i} (M_{ij} + M_{ji}) = 2 \sum_i \sum_{j \neq i} M_{ij}, \quad (74)$$

where in the last step we used that in the second step, we are summing over all elements over the matrix twice (we are only doing it in a different order. In the lecture I explicitly wrote out some terms up to  $i = j = 4$ . It was then clear that each matrix element  $M_{ij}$  appeared twice). We can therefore replace the right hand side of Eq 72 with  $\frac{1}{2} \sum_i \sum_{j \neq i} F_{ij}$ . Therefore, Eq 72 becomes

$$\frac{1}{2} \sum_i m_i \frac{d^2}{dt^2}(\mathbf{r}_i \cdot \mathbf{r}_i) - \sum_i m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i = -\frac{1}{2} \sum_i \sum_{j \neq i} \frac{Gm_i m_j}{|\mathbf{r}_j - \mathbf{r}_i|}. \quad (75)$$

The second term on the left hand side is twice the kinetic energy of the system:  $T = \frac{1}{2} \sum_i m_i v_i^2$ . The term on the left hand side is the total potential energy  $U$ : the factor  $\frac{1}{2}$  accounts for the fact that the double summation actually double counts the potential energy<sup>7</sup>. We therefore have

$$\frac{1}{2} \sum_i m_i \frac{d^2}{dt^2}(\mathbf{r}_i \cdot \mathbf{r}_i) - 2T = U. \quad (76)$$

We have now proven the virial theorem which states that if  $\frac{1}{2} \sum_i m_i \frac{d^2}{dt^2}(\mathbf{r}_i \cdot \mathbf{r}_i) = 0$  then

$$\boxed{2T = |U|} \quad \text{Virial theorem.} \quad (77)$$

<sup>7</sup> To clarify further: in the double sum we add the potential energy of particle ‘i’ due to particle ‘j’ to the same potential energy of particle ‘j’ due to particle ‘i’. We are then double counting the potential energy: if these two particles were the only two particles of the system, the total potential energy of the system would be only  $U = -\frac{Gm_1 m_2}{|\mathbf{r}_1 - \mathbf{r}_2|}$  because this is how much energy we have to insert into the system to increase the distance between the two particles to  $\infty$  at which point the total energy of the system would be zero.

While it is not possible to proof that  $\sum_i m_i \frac{d^2}{dt^2}(\mathbf{r}_i \cdot \mathbf{r}_i) = 0$ , there are good reasons why  $\frac{1}{2} \sum_i m_i \frac{d^2}{dt^2}(\mathbf{r}_i \cdot \mathbf{r}_i) = 0$ , it is possible to argue that in a time averaged sense, this term will go to 0 provided we average over a long enough time. We show this next. We define the quantity  $G = \frac{d}{dt} \sum_i m_i \mathbf{r}_i \cdot \dot{\mathbf{r}}_i = 2 \sum_i m_i \mathbf{r}_i \cdot \dot{\mathbf{r}}_i$ . This quantity is known is the 'virial'. The time average of the time-derivative of the virial is given by

$$\left\langle \frac{dG}{dt} \right\rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt' \frac{dG}{dt'} = \lim_{T \rightarrow \infty} \frac{1}{T} [G(T) - G(0)]. \quad (78)$$

The virial  $G(t)$  can continuously change with time. However, for a system in equilibrium we do not expect  $G(T)$  to blow up to infinity at some point in time (this would not really be an equilibrium). If this is the case then  $[G(T) - G(0)]$  should remain finite, whereas  $T$  can be increased arbitrarily, in which case  $\frac{dG}{dt} \rightarrow 0$  and the virial theorem is satisfied<sup>8</sup>.

## B. Virialization

As we mentioned before (and as was shown in Fig 8), virialization occurs at  $R_{\text{vir}} = 0.5R_{\text{max}}$  in the spherical top-hat model. We can see this as follows: the total energy in particles inside the top-hat is equal to the potential energy of the perturbation at turn-around (because then the kinetic energy is 0). That is,

$$E_{\text{tot}} = -\frac{3GM^2}{5R_{\text{max}}}. \quad (79)$$

When the sphere contracts to  $R = 0.5R_{\text{max}}$  then the total binding energy is  $U_{\text{new}} = -\frac{6GM^2}{5R_{\text{max}}}$ . The additional  $\Delta U = -\frac{3GM^2}{5R_{\text{max}}}$  that the sphere lost has been transformed into kinetic energy  $T$ . We therefore have  $T = \frac{3GM^2}{5R_{\text{max}}}$  and  $U_{\text{new}} = -\frac{6GM^2}{5R_{\text{max}}}$  which satisfies the virial theorem. The radius at which virialization occurs is the 'virial radius', denoted with  $R_{\text{vir}}$ .

We now want to compute the *real* overdensity of the perturbation at *collapse* (which is defined to occur at  $\theta = 2\pi$  in Eq 62, i.e. when  $R = 0$ ). We denote this overdensity with  $\delta_{\text{non-lin}}$ , where 'non-lin' stands for 'non-linear', which is meant to explicitly distinguish it from the linear overdensity at collapse  $\delta_{\text{lin}} = 1.69$  derived earlier (see Eq 67). To get to the answer, we cannot simply plug  $\theta = 2\pi$  into the expression for  $\delta_m$  (see Eq 66), because  $R = 0$  the density diverges, while in reality the sphere virialized at  $R = 0.5R_{\text{max}}$ . The mass density of the sphere thus remains fixed at the value it reached at  $\theta = 3\pi/2$ . We thus must evaluate  $\delta_{\text{non-lin}}$  by evaluating  $\rho_{\text{sphere}}$  at  $\theta = 3\pi/2$  and  $\bar{\rho}$  at  $2\pi$ . The density of the sphere  $\rho_m$  is therefore  $2^3 = 8 \times$  larger than it was at turnaround. Similarly,  $\bar{\rho}$  at  $t_{\text{collapse}} = 2\pi B$  decreased by a factor of  $2^2 = 4$  compared to the value at turn-around (because  $\bar{\rho} \propto t^{-2}$ , see Eq ??). We therefore have

$$\begin{aligned} \delta_{\text{non-lin}}^{\text{collapse}} &= \frac{\rho_{\text{sphere}}^{\text{virialization}}}{\rho_m^{\text{collapse}}} - 1 = \frac{8\rho_{\text{sphere}}^{\text{turn-around}}}{\frac{1}{4}\rho_m^{\text{turnaround}}} - 1 = \\ &32 \frac{\rho_{\text{sphere}}^{\text{turn-around}}}{\rho_m^{\text{turnaround}}} - 1 = 32(\delta_{\text{non-lin}}^{\text{turn-around}} + 1) - 1 \sim 32(5.55) - 1 \sim 178, \end{aligned} \quad (80)$$

where we used that the non-linear overdensity at turn-around was  $\delta = 4.55$  (which can be seen by plugging in  $\theta = \pi$  into Eq 65). An object virializes with  $\delta_{\text{non-lin}} \sim 178$  times the mean density of the Universe at that moment. This allows us to derive some useful physical parameters of dark matter halos as a function of their mass, as discussed below.

<sup>8</sup> Even this argument is not entirely convincing, one can imagine that it is not reasonable to have  $T \gg$  the age of the Universe so in practise even  $T$  will be finite.

### C. Useful Quantities Related to Virialization

From the discussion above we know that the *virial radius* of an object of mass  $M$  that collapsed at redshift  $z$  can be derived from  $M = \delta_{\text{non-lin}} \times \bar{\rho}^{4\pi} R_{\text{vir}}^3$ , i.e:

$$R_{\text{vir}} = 1 \text{ kpc} \left( \frac{M}{10^8 M_\odot} \right)^{1/3} \left( \frac{1+z}{10} \right)^{-1}, \quad (81)$$

where I have substituted some (arbitrary) numbers to get some sense for their magnitude. We can also define the *circular velocity*:

$$v_{\text{circ}} \equiv \sqrt{\frac{GM}{R_{\text{tot}}}} \approx 20 \text{ km s}^{-1} \left( \frac{M}{10^8 M_\odot} \right)^{1/3} \left( \frac{1+z}{10} \right)^{1/2}. \quad (82)$$

which provides a measure of the typical velocity dispersion of dark matter and gas particles inside the dark matter halo of interest. **Assignment: show that  $v_{\text{circ}}$  also represents the velocity at which matter crashes onto the virial radius.** Finally, we can define the *virial temperature*

$$T_{\text{vir}} \equiv \frac{m_p v_{\text{circ}}^2}{2k_B} \approx 2 \times 10^4 \text{ K} \left( \frac{M}{10^8 M_\odot} \right)^{2/3} \left( \frac{1+z}{10} \right). \quad (83)$$

which is the temperature to which the gas gets shock-heated once it reaches the virial radius. I will describe this process in more detail below.

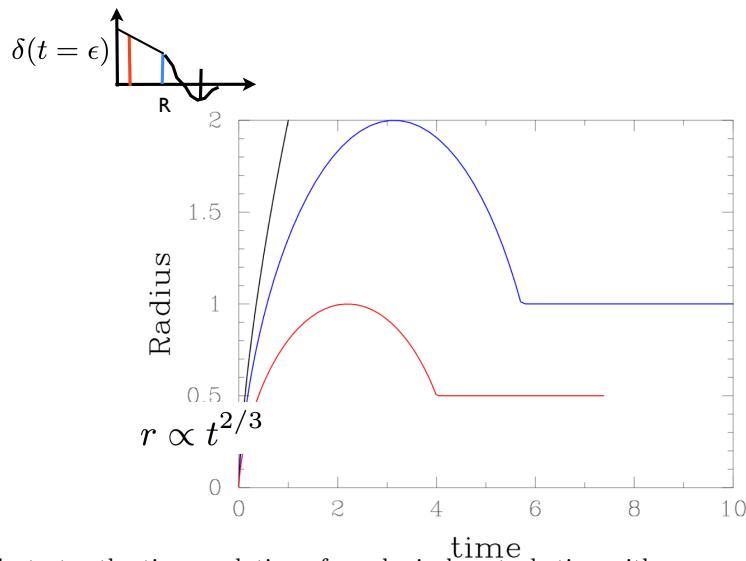


FIG. 9 This Figure illustrates the time evolution of a spherical perturbation with a more general density profile  $\rho(r)$  which decreases with  $r$ . For this perturbation - which is more realistic than the top-hat model - the inner regions of the perturbation collapse and virialize first, and continued accretion occurs onto the central collapsed object.

### D. Virialization beyond the Top-Hat Model

An important aspect of the spherical top-hat model is that all matter within this model turns around, and collapses/virializes at the same time. Figure 9 illustrates that if you have a more general (spherically symmetric) density profile  $\rho(r)$  which peaks at  $r = 0$  and then gradually decreases, that then turn-around and collapse occurs from the inside-out: the enclosed average density also decreases with radius, which means that matter shells initially at some radius  $r_i$  behave as higher density Universe, which turns around and recollapses earlier. In this scenario, which is

more realistic than the spherical top-hat model, a virialized central object exists, and matter continues to ‘rain down’ on it at later times.

### E. Virialization of Baryons: A First Taste of Gastrophysics

The baryons - we refer to the baryonic fluid as ‘gas’ from now on - behave a bit different than the dark matter. Gas is collisional and has a pressure. This modifies its behavior compared to the dark matter at virialization. Consider the more general perturbation  $\rho(r)$  which we discussed in the previous section (see § VI.D). Consider at time  $t$  at which point a central virialized object exists of radius  $R_{\text{vir}}$  and mass  $M$ . You will be asked in the **assignment** to show that the gas will be falling in at speed

$$v_{\text{infall}} = \sqrt{\frac{GM}{R_{\text{vir}}}} \equiv v_{\text{circ.}} \quad (84)$$

Depending on the total enclosed mass, this infall velocity is typically of order a few tens to a few hundreds of km s<sup>-1</sup>. So the inflowing gas will ‘crash’ into the stationary gas. What happens next depends on how  $v_{\text{infall}}$  compares to the sound speed ( $c_s$ ) of the infalling gas:

- If  $v_{\text{infall}} > c_s$ , then infall is ‘*supersonic*’. The infalling gas has no way of ‘knowing’ of the presence of stationary gas, and encounters the static gas at full speed. The ordered inflowing motion of the gas is converted into random thermal motions in a ‘*shock*’. The gas gets ‘shock-heated’ to

$$T_{\text{shock}} = \frac{m_p}{3k_B} v_{\text{infall}}^2, \quad (85)$$

which we obtained by setting  $\frac{3}{2}k_B T \equiv \frac{1}{2}m_p v_{\text{infall}}^2$ . The temperature of the shock heated gas corresponds to the virial temperature introduced earlier (see Eq 83).

- If  $v_{\text{infall}} < c_s$ , then infall is ‘*subsonic*’. Sound waves preceding the infalling gas reflect off the stationary gas, which increases the pressure outside the stationary gas, which causes the infalling gas to gradually slow down before joining the stationary gas.

Once the gas has been shock-heated to the virial temperature, it does not stay in this configuration. Instead, the gas cools. Once it cools, it loses hydrostatic support against gravity, and collapses deeper into the potential well generated by the virialized dark matter halo. Gas cooling thus allows gas to collapse to even higher densities, and eventually form stars & galaxies. We discuss this gas cooling in more detail below.

### F. Cooling of Virialized Gas

The main cooling processes that will be relevant for understanding galaxy formation are *two-body radiative processes*, in which the gas loses thermal energy by emitting photons which are created as a result of two-body interactions (i.e. collisions). There are four two-body processes that are relevant. These include (*Credit*: this discussion follows lectures by F. van den Bosch):

- **Free-free emission/Brehmstrahlung.** Occurs when a free electron is accelerated by a nearby ion. The accelerated electron radiates (power is given by the Larmor formula), which causes the electron to lose energy. Overall, the reduced thermal energy of electrons translates to a reduced thermal energy of ions and atoms, and thus cools the gas.
- **Free-bound/recombination emission.** When a free electron is captured by an ion, its kinetic + binding energy is radiated away. Only the loss of kinetic energy counts towards cooling of the gas, as the binding energy was counted already when the electron was detached from the atom.
- **Bound-free/collisional ionization emission.** A free electron collides with a hydrogen atom, which knocks the electron off the atom, at the expense of the kinetic energy of the free electron.
- **Bound-bound/collisional excitation emission.** A free electron collides with a hydrogen atom, which puts the atom from its ground state into one of its excited states, again at the expense of the kinetic energy of the free electron. The excited atom radiatively (i.e. by emitting radiation) transitions back to its ground state.

## The Cooling Function

The cooling function denotes the rate at which radiation cools the gas.

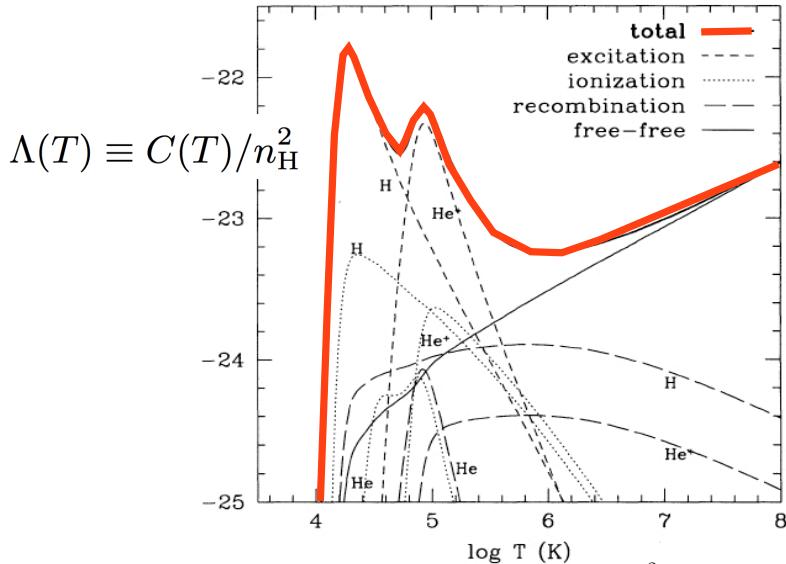


FIG. 10 The red solid line shows the total cooling function  $\Lambda(T) \equiv C(T)/n_H^2$ , where  $C(T)$  denotes the cooling rate per unit volume ,and  $n_H$  denotes the number density of H-nuclei irrespective whether the nucleus is bound to an electron or not (i.e.  $n_H = n_p + n_{HI}$ ). Short-dashed lines show the bound-bound cooling functions (there are multiple lines, each representing a different element like H, He, or HeII), dotted lines show the bound-free cooling functions, long-dashed lines show the free-bound cooling functions, while the thick solid line shows the free-free cooling function. These cooling curves assume collisional ionization equilibrium. Under this assumption cooling is generally dominated by bound-bound and free-free cooling.

All of these processes involve interactions of two particles. The rate of these two-body processes scales as the product of the number density of particle species. For example, free-free emission scales as  $\propto n_e^2$ , where  $n_e$  denotes the number density of electrons, while bound-free emission scales as  $\propto n_e n_{HI}$ , in which  $n_{HI}$  denotes the number density of hydrogen atoms.

We can now look at the cooling properties of a gas of primordial composition (i.e.  $\sim 75\%$  H by mass,  $\sim 25\%$  He). The cooling properties of a gas are usually characterized by its ‘cooling function’,  $\Lambda(T)$ , which denotes the cooling rate of gas per hydrogen nucleus per unit density (in  $\text{erg s}^{-1} \text{cm}^3$ ). This way of quantifying the cooling function is not very intuitive but removes the density dependence of the cooling processes. The cooling rate,  $C(T)$ , which denotes the rate at which the gas radiates thermal energy away per unit volume is  $C(T) \equiv \Lambda(T)n_H^2$ . Here,  $n_H$  denotes the number density of hydrogen nuclei, irrespective of whether the hydrogen is neutral or ionized. We can decompose both  $C(T)$  and  $\Lambda(T)$  into separate components, and write the total cooling rate as the sum of bound-bound, bound-free, free-bound, and free-free cooling rates, i.e.  $C(T) = C_{bb}(T) + C_{bf}(T) + C_{fb}(T) + C_{ff}(T)$ . We can do something similar for the cooling function. We discuss the cooling function in more detail below.

Figure 10 shows the total cooling function  $\Lambda(T)$  as a function of  $T$  for a primordial gas. The cooling rate at a fixed  $T$  must also depend on the ionization state of the gas (since electrons play a key role in all cooling processes), and this plot assumes that the gas temperature  $T$  uniquely sets the ionization state of the gas via collisional processes. That is, the gas is assumed to be in *collisional ionization equilibrium* (CIE). The figure also shows cooling functions associated with bound-bound emission etc.

We can understand several features in this plot as follows:

- At  $T \lesssim 10^4 \text{ K}$ ,  $\Lambda(T)$  is very small. At these temperatures, bound-bound cooling dominates (i.e.  $\Lambda(T) \sim \Lambda_{bb}(T)$ ) because the gas is mostly neutral, and the small number of electrons that are around generally do not have enough energy to excite any transitions in either H or He. Collisional ionization of either H or He is even less likely as this requires higher energy collisions. Free-free cooling is also negligible. The bound-bound cooling rate

$C_{\text{bb}}(T) \propto n_e n_{\text{HI}} \Lambda_{\text{bb}}(T) = n_{\text{H}}^2 x_{\text{e}} x_{\text{HI}} \Lambda_{\text{bb}}(T)$ . The electrons come mostly from hydrogen and we have  $n_{\text{p}} \sim n_{\text{e}}$ , i.e.  $x_{\text{e}} \sim x_{\text{HII}}$ . Since  $x_{\text{HI}} = 1 - x_{\text{HII}}$ , we can write  $C_{\text{bb}}(T) = n_{\text{H}}^2 x_{\text{HI}} (1 - x_{\text{HI}}) \Lambda_{\text{bb}}(T)$ .

- At  $T \sim 10^4$  K,  $\Lambda(T)$  suddenly rises rapidly, due to the rise in both the number density of electrons, and the fraction which is energetic enough (**Why would this go hand-in-hand?**) to collisional excite levels inside the hydrogen atoms. The efficiency of bound-bound cooling therefore increases rapidly with  $T$ . Note that the *black dotted line*, which shows the bound-free cooling rate (cooling as a result of collisional ionization) also rises rapidly with  $T$ .
- $\Lambda(T)$  peaks at  $T \sim 2 \times 10^4$  K. The rapid rise of cooling by bound-bound emission cannot increase indefinitely. At  $T \sim 2 \times 10^4$  K, there are enough electrons around which can collisionally ionize hydrogen, at which point most of the hydrogen atoms will be ionized away. With hydrogen atoms disappearing, cooling by both bound-bound and bound-free emission is decreasing rapidly.
- $\Lambda(T)$  reaches a minimum at  $T \sim 3 - 4 \times 10^4$  K. The decrease in cooling by bound-bound and bound-free emission is counteracted by the increase in cooling by bound-bound emission associated with singly ionized Helium ions. Collisionally exciting and ionizing  $\text{He}^+$  requires higher energy electrons, and therefore higher  $T$ . Collisional excitation of  $\text{He}^+$ , and the resulting cooling, rapidly kicks in at  $T \gtrsim 3 - 4 \times 10^4$  K, analogous to our explanation of the rapid rise of  $C(T)$  at  $T \sim 10^4$  K.
- $\Lambda(T)$  again peaks at  $T \sim 10^5$  K, after which  $\text{He}^+$  is mostly doubly ionized to  $\text{He}^{++}$  and cooling by bound-bound and bound-free emission associated with  $\text{He}^+$  becomes less efficient.
- At higher temperatures the gas is basically fully ionized, and free-free emission cools the gas, which causes cooling function to be dominated by the free-free cooling function, which scales as  $\Lambda(T) \sim \Lambda_{\text{ff}}(T) \propto T^{1/2}$ . Note that for free-free cooling, the cooling rate scales<sup>9</sup> as  $C_{\text{ff}}(T) = \Lambda_{\text{ff}}(T) n_e n_p = \Lambda_{\text{ff}}(T) n_{\text{H}}^2 x_{\text{HII}}^2$ .

Now that we have discussed the cooling function, we can introduce another key quantity: the *cooling time* of an astrophysical gas,  $t_{\text{cool}}$ , which is defined as

$$t_{\text{cool}} = \frac{3n_{\text{H}}k_{\text{B}}T}{3n_{\text{H}}^2\Lambda(T, \mathcal{Z})}, \quad (86)$$

where  $3n_{\text{H}}k_{\text{B}}T/2$  denotes the thermal energy of the gas (per unit volume). The cooling time can be thought of as the characteristic time over which gas loses its thermal energy, and hence its hydrostatic support against gravity. Note that we also explicitly mentioned that  $\Lambda(T)$  also depends on the presence of elements other than H and He, which are collectively referred to as ‘metals’ in astrophysics. The ‘metallicity’ of an astrophysical gas is denoted with  $\mathcal{Z}$ . The metallicity is often expressed in terms of solar units, i.e.  $\mathcal{Z} = \mathcal{Z}_{\odot}$  corresponds to the ‘solar’ metallicity. The interstellar medium of most galaxies in the present-day Universe is enriched to a level  $\mathcal{Z} \sim 0.1 - 1\mathcal{Z}_{\odot}$ . Plugging in some numbers we can get some feel for the order of magnitude of the cooling time:

$$t_{\text{cool}} = \frac{\frac{3}{2}n_{\text{H}}k_{\text{B}}T}{n_{\text{H}}^2\Lambda(T, \mathcal{Z})} \approx 3.3 \times 10^9 \text{ yr} \left[ \frac{T}{10^6 \text{ K}} \right] \left[ \frac{n_{\text{H}}}{10^{-3} \text{ cm}^{-3}} \right]^{-1} \left[ \frac{\Lambda(T, \mathcal{Z})}{10^{-23} \text{ erg s}^{-1} \text{ cm}^3} \right]^{-1}. \quad (87)$$

For now this number may not say much, but it soon will as we start talking about galaxy formation.

## G. Gas Cooling & Galaxy Formation

In the previous section, we introduce the cooling time,  $t_{\text{cool}}$ , which represents the characteristic timescale on which gas cools and loses its hydrostatic support against gravity. We introduced another time earlier, namely the collapse

<sup>9</sup> We previously decomposed the cooling rate into its separate components as  $C(T) = C_{\text{bb}}(T) + C_{\text{bf}}(T) + C_{\text{fb}}(T) + C_{\text{ff}}(T)$ . We showed that  $C_{\text{ff}}(T) = \Lambda_{\text{ff}}(T) n_{\text{H}}^2 x_{\text{HII}}^2$ , and that  $C_{\text{bb}}(T) = n_{\text{H}}^2 x_{\text{HI}} (1 - x_{\text{HI}}) \Lambda_{\text{bb}}(T)$ . Similarly, we have that  $C_{\text{bf}}(T) = n_{\text{H}}^2 x_{\text{HI}} (1 - x_{\text{HI}}) \Lambda_{\text{bf}}(T)$ , and  $C_{\text{fb}}(T) = n_{\text{H}}^2 x_{\text{HII}}^2 \Lambda_{\text{fb}}(T)$ . So, we have that  $C(T) = n_{\text{H}}^2 [\Lambda_{\text{bb}}(T)x_{\text{HI}}(1-x_{\text{HI}}) + \Lambda_{\text{bf}}(T)x_{\text{HI}}(1-x_{\text{HI}}) + \Lambda_{\text{fb}}(T)(1-x_{\text{HI}})^2 + \Lambda_{\text{ff}}(T)(1-x_{\text{HI}})^2] \equiv n_{\text{H}}^2 \Lambda(T)$ . Note that I freely used that  $x_{\text{HII}} = 1 - x_{\text{HI}}$ .

time,  $t_{\text{coll}} \equiv 1/\sqrt{4\pi G\rho}$  (see below Eq 41). If we substitute some numbers we get

$$t_{\text{coll}} \equiv t_{\text{dyn}} \approx 0.9 \times 10^9 \text{ yr} \left[ \frac{n_{\text{H}}}{10^{-3} \text{ cm}^{-3}} \right]^{-1/2}. \quad (88)$$

This timescale can be thought of as the characteristic time-scale over which ‘structural changes’ occur within the gas: it entered as the characteristic time-scale in the exponential growth rate of density perturbation initially above the Jeans mass (see discussion below Eq 41). Note that the subscript ‘dyn’ in  $t_{\text{dyn}}$  reflects that the collapse time is also often referred to as the ‘*dynamical time*’ of the gas. Let us discuss three possible scenarios:

- **Case I**  $t_{\text{cool}} < t_{\text{dyn}}$ : Gas loses its hydrostatic support before any structural changes can occur in the gas. That is, the gas cools faster than system can ‘respond and try to re-establish hydrostatic equilibrium. With no hydrostatic support, the gas can collapse freely towards center of dark matter halo.
- **Case II**  $t_{\text{cool}} > t_{\text{dyn}}$ : Gas cools, but as it slowly loses pressure, gas reorganizes to establish hydrostatic equilibrium (**How would this happen?**). Gas is in ‘quasi hydrostatic equilibrium. This means that the gas still collapses, but on the longer cooling time.
- **Case III**:  $t_{\text{cool}} > t_{\text{hub}}$ : Gas cooling takes longer than the Hubble time,  $t_{\text{hub}} \equiv \frac{1}{H(z)}$ , which can be thought of as the characteristic timescale over which changes in the Universe as a whole happen. The Hubble time close to the age of the Universe at all times. When  $t_{\text{cool}} > t_{\text{hub}}$  the gas loses its thermal support over a time-scale such that it has no physical implications.

The previous highlights that the condition  $t_{\text{cool}} < t_{\text{dyn}}$  is a key requirement to allow gas to collapse to the center of a dark matter halo, which is what is required for gas to reach high enough densities to start forming stars, etc. **Assignment: show/discuss that when this condition is satisfied, that then the condition  $t_{\text{cool}} < t_{\text{hub}}$  is also satisfied** This condition depends on gas density, and temperature (and to some extent on gas metallicity). We can now understand a basic aspect of galaxy formation in a number of steps:

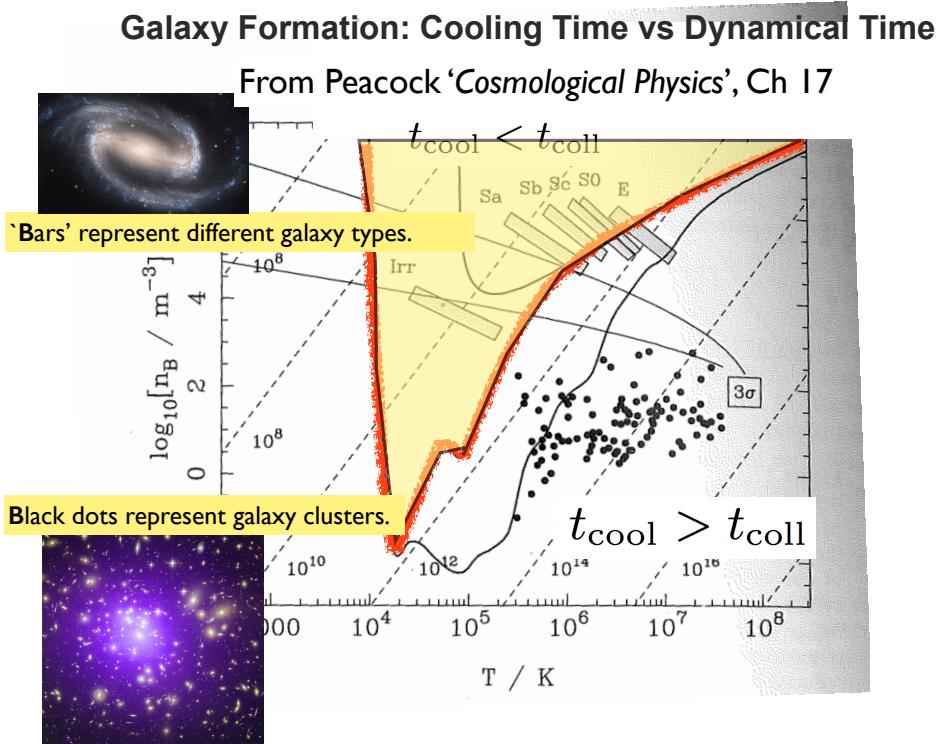


FIG. 11

- Consider gas at some temperature  $T$  (and fixed metallicity  $\mathcal{Z}$ ). Because  $t_{\text{cool}} \propto \rho^{-1}$  and  $t_{\text{dyn}} \propto \rho^{-1/2}$ , we must have  $t_{\text{cool}} < t_{\text{dyn}}$  for  $\rho > \rho_{\text{cool}}$ , where  $\rho_{\text{cool}}$  defines the density at which  $t_{\text{dyn}} = t_{\text{cool}}$ .

- We can calculate  $\rho_{\text{cool}}$  as a function of  $T$ . Points in the  $(\rho, T)$ -plane that lie above the line defined by  $\rho_{\text{cool}}$  represent an astrophysical gas which can cool and collapse. Put differently, these points represent astrophysical gases which are suitable for collapsing in a way that allows for star & galaxy formation.
- The *red solid curve* in Figure 11 shows the line defined by  $\rho_{\text{cool}}$ . The yellow-colored region indicates region in which star and galaxy formation is allowed to occur. Also indicated within this region are densities and temperatures inferred for different type of galaxies ('Irr' stands for 'irregulars', 'E' stands for 'elliptical galaxies', and 'S' stands for spiral galaxies). On the other hand, *filled solid circles* represent the physical conditions associated with clusters and groups of galaxies. This figure shows that the condition  $t_{\text{cool}} < t_{\text{dyn}}$  immediately separates galaxies from larger, more massive structures such as groups of galaxies. It explains - at a basic level - why galaxies have the masses they have, and why for example no spiral galaxies exist that are (say) 100 times more massive than our Milky Way. Instead, these massive structures typically exist of gigantic, glowing plasma spheres with some galaxies sprinkled throughout<sup>10</sup>.

The main goal of this discussion was to link the properties of galaxies to some basic statistical and atomic physics: the mass of gas is set largely by the ability of astrophysical gases to cool, which depends on microscopic interactions between ions, electrons and atoms. It is even possible to get an order of magnitude estimate for the mass of galaxies expressed entirely in fundamental constants.

---

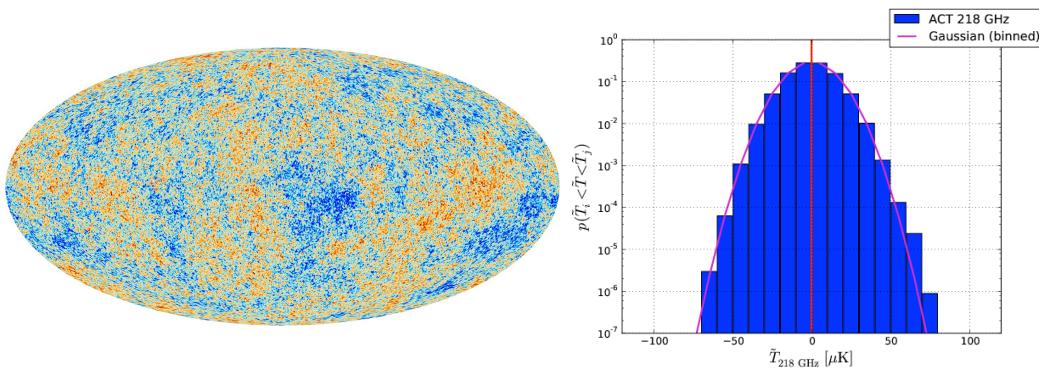
<sup>10</sup> The sheer existence of galaxies within this hot gas cannot be understood in our simple picture. We will return to this in later lectures

The previous discussions focussed on how initially small density perturbations in a ‘cosmological fluid’ (a fluid containing a collisionless dark matter component, and collisional baryons, which only interact with each other through gravity) evolve with time under the influence of gravity and pressure. In the next section we look at the statistical properties of the initial density field, and how we can relate these to observed statistical properties of galaxies. This connection further illustrates that we seem to understand the basic aspects of galaxy formation.

## VII. GAUSSIAN RANDOM FIELDS

### A. Introduction

A Gaussian random field (GRF) is a [random field](#) involving [Gaussian probability density functions](#) of the variables.



Example, the Cosmic Microwave Background

Temperature fluctuations are ‘sourced’ by density fluctuations of baryons (recall that these represent a ‘suppressed’ version of the dark matter fluctuations.)

FIG. 12

Observations of the temperature fluctuations in the cosmic microwave background (CMB) indicate that these fluctuations are Gaussian (see Fig 12): the CMB intensity and - by relating intensity to temperature in the Rayleigh Jeans limit -CMB temperature is slightly different in different directions. If we were to make a histogram of the CMB temperature in different directions, we would find that the temperature distribution would be described perfectly by a Gaussian. The *width* of the Gaussian distribution depends on how we measure temperature in a certain direction: for example, we can divide the sky up into  $N$  patches of equal solid angle  $d\Omega = 4\pi/N$ , and define the temperature to be

1. the value of the temperature at the (geometric) center of each patch.
2. the value of temperature *averaged* over the entire patch. In this case, we expect the width of the Gaussian temperature distribution to decrease if we increase the patch size (i.e. if we decrease  $N$ ), because if we take larger patches, we average over a larger number of temperature measurements. That is,  $\sigma(T)$  decreases with the ‘smoothing scale’ of the temperature field. We will return to this concept later.

In the standard cosmological model the temperature-fluctuations are sourced by density fluctuations, which must then also be Gaussian. For this reason, gaussian random fields play a key role in theory of structure formation.

In this chapter, we summarise some properties of Gaussian random fields. First, we consider a single-variate Gaussian distribution. This has many of the features which return in the more general (and cosmologically relevant)

multivariate Gaussian distribution. We then summarise properties of a Multi-variate Gaussian. This is followed by discussing properties of Gaussian random fields in Fourier space. Here, we introduce the power spectrum, which plays a central role in structure formation theories. This discussion follows that in the lecture of Eiichiro Komatsu: [http://www.mpa-garching.mpg.de/~komatsu/cmb/lecture\\_NG\\_iucaa\\_2011.pdf](http://www.mpa-garching.mpg.de/~komatsu/cmb/lecture_NG_iucaa_2011.pdf).

## B. Single-Variate Gaussian Random Fields

We have a single variable  $x$  drawn from the probability distribution function (PDF)  $p(x)$ , which is normalised to  $\int_{-\infty}^{\infty} dx p(x) = 1$ . If  $x$  obeys Gaussian statistics then

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (89)$$

, where  $\sigma^2$  denotes the variance. The PDF contains the full information of the field. Various useful statistical quantities of the field are 'moments' of the PDF. The first four are:

$$\begin{aligned} \langle x \rangle &= \int_{-\infty}^{\infty} dx x p(x) = 0 \\ \langle x^2 \rangle &= \int_{-\infty}^{\infty} dx x^2 p(x) = \sigma^2 \\ \langle x^3 \rangle &= \int_{-\infty}^{\infty} dx x^3 p(x) = 0 \\ \langle x^4 \rangle &= \int_{-\infty}^{\infty} dx x^4 p(x) = 3\sigma^4. \end{aligned} \quad (90)$$

These 4 moments indicate that the variable  $x$  has zero mean (first moment), variance  $\sigma^2$  (second moment), zero skewness (third moment), zero kurtosis (the kurtosis  $\kappa_4 \equiv \langle x^4 \rangle - 3\langle x^2 \rangle^2 = 0$ ). For a single-variate Gaussian random field, all odd moments vanish, all even moments are given in terms of  $\sigma^{2n}$ .

## C. Multi-Variate Gaussian Random Fields

The general expression for a multi-variate (here there are  $N$  variables) Gaussian-PDF is given by

$$p(x_1, x_2, \dots, x_N) = \frac{1}{(2\pi)^{N/2} |\xi|} \exp\left(-0.5 \sum_{ij} x_i (\xi_{ij}^{-1}) x_j\right), \quad (91)$$

where  $\xi_{ij}$  denotes the covariance matrix or the two-point correlation function. Moments of this PDF are

$$\begin{aligned} \langle x_i \rangle &= \underbrace{\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \dots \int_{-\infty}^{\infty} dx_N}_{\int d^N x} x_i p(x_1, x_2, \dots, x_N) = 0 \\ \langle x_i x_j \rangle &= \int d^N x x_i x_j p(x_1, x_2, \dots, x_N) = \xi_{ij} \\ \langle x_i x_j x_l \rangle &= \int d^N x x_i x_j x_l p(x_1, x_2, \dots, x_N) = 0 \\ \langle x_i x_j x_l x_m \rangle &= \int d^N x x_i x_j x_l x_m p(x_1, x_2, \dots, x_N) = \xi_{ij} \xi_{lm} + \xi_{il} \xi_{jm} + \xi_{im} \xi_{jl}. \end{aligned} \quad (92)$$

This is very similar to the single variate case: odd moments vanish. Even moments are given in terms of  $\xi$ . Note that  $\langle x_i x_i \rangle = \xi_{ii} = \sigma^2$ . Moreover,  $\langle x_i x_i x_i x_i \rangle = 3\sigma^4$  exactly as in the single-variate case.

Next, we will use that the Universe is isotropic (invariant under rotation) and homogeneous (invariant under translation). Let  $x$  now be a variable that depends on position  $\mathbf{q}$ . The two-point correlation function  $\xi_{ij}$  is

$$\xi_{ij} = \langle x(\mathbf{q}_i)x(\mathbf{q}_j) \rangle = \langle x(\mathbf{q}_i)x(\mathbf{q}_i + \mathbf{r}_{ij}) \rangle, \quad (93)$$

where  $\mathbf{r}_{ij}$  is a vector that connects the two points. Homogeneity (translational invariance) requires that  $\xi_{ij}$  is a function of  $\mathbf{r}_{ij}$  alone, and not of  $\mathbf{q}_i$ . Homogeneity and isotropy then requires that  $\xi_{ij}$  is a function of  $|\mathbf{r}_{ij}|$  alone.

#### D. Multi-Variate Gaussian Random Fields in Fourier Space

Working in Fourier space may be less intuitive, but it has mathematical advantages advantages. We will encounter one at the end of this section. The Fourier transform of  $x(\mathbf{q})$  is given by

$$\tilde{x}(\mathbf{k}) = \int d^3\mathbf{q} \exp(-i\mathbf{k} \cdot \mathbf{q}) x(\mathbf{q}). \quad (94)$$

The PDF of the Fourier components is given by

$$p(\hat{x}(\mathbf{k}_1), \hat{x}(\mathbf{k}_2), \dots, \hat{x}(\mathbf{k}_N)) = \frac{1}{(2\pi)^{N/2} |C|^{1/2}} \exp \left( -0.5 \sum_{ij} \tilde{x}(\mathbf{k}_i)(C_{ij}^{-1}) \tilde{x}^*(\mathbf{k}_j) \right), \quad (95)$$

where  $\tilde{x}^*$  denotes the complex conjugate of  $\tilde{x}$ . Now, similar to the previous cases

$$\langle \tilde{x}(\mathbf{k}_i)\tilde{x}^*(\mathbf{k}_j) \rangle = C_{ij}. \quad (96)$$

Assuming spatial homogeneity, we can show that

$$\begin{aligned} C_{ij} &= (2\pi)^3 \delta_D(\mathbf{k}_i - \mathbf{k}_j) P(\mathbf{k}_j) \\ P(\mathbf{k}_j) &= \int d^3\mathbf{r} \exp(-i\mathbf{k} \cdot \mathbf{r}) \xi(\mathbf{r}). \end{aligned} \quad (97)$$

Here,  $P(\mathbf{k}_j)$  is known as the 'power spectrum'. Note that  $C_{ij} = 0$  if  $i \neq j$ . That is, different Fourier components of a Gaussian random field are completely uncorrelated. Phrased differently, different Fourier modes of a Gaussian random field are completely independent. Assuming additional isotropy, we have that  $P(\mathbf{k}_j) = P(|\mathbf{k}_j|)$ . Using these expressions can 'simplify'<sup>11</sup> the PDF of the Fourier components:

$$p(\hat{x}(\mathbf{k}_1), \hat{x}(\mathbf{k}_2), \dots, \hat{x}(\mathbf{k}_N)) = \frac{1}{(2\pi)^{N/2} |\prod_i P(|\mathbf{k}_i|)|^{1/2}} \exp \left( -0.5 \sum_i \frac{|\tilde{x}(\mathbf{k}_i)|^2}{P(|\mathbf{k}_i|)} \right). \quad (98)$$

The reason that we call this simplified is that there is no  $N \times N$  matrix anymore (this is a direct consequence of the fact that different Fourier models are independent. In this case, the  $N \times N$  covariance matrix only has non-zero terms on its diagonal). Instead, all information is now encoded in the  $N$ -dimensional vector  $P(\mathbf{k}_i)$ , the **power spectrum**. This power spectrum is a popular and powerful way to describe the information of a (Gaussian) density field.

#### E. $\sigma(M)$ : Time evolution & Relation to $P(k)$

The previous discussion indicates that the variance  $\sigma^2$  plays a key role in characterizing Gaussian random fields. We previously discussed how the CMB temperature (or intensity) in a certain direction of the sky depends on the

<sup>11</sup> It still looks a bit ugly.

scale over which we smooth the density field. Similarly, the variance of the mass density field must depend on smoothing scale  $S_c$ , i.e.  $\sigma \equiv \sigma(S_c)$ . It is common to not define the variance in terms of a smoothing scale, but instead in terms of a total enclosed average mass  $M$ : the comoving smoothing scale  $S_c$  encloses a fixed amount of average mass  $M = \frac{4\pi}{3} \bar{\rho} S_c^3$ . We generally write  $\sigma \equiv \sigma(M)$ .

A useful think to know is that when density fluctuations are small (i.e. in the linear regime)  $\delta \propto a$  (see Eq 55) for all relevant  $\delta$ , then we have  $\sigma(M) \propto a$  as well. That is, gravity amplifies density fluctuations, which can be quantified as a continuous increase in  $\sigma(M)$  of the Gaussian random density field.

Recall that  $\sigma^2(M)$  denotes the variance of the matter density field smoothed on mass-scale  $M$  (or comoving smoothing scale  $S_c$ ). In practise, the (over)density at some location  $\mathbf{r}_c$   $\delta_{\text{sm}}(\mathbf{r}_c)$  is given by

$$\delta_{\text{sm}}(\mathbf{r}_c, M) = \frac{1}{V} \int_{|\mathbf{r}' - \mathbf{r}_c|} d^3 \mathbf{r}' \delta(\mathbf{r}') \equiv \frac{1}{V} \int d^3 \mathbf{r}' \delta(\mathbf{r}') W_M(|\mathbf{r}' - \mathbf{r}_c|) \quad (99)$$

, with  $V = \frac{4\pi}{3} S_c^3$ . We have also introduced the window function  $W_M(x)$  which is 1 for  $x < S_c$  and 0 otherwise (the mass dependence through  $S_c$  is represented by the subscript  $M$ ). This equation shows explicitly that evaluating the smoothed density at some location  $\mathbf{r}_c$  involves a convolution of the 3-D density field with the window function, which requires 3-dimensional integral. This is computationally expensive to repeat for a large number of  $\mathbf{r}_c$ . Fortunately, the convolution theorem states that a convolution in real space corresponds to a product in Fourier space. That is, the Fourier components of the smoothed density field can be obtained as

$$\tilde{\delta}_{\text{sm}}(\mathbf{k}_c, M) = \tilde{\delta}(\mathbf{k}) \tilde{W}(\mathbf{k}_c, S_c), \quad (100)$$

where quantities with a tilde indicate Fourier components/transforms. In the **Assignment** you will show that

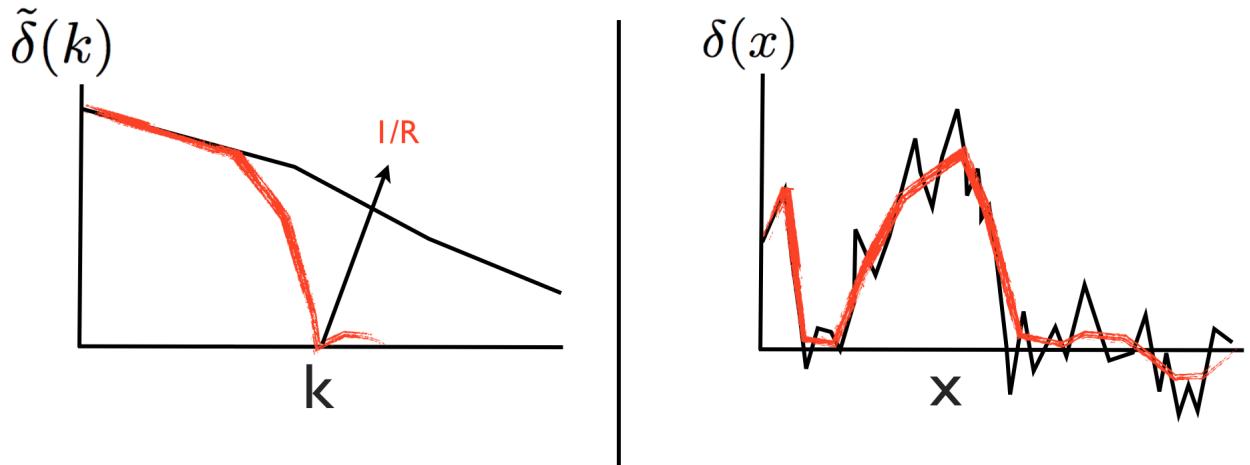


FIG. 13

$\tilde{W}(\mathbf{k}_c, S_c)$  is a function that is peaked on  $k = 0$ , and that its characteristic width is  $\propto 1/S_c$ . That is, smoothing the density field on some comoving scale  $S_c$  corresponds to suppressing the amplitude of the Fourier components at  $|\mathbf{k}_c| \gtrsim 1/S_c$ . Another way of saying this is that *smoothing the density field on some comoving scale  $S_c$  corresponds to suppressing the power-spectrum at  $|\mathbf{k}_c| \gtrsim 1/S_c$* . This statement will be useful throughout these lectures. Figure 13 schematically shows this: the *black lines* show a 1D density field,  $\delta(x)$  (*left*), and its Fourier transform  $\tilde{\delta}(k)$  (*right*). The *red lines* show the same density field, but smoothed over a larger scale. The density field looks smoother in real space, and its Fourier-transform is suppressed for larger  $k$ . Another way to see this is that by adding Fourier modes with increasing  $k$ , we are adding increasingly fine structure to the density field.

As a first demonstration of how useful this is, we discuss how quantity  $\sigma(M)$  relates to the power spectrum  $P(k)$ .

We know that  $\xi(\mathbf{r})$  and  $P(\mathbf{k})$  are Fourier transforms of each other (see Eq 97):

$$\xi(r) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} P(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (101)$$

We also know from our analysis on Gaussian random fields that  $\sigma^2 = \xi(\mathbf{r} = 0)$ . Therefore,

$$\sigma^2 = \int \frac{d^3\mathbf{k}}{(2\pi)^3} P(\mathbf{k}) \underset{\text{isotropy}}{=} \frac{1}{2\pi^2} \int_0^\infty dk k^2 P(k). \quad (102)$$

Smoothing on a mass scale  $M$  - i.e. comoving length scale  $S_c$  - suppresses the power at  $k \gtrsim 1/S_c$ . Let us assume for simplicity that smoothing completely suppresses the power for  $k \gtrsim 1/S_c$ :

$$P_{\text{sm}}(k) = \begin{cases} P(k) & k \leq k_{\text{sm}} \\ 0 & k > k_{\text{sm}}, \end{cases} \quad (103)$$

where  $k_{\text{sm}} \equiv S_c^{-1} \propto M^{-1/3}$ . Substituting this we get

$$\sigma^2(M) = \frac{1}{2\pi^2} \int_0^\infty dk k^2 P_{\text{sm}}(k) \propto \frac{1}{2\pi^2} \int_0^{k_{\text{sm}}} dk k^2 P(k) \propto \int_0^{k_{\text{sm}}} dk k^{2+n}, \quad (104)$$

where we assumed that  $P(k) = Ak^n$ . We therefore have

$$\sigma^2(M) \propto k_{\text{sm}}^{3+n} \propto S_c^{(-n-3)} \propto M^{(-n-3)/3}. \quad (105)$$

The main point is that we can compute  $\sigma(M)$  once we specified  $P(k)$ . This is extremely useful as we will see next. As we will discuss in later lectures, the *primordial* power spectrum has a slope  $n \sim 1$  on very large scales (small  $k$ ), but turns-over and approaches  $n \sim -3$  on smaller scales (large  $k$ ). This shows that  $\sigma(M)$  increases as we decrease  $M$ , but that it flattens out towards very small  $M$ .

## VIII. STATISTICAL DESCRIPTIONS OF OBSERVATIONS OF GALAXIES

Before going deeper into Gaussian random field theory, we discussed some basic statistical descriptions of galaxies. In particular, we focussed on the *galaxy luminosity function*, and *galaxy clustering*. These are discussed in the following two subsections. For some examples of measurements of both quantities, I refer you to the lecture slides. There are other useful measures such as the *stellar mass function*, but we have not discussed this here.

### A. The Galaxy Luminosity Function (LF)

The Luminosity Function (LF) quantifies the number density of galaxies as a function of their luminosity (total radiative output). Luminosity functions are always measured in a specific band. A popular luminosity function is the rest-frame UV luminosity function (UV-LF) which quantifies the number density of galaxies as a function of the inferred rest-frame UV-luminosity. The UV-luminosity is usually defined to correspond to wavelengths in the range  $\lambda = 1300 - 1700$  Å. As we will see in later lectures, the rest-frame UV is an interesting band to work in, as radiation that comes out of galaxies at these wavelengths originates in young, hot stars. As a result, it has been shown that UV-luminosity provides a good measure of the rate at which galaxies are forming stars. The UV-LF has been measured very well out to  $z \sim 6$ , and constraints exist at redshifts as high as  $z \sim 10$ . The UV-LF, but also LFs in other bands can be described very well by a so-called *Schechter function*:

$$\phi(L) \equiv \frac{dn}{dL} = \phi_* \left( \frac{L}{L_*} \right)^\alpha \exp \left( -\frac{L}{L_*} \right), \quad (106)$$

where

- $\phi_*$  is the overall normalization.

- $L - *$  defines a characteristic luminosity where the LF breaks from a power-law into an exponential function
- $\alpha$  defines the faint-end slope.

Observations have constrained all three parameters well out to  $z \sim 6$ . The deepest images with Hubble's 'Wide-Field Camera 3' have recently provided us with constraints at higher  $z$ .

## B. Clustering

Galaxies are not distributed randomly throughout the Universe. A simple analogy is the people on the Earth surface. The surface of the earth is about  $4 \times 10^8 \text{ km}^2$  with  $\sim 6 \times 10^9$  people on it. This gives an average of  $\Sigma_{\text{people}} \sim 15 \text{ km}^{-2}$ . If you take a random person on Earth, and ask what the probability is that there will be a second person within  $R = 10$  meters, then if people were truly randomly distributed, this probability would be  $P = \pi \Sigma_{\text{people}} R^2 \approx 4 \times 10^{-3}$ . We know however, that in practise this probability must be significantly greater. This excess probability over the random probability indicates that the spatial distribution of people is *clustered*. The same is the case for galaxies. If  $n_g(\mathbf{r})$  denotes the number density of galaxies at  $\mathbf{r}$ , then  $\langle n_1 \rangle \neq 0$ . To connect more closely to the formalism discussed previously (see Eq ), we instead work with galaxy over-number density, i.e.  $\delta n \equiv \frac{n - \bar{n}}{\bar{n}}$ , in which case  $\langle \delta n \rangle = 0$ . Note that

$$\xi_{12} \equiv \xi(|\mathbf{r}_{12}|) = \frac{\langle n_g(\mathbf{r}_1) n_g(\mathbf{r}_2) \rangle}{\bar{n}_g(\mathbf{r}_1) \bar{n}_g(\mathbf{r}_2)} - 1 \quad (107)$$

We will return to this two-point correlation function in more detail later. In the lecture, we described a few observed aspects of galaxy clustering. These include:

- The two-point correlation function of the spatial distribution of galaxies is given by a power-law:  $\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$ , where  $r_0$  denotes the correlation length.
- Observations indicate that  $\gamma \sim 1.7 - 1.8$ .
- Brighter galaxies are more strongly clustered, i.e. brighter galaxies have a longer correlation length.

## IX. THE PRESS SCHECHTER FORMALISM

The goal of this section is to show that the previously described observations can broadly be well understood from Gaussian random field theory. Specifically, the Press-Schechter (PS) theory can be thought of as a first attempt to connect statistical aspects of a Gaussian random density field, to observed statistical descriptions of galaxies. As we will discuss later, the PS formalism is flawed in many ways, yet somehow works remarkably well.

### A. The Basic Press-Schechter Formalism

From our previous discussions, we know that galaxies form inside virialized (i.e. collapsed) dark matter halos. To connect the random Gaussian density field to galaxies, we will first describe some statistical properties of regions in the density field that have collapsed. We will focus on the probability that mass at  $\mathbf{x}$  is part of a virialized object with mass  $> M$ . This probability is denoted with  $P(> M)$ . The Press-Schechter formalism is based on the following assumptions

1. The density field evolves linearly at all times, i.e.  $\delta \propto a$ .
2. An object has collapsed (or virialized) once its linearly extrapolated overdensity exceeds  $\delta_{\text{crit}} = \delta_{\text{lin}} = 1.69$  (see Eq 67 for the origin of this number).
3. The probability that mass at point  $\mathbf{x}$  is part of a collapsed object with mass  $> M$  equals:

$$P(> M) = P(\delta > \delta_{\text{crit}} | M) \quad (108)$$

where  $P(\delta | M)$  denotes the PDF of the density field smoothed on some scale  $S_c$  - which corresponds to a mass scale  $M$ . The idea is that if  $\delta > \delta_{\text{crit}}$  on some scale  $M$ , then  $\delta = \delta_{\text{crit}}$  on some larger mass-scale  $M' > M$  and

that  $\mathbf{x}$  would be part of the larger collapsed mass  $M'$ .

Because the mass-density field smoothed on mass-scale  $M$ , indicated with  $\delta_{\text{sm}}(\mathbf{x})$ , is also a Gaussian random field with zero mean and variance  $\sigma_{\text{sm}}^2(M)$ . We can compute the probability

$$P(\delta_{\text{sm}} > \delta_{\text{crit}} | M) = \frac{1}{\sigma(M)\sqrt{2\pi}} \int_{\delta_{\text{crit}}}^{\infty} d\delta' \exp\left(-\frac{\delta'^2}{2\sigma^2(M)}\right). \quad (109)$$

This integral can be re-expressed as an error-function<sup>12</sup>:

$$P(> M) = P(\delta > \delta_{\text{crit}} | M) = \frac{1}{2} \left[ 1 - \text{erf}\left(\frac{\nu}{\sqrt{2}}\right) \right], \quad (110)$$

where  $\nu \equiv \frac{\delta_{\text{crit}}}{\sigma(M)}$ . This idea has problem with it, if we take  $M \rightarrow 0$ , then  $P(> M) = 0.5$ . This is because for a Gaussian random field, half the density fluctuations are under dense. This was a well known problem as soon as the theory was proposed, and was fixed simply by multiplying by a factor<sup>13</sup> of 2. We will do that as well, i.e.

$$\boxed{P(> M) = P(\delta > \delta_{\text{crit}} | M) = \left[ 1 - \text{erf}\left(\frac{\nu}{\sqrt{2}}\right) \right].} \quad (111)$$

Because  $P(> M)$  denotes the probability that mass at  $\mathbf{x}$  is part of collapsed structure with mass  $> M$ , it also corresponds to the fraction of mass in the Universe that is in collapsed structures of this mass (this assumes ‘ergodicity’, i.e. when sampled over a large enough volume of the Universe, the density fluctuations in the Universe closely follow that of the underlying statistical distribution).

If we take  $P(> M)$  to denote the fraction of mass locked up in an object of mass  $> M$ , then  $\frac{\partial P}{\partial M} dM$  denotes the mass fraction of the Universe locked up into collapsed dark matter halos of masses in the range  $M \pm dM/2$ . This fraction can be related to the dark matter halo mass function,  $n(M)dM$ , which denotes the number density of dark matter halos with masses in the range  $M \pm dM/2$ . To see this connection, we note that

1. If  $n(M)dM$  denotes the *number density* of dark matter halos with mass range  $M \pm dM/2$ , the  $M \times n(M)dM$  denotes the *mass density* in dark matter halos with masses in the range  $M \pm dM/2$ .
2.  $Mn(M)dM/\rho_m$  denotes the fraction of the mass density (i.e. mass fraction) in dark matter halos with masses in the range  $M \pm dM/2$ .

We therefore have

$$\begin{aligned} Mn(M)dM/\rho_m &= \frac{\partial P}{\partial M} dM \Rightarrow n(M) = \frac{\rho_m}{M} \frac{\partial P}{\partial M} = \frac{\rho_m}{M} \frac{\partial}{\partial M} \left[ 1 - \text{erf}\left(\frac{\nu}{\sqrt{2}}\right) \right] = \\ &= \frac{\rho_m}{M} \frac{\sqrt{2}}{\sqrt{2}} \frac{\partial}{\partial \nu} \left[ 1 - \text{erf}\left(\frac{\nu}{\sqrt{2}}\right) \right] \frac{\partial \nu}{\partial M} = \frac{\rho_m}{M} \frac{1}{\sqrt{2}} \frac{\partial}{\partial x} \left[ 1 - \text{erf}(x) \right] \frac{\partial \nu}{\partial M} = \\ &= -\frac{\rho_m}{M} \frac{1}{\sqrt{2}} \frac{2}{\sqrt{\pi}} \exp(-x^2) \frac{\delta_{\text{crit}}}{\sigma^2(M)} \frac{\partial \sigma}{\partial M} \end{aligned} \quad (112)$$

What remains to be done is to characterise our Gaussian random field, which we do with the power spectrum  $P(k)$  which enters our expression through its influence on  $\sigma(M)$ . In the lecture we recast this expression as

$$n(M) = \frac{\rho_m}{M^2} \sqrt{\frac{2}{\pi}} \exp(-\nu^2/2) \nu \left| \frac{\partial \log \sigma}{\partial \log M} \right| \Rightarrow \boxed{n(M) \propto M^{\frac{n-9}{6}} \exp(-\nu^2)}. \quad (113)$$

where we used our earlier derived result that  $\sigma^2(M) \propto M^{-\frac{n-3}{3}}$ . In later lectures, we will derive that  $n$  ranges from  $n \sim -3$  on small scales to  $n = 1$  on large scales. This implies that  $\sigma(M)$  decreases with mass, at a rate that depends

<sup>12</sup> The error function is define as  $\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^\infty \exp(-t^2) dt$ .

<sup>13</sup> This factor of 2 follows naturally from extension of the theory as we discuss later.

on  $M$ .

Notice the structure of the formula for  $n(M)$ :

- The expression for  $n(M)$  contains a power of  $M$  that is multiplied by an exponential function of a power of  $M$  (because  $\nu^2 \propto \sigma^{-2}(M) \propto M^{-\frac{n-3}{3}}$ ). Notice the apparent similarity of this function to the Schechter function (see Eq 106), which represents *an observed* quantity.
- Low  $M$  corresponds to small-scale fluctuations, and we expect  $\sigma(M)$  to be large. If  $\sigma(M) \gg \delta_{\text{crit}} \Rightarrow \nu \ll 1$ , then we can ignore the exponential term, and  $n(M)$  is a power-law function of  $M$ ,
- The exponential function becomes important at  $M_{\text{crit}}$  which is defined as the mass at which  $\nu = 1$ , i.e. when  $\sigma(M_{\text{crit}}) = 1.69$ .
- For  $M > M_{\text{crit}}$ ,  $n(M)$  decreases exponentially with  $M$ .
- The characteristic mass-scale  $M_{\text{crit}}$  evolves with cosmic time, because in general  $\sigma(M) \propto a$ . That is, the characteristic 'turn-over' mass-scale increases with  $a$ , i.e. it decreases with redshift.

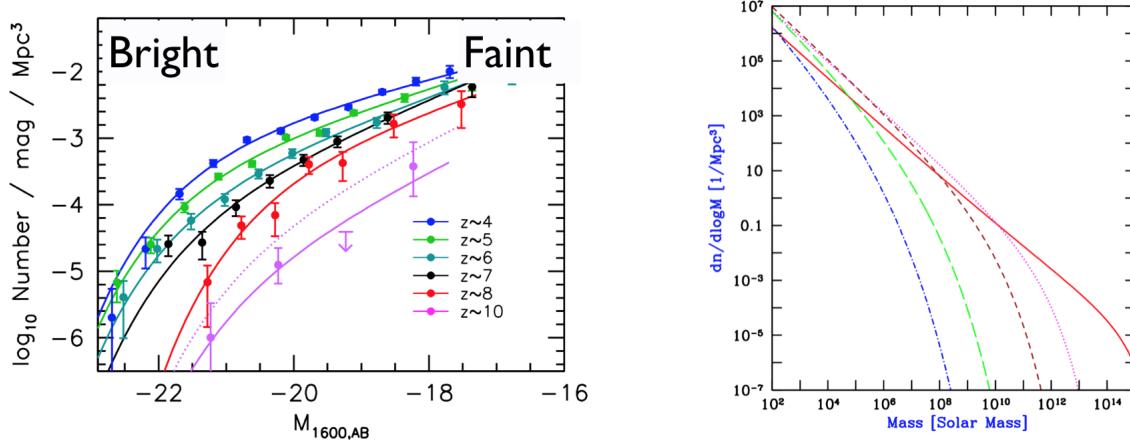


FIG. 14

This behavior of  $n(M)$  already allows us to *qualitatively* understand the observed redshift evolution of the UV-LF, which provides a good quantitative description of the number density of star forming galaxies as a function of their star formation rate (see § VIII.A): if the rate at which galaxies form stars relates to the dark matter halo mass hosting the galaxies (**Think about why this may be the case**), then we can understand (i) why the number density of galaxies decreases exponentially for the most UV-luminous systems, and (ii) why most of the evolution in the UV-LF is in the evolution of the characteristic turn-over luminosity  $L_*$  (see Eq 106, and see the *left panel of Fig 14*).

The *right panel* Figure 14 also shows that the number density of low-mass dark matter halos ( $M \lesssim 10^{5-6} M_\odot$ ) evolves only weakly with cosmic time, just like the number density of UV-faint galaxies. In contrast, the number density of more massive dark matter halos evolves rapidly with redshift, and in a way that increasingly massive systems appear while the number density of lower-mass dark matter halos remains more constant after they first appeared. This is referred to as *hierarchical* structure formation, in which low mass dark matter halos collapse first and increasingly massive halos appear as cosmic time passes by.

This hierarchical build-up of structure implies that the first galaxies formed in the lowest mass dark matter halos. *How low* depends on the previously introduced Jeans Mass,  $M_J$  (see Eq 42). Only if  $M > M_J$  is gas able to overcome pressure forces and collapse following the dark matter. In **the assignment** you will be asked to show that the cosmological Jeans mass is

$$M_J = \frac{\pi}{6} \left( \frac{\pi k_B}{\mu m_p G} \right)^{3/2} \frac{T^{3/2}}{\rho^{1/2}} \approx 10^4 - 10^5 M_\odot, \quad (114)$$

at  $z \approx 10 - 20$ . But as we discussed previously, the formation of galaxies does not only require that  $M > M_J$ , it also requires that the gas is able to cool efficiently enough (i.e that  $t_{\text{cool}} < t_{\text{coll}}$ ). The virial temperature for dark matter halos just above the cosmological Jeans mass is  $T_{\text{vir}} \ll 10^4$  K, and gas has no way of cooling (see Fig 10). It therefore appears that halos with  $M$  just above the cosmological Jeans mass can collapse, but form no stars. This is not far from our current understanding of the formation of the first galaxies. It turns out that at the cooling efficiency of primordial gas does not plummet all the way to zero for  $T \ll 10^4$  K. Instead, at high redshifts, molecular hydrogen ( $\text{H}_2$ ) can form, which can also cool gas in dark matter halos with  $T_{\text{vir}} \gtrsim 300$  K. These halos, with 300 K  $\lesssim T_{\text{vir}} \lesssim 10^4$  K are called 'mini'-halos. It is generally thought that these minihalos hosted the first stars that formed in our Universe. These halos likely did not form 'galaxies': as we will discuss in later lectures, the radiation from the first stars likely heated the gas inside minihalos to a temperature of  $T = 10^4$  K. The thermal speed of atoms at this temperature ( $\sim 10 \text{ km s}^{-1}$ ) greatly exceeds the escape velocity of minihalos, which is of order the circular velocity,  $v_{\text{circ}}$ . Radiation from the first stars would likely unbind all baryons, and shut-down further star formation. Star formation is therefore thought to have proceeded very inefficiently in low mass minihalos. The suppression by subsequent star formation after the first stars formed represents a form of 'feedback'. Understanding feedback processes plays a key role in understanding galaxy formation in general as we will discuss in later lectures.

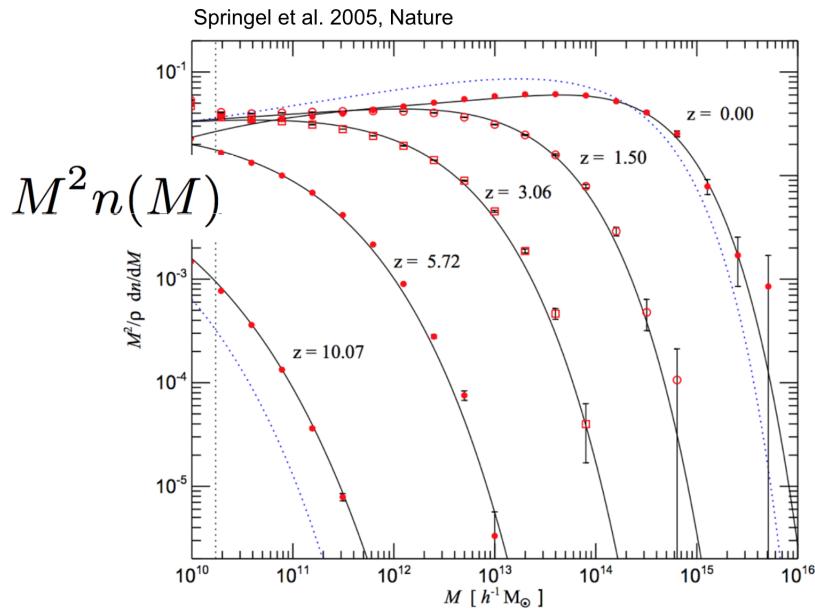


FIG. 15

It is also instructive to compare  $n(M)$  predicted by the Press-Schechter formalism to that obtained from full numerical simulations, which solve the Poisson, Euler & continuity equations with as little approximations as numerically possible. It is worth keeping in mind that identifying & counting 'collapsed' objects in a cosmological numerical simulation is not trivial. Figure 15 compares simulated number density of dark matter halos (red points) to the Press-Schechter mass function (blue dotted lines). While clearly the agreement is not perfect, given the simplicity of the Press-Schechter formalism, which is based entirely on Gaussian random fields, it is remarkable how it reproduces the observed number density of DM halos in the present-day Universe, which has evolved well into the non-linear regime on a wide range of scales. The *black solid lines* show the Sheth-Tormen modification of the Press-Schechter mass function, which is based on a modification of the original PS formalism which allows for non-spherical (ellipsoidal) collapse. The Sheth-Tormen mass function provides a practically perfect description of the halo mass function at all  $z$ .

## B. The Origin of the Fudge Factor of 2

Perhaps one of the most disturbing aspects of the PS theory, is the arbitrary introduction of the factor of 2 in the mass-fraction in bound objects with mass  $> M$ :

$$P(> M) = 2P(\delta > \delta_{\text{crit}} | M) = 1 - \text{erf}(\nu/\sqrt{2}), \quad (115)$$

entirely to guarantee that  $P(> 0) = 1$ . Without this fudge factor, PS theory would underestimate  $P(> M)$ . This is because it actually is possible for  $\delta$  to be less than  $\delta_{\text{crit}}$  on a mass-scale  $M$  -centered on  $\mathbf{x}$  - while it is part of a collapsed object with a mass  $> M$ . Visually, this is shown in Figure 16. The assumption that  $P(> M) = P(\delta > \delta_{\text{crit}} | M)$  is

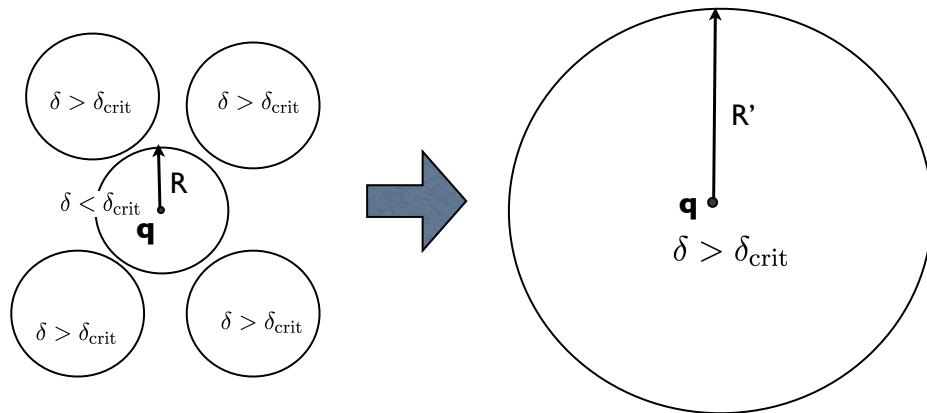


FIG. 16 The density at  $\mathbf{x}$  (in the figure mistakenly indicated with  $\mathbf{q}$ ) smoothed on a scale  $R$  is  $\delta < \delta_{\text{crit}}$ . The PS formalism would say that this point cannot be part of a collapsed object of mass  $> M$ . Note however, that neighbouring regions may all have  $\delta > \delta_{\text{crit}}$ . If we smooth the density field on some larger scale  $R' > R$ , then it is possible that the smoothed density at  $\mathbf{x}$   $\delta > \delta_{\text{crit}}$  and the point would be part of a collapsed object of mass  $M' > M$ . The original PS formalism (without the factor of 2) failed to account for these objects.

thus not correct. We can fix this, and interestingly the fix provides the infamous factor of 2. We show this next.

Again consider position  $\mathbf{x}$  (i.e.  $\mathbf{q}$  in Fig 16). We are going to evaluate the density at  $\mathbf{x}$  but smoothed over the largest possible scale, and then at increasingly smaller scales. We are going to do this in a number of steps:

- First, we smooth the density field in a sphere of radius  $R \rightarrow \infty$  centered on  $\mathbf{q}$ . Eq 104 then shows that  $\sigma(M) = 0$ . On this scale the density field is Gaussian with zero variance, and if we draw  $\delta$  from a Gaussian with zero variance, then it must be zero, i.e.  $\delta = 0$ . This point is labeled as '1' in Figure .
- Next, we reduce  $R$ . This corresponds to adding independent Fourier modes to the integral over  $k$ . We now have a finite  $\sigma^2(M_2)$ . The subscript '2' refers to the step number in this analysis. We now draw a random  $\delta$  from the Gaussian PDF with variance  $\sigma^2(M_2)$ . This  $\delta_2$  is labeled in Figure 17 with '2'.
- Next, we reduce  $R$  further. This again corresponds to adding independent Fourier modes. The new variance  $\sigma^2(M_3) \equiv \sigma^2(M_2) + d\sigma_{23}^2$ , where  $d\sigma_{23}^2$  is the variance we added by adding the new Fourier modes. We generate the new  $\delta_3 = \delta_2 + d\delta_{23}$  where  $d\delta_{23}$  is generated from a Gaussian distribution with variance  $d\sigma_{23}^2$
- Repeat....

This analyse shows that  $\delta$  makes a random change for each reduction in  $R$  (the r.m.s size of this change in step  $n$  is given by the newly added variance  $d\sigma_{n,n-1}$ ). Each point in the Universe has its own random trajectory - each random trajectory is called a '**random walk**'.

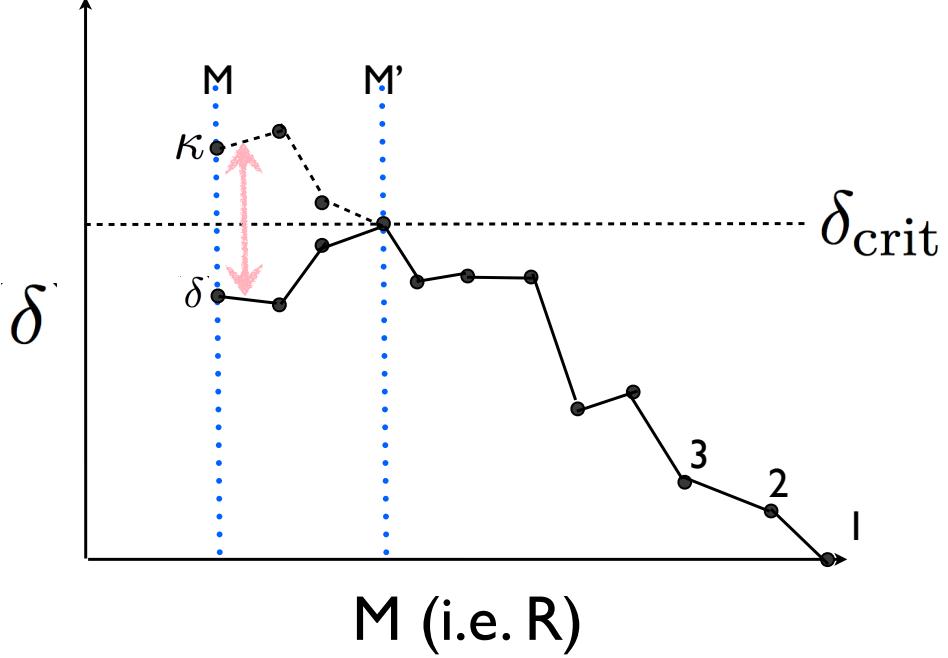


FIG. 17

Figure 17 shows the random walk associated with point  $\mathbf{x}$  shown in Figure 16. The random walk crosses the barrier at mass scale  $M'$ , and  $\mathbf{x}$  is thus part of a collapsed object of mass  $M'$ . We continue the random walk to smaller  $R$  (and thus  $M$ ). In the case of Figure 16, we know that the random walk brings  $\delta$  to  $\delta < \delta_{\text{crit}}$  at mass scale  $M$ . Note that because each step is drawn from a Gaussian distribution, which is symmetric around zero, we could have generated the random walk- mirrored around the line  $\delta_{\text{crit}}$ , and which ends at  $\kappa = 2\delta_{\text{crit}} - \delta$ , with equal probability. We therefore know that the probability of ending at  $\delta$  on mass scale  $M$ , after crossing  $\delta_{\text{crit}}$  on mass scale  $M'$  is equal to ending at  $\kappa$  on mass scale  $M$ .

We know that  $P(\delta|M)$  is a Gaussian with variance  $\sigma^2(M)$ . We would like to differentiate between points at  $\delta|M$  (this is short-hand notation for density  $\delta$  smoothed on mass scale  $M$ ) that did, and that did not cross the barrier at some larger mass-scale  $M'$ . We can obtain the probability that a point has  $\delta|M$  without having crossed any barrier on larger scales,  $P_{\text{nc}}$  ('nc' stands for 'no crossing') from

$$P_{\text{nc}}(\delta|M) = P(\delta|M) - P_{\text{cross}}(\delta|M), \quad (116)$$

where  $P_{\text{cross}}(\delta|M)$  denotes the probability that a point reached  $\delta|M$  after having crossed the barrier. We know from our random-walk analysis (Fig 17) that  $P_{\text{cross}}(\delta|M) = P(\kappa|M)$ . We therefore have

$$P_{\text{nc}}(\delta|M) = P(\delta|M) - P([2\delta_{\text{crit}} - \delta]|M) = \frac{1}{\sqrt{2\pi}\sigma(M)} \left( \exp \left[ -\frac{\delta^2}{2\sigma^2(M)} \right] - \exp \left[ -\frac{[2\delta_{\text{crit}} - \delta]^2}{2\sigma^2(M)} \right] \right). \quad (117)$$

The probability that a mass element at  $\mathbf{q}$  is therefore embedded within a collapsed object of mass  $> M$  is then

$$P(> M) = 1 - \int_{-\infty}^{\delta_{\text{crit}}} d\delta P_{\text{nc}}(\delta|M) \quad (118)$$

I left it as an exercise to show that now we are left with

$$P(> M) = 2P(\delta > \delta_{\text{crit}}|M) = 1 - \text{erf}(\nu/\sqrt{2}). \quad (119)$$

That is, the factor of 2 now comes in naturally.

## X. GALAXY CLUSTERING IN THE PRESS-SCHECHTER FORMALISM

We previously used the Press-Schechter Formalism to convert a Gaussian random density field, combined with the linear theory of growth of structure, to predict number density of collapsed objects as a function of their mass  $M$ . We denoted this number density with  $n(M)$ , which is also referred to as the halo mass function. This halo mass function give some very useful insights into the origin of ‘hierarchical’ structure formation, and the shape and redshift evolution of the (UV) luminosity function of galaxies. We mentioned previously that there is another important statistical description of galaxies, namely their clustering properties (see § VIII.B). Our goal is to gain insight into the clustering properties of galaxies by studying the properties of collapsed dark matter halos in the context of Press-Schechter formalism.

Before doing so, we return to the two-point correlation function  $\xi(t)$ , which quantifies the clustering. Pick two volume elements  $dV$  labelled ‘1’ and ‘2’, separated by a distance  $r_{12}$ , and correlate the number of galaxies in both:

$$\begin{aligned} dV^2 \langle n_{\text{gal},1} n_{\text{gal},2} \rangle &= dV^2 \langle \bar{n}_{\text{gal}}(1 + \delta_{\text{gal},1}) \bar{n}_{\text{gal}}(1 + \delta_{\text{gal},2}) \rangle = \\ &= dV^2 \bar{n}_{\text{gal}}^2 \langle 1 + \delta_{\text{gal},1} + \delta_{\text{gal},2} + \delta_{\text{gal},1}\delta_{\text{gal},2} \rangle = \\ &= dV^2 \bar{n}_{\text{gal}}^2 \left( \langle 1 \rangle + \langle \delta_{\text{gal},1} \rangle + \langle \delta_{\text{gal},2} \rangle + \langle \delta_{\text{gal},1}\delta_{\text{gal},2} \rangle \right). \end{aligned} \quad (120)$$

If the galaxy distribution is Gaussian like the mass density field (which is a reasonable assumption as we will see at the end of § 1.3), then from our lectures on Gaussian random fields we know that  $\langle \delta_{\text{gal}} \rangle = 0$ , and  $\langle \delta_{\text{gal},1}\delta_{\text{gal},2} \rangle = \xi_{\text{gal}}(r_{12})$ . We therefore have

$$dV^2 \langle n_{\text{gal},1} n_{\text{gal},2} \rangle = dV^2 \bar{n}_{\text{gal}}^2 \left( 1 + \langle \delta_{\text{gal},1}\delta_{\text{gal},2} \rangle \right) \equiv dV^2 \bar{n}_{\text{gal}}^2 (1 + \xi_{\text{gal}}(r_{12})). \quad (121)$$

If we pick the volume element to be sufficiently small that the number density of galaxies  $n_{\text{gal}}dV \ll 1$ , then  $n_{\text{gal}}dV$  denotes the probability that a volume element contains a galaxy. We can see that  $\xi_{\text{gal}}(r_{12})$  denotes *excess probability* over random that the two volume elements have galaxies inside them (recall the analogy of the excess probability of finding another person within 10m from a randomly chosen person on Earth).

### A. The Two-Point Function of Collapsed Objects and ‘Bias’ in the PS Formalism

Our goal is now to compute  $\xi(r)$  of mass inside collapsed objects. These collapsed objects is where galaxies are thought to form. Understanding the clustering properties of mass in collapsed objects therefore can help us understand clustering properties of galaxies.

First recall that in the Press-Schechter formalism a mass element is part of a collapsed object when its linear over density  $\delta_{\text{lin}} \geq \delta_{\text{crit}} = 1.69$ . How does this requirement affect  $\xi(r)$ ? The number density of collapsed objects (in other words, dark matter halos) within the mass range  $M \pm dM/2$  is denoted by  $n(M)dM$ . Press-Schechter theory provided us with an expression for  $n(M)$  (see Eq 113):

$$M^2 n(M) = \nu \exp(-\nu^2/2), \quad \nu \equiv \delta_{\text{crit}}/\sigma(M), \quad (122)$$

where  $k = \rho_m |\frac{\partial \log \sigma}{\partial \log M}| \sqrt{\frac{2}{\pi}}$  is a constant.

We first address how the number density of collapsed objects changes if we move into a part of the Universe where the *average* overdensity with some (arbitrary) volume  $V$  equals  $\delta = \epsilon > 0$ . In this region we expect a boost in the number density of objects because of two effects:

1. If the local overdensity (averaged over the volume  $V$ ) is  $\delta = \epsilon$ , then the local density of dark matter (and baryonic) particles - and consequently dark matter particles in collapsed objects - is higher by a factor of  $(1 + \epsilon)$ .

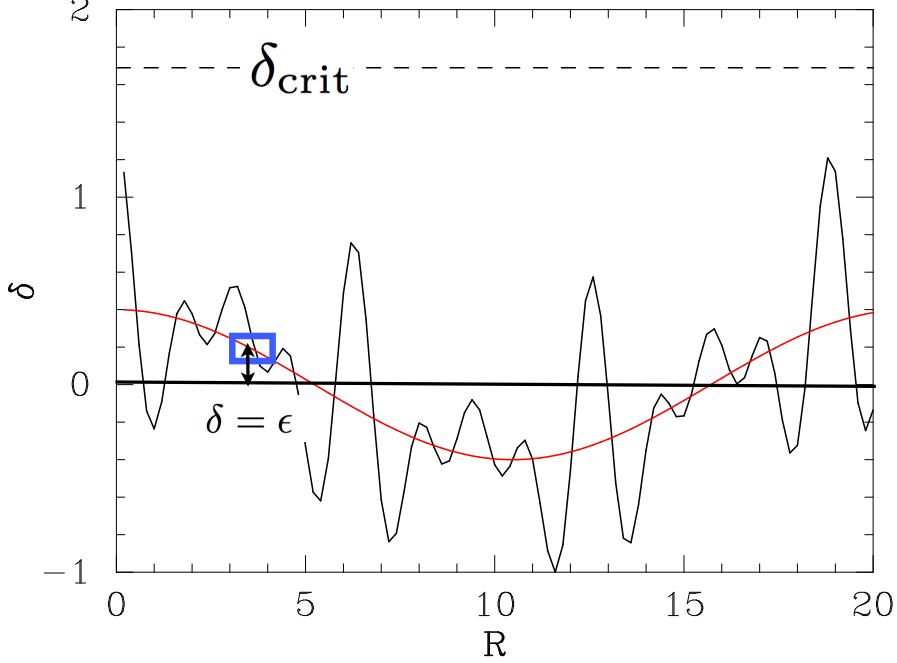


FIG. 18 This Figure shows the mass overdensity  $\delta$  as a function of position  $R$  (arbitrary units) along a random line through the universe. The mass density field fluctuates, and the red solid line denotes a long-wavelength fluctuation. Because of this long-wavelength fluctuations, there are large regions in Universe for which the average overdensity is not 0 (as is the case for the universe as a whole, recall that  $\langle \delta \rangle = 0$  for a Gaussian random density field). For example, in the blue box the mean overdensity is  $\delta = \epsilon$ . The number density of collapsed objects in a volume where the mean overdensity is elevated by an amount known as ‘bias’.

2. If the local mean overdensity is  $\delta = \epsilon$ , then the collapse barrier for most dark matter particles lies closer on average, at  $\delta_{\text{crit}} - \epsilon$  (see Figure 1).

We consider two effects separately: the first effect boosts  $n \rightarrow n(1 + \epsilon)$ . The second effect requires more work: the barrier for collapse  $\delta_{\text{crit}}$  lies lower at  $\delta_{\text{crit}} - \epsilon$ . The resulting change in  $n(M)$  at a fixed mass can be obtained by applying the chain rule:

$$n \rightarrow n + \frac{\partial n}{\partial \nu} \frac{\partial \nu}{\partial \epsilon} \epsilon. \quad (123)$$

We know that the ‘local’ collapse barrier is lowered to  $\delta_{\text{crit}} - \epsilon$ , and therefore

$$\nu = [\delta_{\text{crit}} - \epsilon]/\sigma \Rightarrow \frac{\partial \nu}{\partial \epsilon} = -\frac{1}{\sigma}. \quad (124)$$

Substituting this into Eq 3 we get

$$\begin{aligned} n &\rightarrow n - \frac{\epsilon}{\sigma} \frac{\partial n}{\partial \nu} = n - \frac{\epsilon}{\sigma} \left( \frac{k \exp(-\nu^2/2)}{M^2} - \frac{k \nu^2 \exp(-\nu^2/2)}{M^2} \right) = \\ &= n - \frac{\epsilon}{\sigma} \frac{n}{\nu} \left( 1 - \nu^2 \right) = n \left[ 1 - \frac{\epsilon}{\sigma} \left( \frac{1 - \nu^2}{\nu} \right) \right] = n \left[ 1 + \frac{\epsilon}{\sigma} \left( \frac{\nu^2 - 1}{\nu} \right) \right] \equiv n + \delta n. \end{aligned} \quad (125)$$

So we have

$$\frac{\delta n}{n} \equiv \left( \frac{\nu^2 - 1}{\sigma \nu} \right) \epsilon \equiv b_L \epsilon \quad (126)$$

The quantity  $b_L$  denotes the ‘Lagrangian’ bias. This bias accounts for the boost in the local number density of collapsed objects as a result of the lowering of the collapse threshold.

The total boost in the number density of collapsed object now changes as

$$n \rightarrow n(1 + \epsilon)(1 + b_L\epsilon) \approx n(1 + \epsilon[1 + b_L]) \equiv n(1 + b_E\epsilon) \quad (127)$$

, where we have defined the Eulerian bias  $b_E$

$$b_E = 1 + b_L = 1 + \frac{\nu^2 - 1}{\sigma\nu}.$$

(128)

The Eulerian bias parameter includes the boost in the number density of collapsed objects - in response to a an enhancement of the local density by a factor of  $1 + \epsilon$  - due to booth the increase in the local density, and due to the local lowering of the collapse barrier. So simply speaking, this accounts for the total boost.

Now we return to the galaxy correlation function

$$\xi_{\text{gal}}(r_{12}) = \langle \delta_{\text{gal},1} \delta_{\text{gal},2} \rangle. \quad (129)$$

Note that here we defined  $n_{\text{gal}} = \bar{n}(1 + \delta_{\text{gal}})$ . From the analysis above we know that  $\delta_{\text{gal}} = b_E\epsilon$  (here we used that galaxies reside in collapsed objects). We therefore have

$$\xi_{\text{gal}}(r_{12}) = b_E^2 \langle \epsilon_1 \epsilon_2 \rangle = b_E^2 \xi(r_{12}),$$

(130)

where  $\xi(r_{12})$  denotes the 2-point correlation function of the overall mass density field<sup>14</sup>. This is an important result: the correlation function of galaxies - and more generally of mass that resides inside collapsed objects - is the same as that for the overall mass distribution **multiplied by the bias parameter squared**<sup>15</sup>.

There are four regimes that we look at

1.  $\nu = 1 \Rightarrow b_E = 1$ . Collapsed objects with  $\nu = 1$  have the same 2-pt function  $\xi(r)$  as all matter (irrespective of whether it is part of a bound object or not). Note that  $\nu = 1$  corresponds to objects with masses where the PS mass function stars to turn-over from power-law to exponential decrease: i.e.  $\nu = 1$  corresponds to the characteristic mass of the PS mass-function.
2.  $\nu < 1 \Rightarrow b_E < 1$ . Collapsed objects with  $\nu < 1$  have a 2-pt function  $\xi(r)$  that is suppressed to that of all matter. These objects are said to ‘cluster less’, or said to be ‘anti-biased’ tracers of the total matter distribution. Objects with  $\nu < 1$  lie in the power-law low-mass end of the PS mass function.
3.  $\nu > 1 \Rightarrow b_E > 1$ . Collapsed objects with  $\nu > 1$  have a 2-pt function  $\xi(r)$  that is enhanced to that of all matter. These objects are said to ‘cluster stronger’, or said to be ‘biased’ tracers of the total matter distribution. These objects lie in the exponential tail of the halo mass function and are therefore more rare.
4.  $\nu \gg 1 \Rightarrow b_E \sim \frac{\nu}{\sigma}$ . Note that  $\nu \equiv \delta_{\text{crit}}/\sigma(M) \gg 1$ . i.e.  $\sigma(M) \ll 1$ . Therefore  $b_E \gg \nu \gg 1$ . The rarest objects are very strongly clustered!

In the lectures, we discussed some applications of bias:

1. As we mentioned earlier (see § VIII.B), the more luminous a galaxy, the more clustered the have been observed to be. This implies that more massive halos, host more luminous galaxies. This perhaps not surprising from an intuitive point of view, and is also consistent with the qualitative agreement between the shape of the dark matter halo mass function, and the observed galaxy luminosity functions.

<sup>14</sup> Note that  $\epsilon$  was just another symbol that denoted local mass overdensity, we used  $\delta$  for this in many other lectures.

<sup>15</sup> This also justifies that we assumed in § 1.2 that the distribution of galaxies was described by a Gaussian random field.

2. The very strong clustering of very rare objects, implies for example that the very first galaxies (which formed in the rare first minihalos). The correlation length of such objects is therefore very large. This implies that if you want to simulate certain processes in the Universe, such as the ‘reionization’ process of the Universe (we will return to this later), we need large simulations that contain fluctuations on these very large scales.
3. the bias parameter is mostly a function of mass: i.e.  $\xi_{\text{coll}}(r|M)$  *mostly* has the same dependence on radius as  $\xi(r)$ , i.e. there is *little* scale-dependence to the bias (which can be important if you want to do precision cosmology with galaxy-clustering).

## XI. COSMOLOGICAL USES OF GALAXY CLUSTERING

The previous discussion illustrates that galaxies (especially luminous galaxies) are biased tracers of the underlying density field. The statistical properties of the cosmological density field contains a wealth of cosmological information. We next discuss a few applications of how clustering measurements (mostly of galaxies, but the discussion applies to any tracer of the underlying cosmological density field) can constrain cosmological parameters, and that these constraints are affected only weakly by uncertainties in the galaxy formation process.

### A. Acoustic Scale and Baryon Acoustic Oscillations

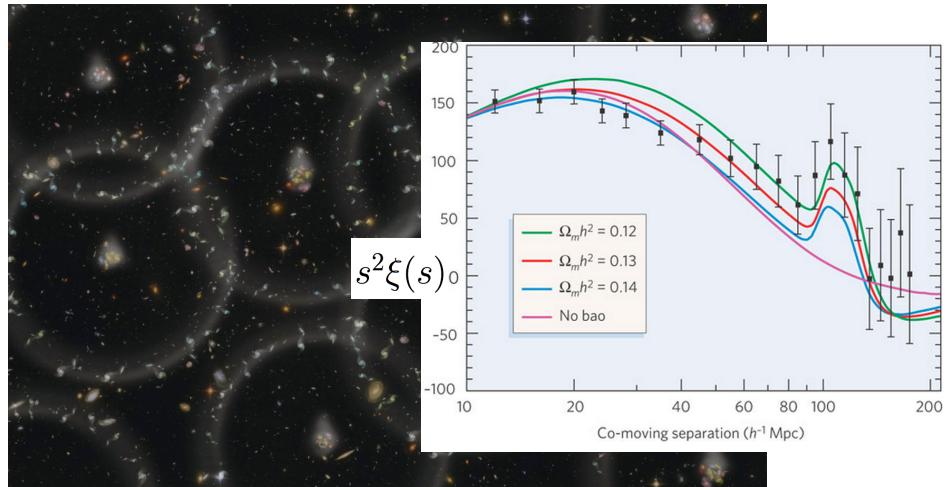


FIG. 19 .

Density perturbations that were generated during the very first moments after the Big Bang generated soundwaves that propagated through the primeval plasma. In the lecture we discussed how these soundwaves left a feature in  $\xi(r)$ . The description I gave in the lecture was based on that by D. Eisenstein (who is an expert in this area). The website can be found at [https://www.cfa.harvard.edu/~deisenst/acousticpeak/acoustic\\_physics.html](https://www.cfa.harvard.edu/~deisenst/acousticpeak/acoustic_physics.html). I have uploaded a PDF version of this site as well (EisensteinAcousticpeak.pdf). The so-called ‘acoustic peak’ is a feature at  $R_{\text{ac}}$  100 cMpc (cMpc stands for comoving Mpc) in the two-point correlation function of matter, and to the total distance a pressure wave could travel from a point perturbation in the cosmological density-field from  $t = 0$  to  $t = t_{\text{rec}}$ , where  $t_{\text{rec}}$  denotes the age of the Universe at recombination.

Because galaxies provide biased tracers of  $\xi(r)$  the acoustic peak should be visible in the clustering of galaxies as well. The *left panel* of Figure 19 shows an artist’s impression of the feature in  $\xi(r)$  as a characteristic clustering length. The *right panel* shows the actually observed clustering of luminous galaxies. Note that this plot shows  $r^2 \times \xi(r)$  to emphasize the feature. The presence of this acoustic peak is extremely interesting for cosmological purposes. We can compute what size  $R_{\text{ac}}$  should be, which requires ‘basic’ plasma physics and depends on the cosmological parameters.

Observations of galaxy clustering give us an angular size of the acoustic peak, and as a function of redshift. This angular size,  $\theta_{\text{ac}}$  relates to  $R_{\text{ac}}$  through the standard relation (see Eq 12):

$$R_{\text{ac}} \equiv \theta_{\text{ac}}(z) d_{\text{A}}(z). \quad (131)$$

Measuring the angular size, and as a function of redshift provides robust constraints on cosmological parameters.

### B. Two 2D-Two-Point Function I: The Alcock-Paczynski Test

We can get other cosmological constraints from  $\xi(r)$ , namely its properties in 2-dimensions. The two-point correla-

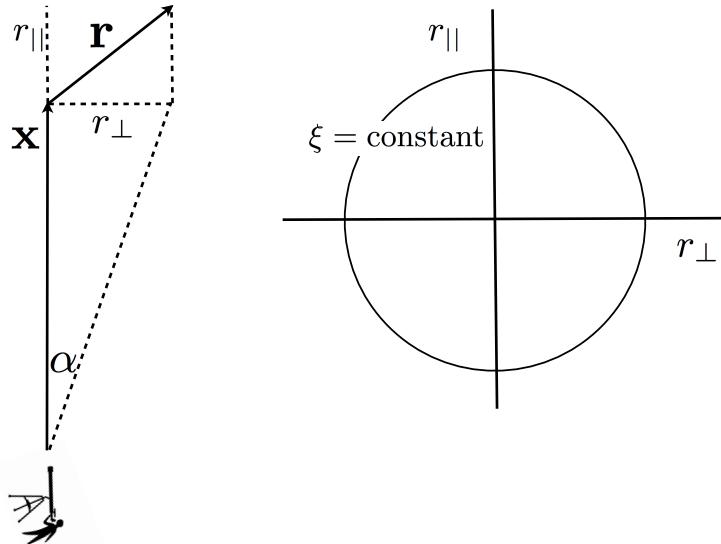


FIG. 20 This Figure shows how a 3D separation  $\mathbf{r}$  is usually separated into a transverse ( $r_{\perp}$ ) and line-of-sight ( $r_{\parallel}$ ) separation (on the left). The observed is located in the bottom of the figure. Isotropy requires that contours of constant  $\xi$  are circular in the  $r_{\perp} - r_{\parallel}$  plane (shown on the right).

tion function  $\xi(\mathbf{r})$  gives us the excess probability over random of finding two objects (e.g. galaxies) at some separation  $\mathbf{r}$ . However, when observing astrophysical objects, we do not know the 3D separation  $\mathbf{r}$  perfectly. In practise we write  $\xi(\mathbf{r}) \equiv \xi(r_{\perp}, r_{\parallel})$ . Here,  $r_{\perp}$  denotes the ‘transverse’ separation and  $r_{\parallel}$  denotes the line-of-sight separation. These are indicated schematically in Figure 20. The transverse separation corresponds to the angular separation (in radians) times the angular diameter distance  $d_{\text{A}}(z)$  to redshift  $z$ .

$$r_{\perp} \equiv \alpha d_{\text{A}}(z). \quad (132)$$

The line-of-sight separation is inferred from the velocity separation as:

$$r_{\parallel} \equiv \Delta v / H(z). \quad (133)$$

Isotropy requires that  $\xi(r_{\perp}) = \xi(r_{\parallel})$  if  $r_{\perp} = r_{\parallel}$ . In other words, isotropy requires that contours of constant  $\xi$  should be circles in the  $r_{\perp} - r_{\parallel}$  plane. Recall that the conversion from the observable quantities ( $\alpha, \Delta v$ ) into ( $r_{\perp}, r_{\parallel}$ ) requires knowledge of  $H(z)$  and  $d_{\text{A}}(z)$ . Adopting the wrong set of cosmological parameters will give the wrong conversion, and will generally not yield circular contours. If we enforce the contours of  $\xi$  in the  $r_{\perp} - r_{\parallel}$  plane to be circular, then this translates to constraints<sup>16</sup> on  $d_{\text{A}}(z)$  and  $H(z)$ , and therefore on cosmological parameters. This cosmological test

---

<sup>16</sup> We get circular contours not only for the correct cosmological parameters, but also if we simultaneously adopt the incorrect  $d_{\text{A}}(z)$  by

is known as the Alcock-Paczynski test.

### C. Two 2D-Two-Point Function II & 2D Power Spectrum: Redshift Space Distortions

The Alcock-Paczynski test is strictly applicable only accurate when peculiar velocities are negligible. For large values of  $r$  in  $\xi(r)$  the expansion of space overwhelms peculiar velocities (as  $v_0 = Hr$ ), and this is the case. However, this is not the case for smaller  $r$ . As in previous lectures, we decompose the velocity  $\mathbf{v}$  of an object into an ‘unperturbed’ velocity  $\mathbf{v}_0$  and a peculiar velocity  $\mathbf{u}$ . That is,  $\mathbf{v} = \mathbf{v}_0 + \mathbf{u}$ . Peculiar velocities cause so called *redshift-space distortions*: consider an overdensity  $\delta > 0$ . The gravitational potential generated by this overdensity attracts nearby matter. Recall from linear theory that  $\delta = -\nabla \cdot \mathbf{u}$ . From linear theory we know that if  $\delta > 0$ , then  $\delta > 0$  and therefore  $\nabla \cdot \mathbf{u} < 0$ . In other words, the overdensity is surrounded by a peculiar velocity vectors pointing at the central over density  $\delta$ . This is illustrated visually in the *left panel* of Figure 21.

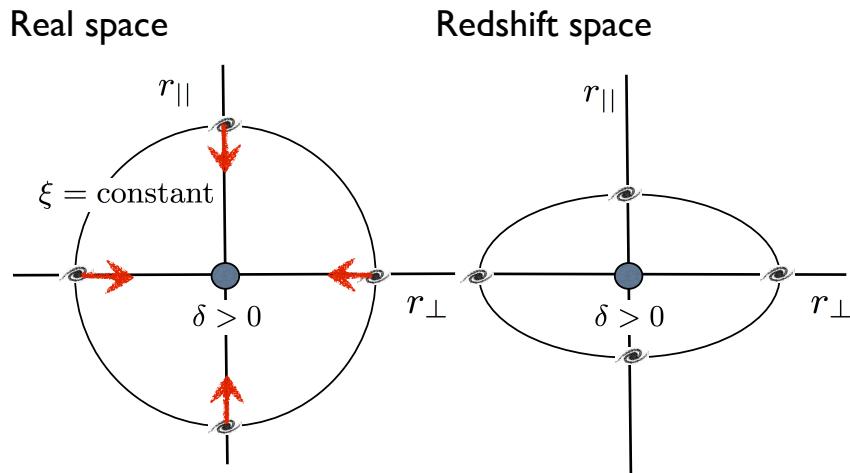


FIG. 21 .

*Redshift space* corresponds to the space in which the radial coordinate of an object comes entirely from its measured velocity. Peculiar velocities thus induce distortions in the distribution of galaxies in redshift space in the line-of-sight direction. This is illustrated visually in the *right panel* of Figure 21: the peculiar velocity of a galaxy on the far side of the perturbation is pointing toward us: hence, the total recessional velocity of this galaxy appears smaller. It would therefore appear closer to the overdensity in redshift space than it is in reality. Peculiar velocities induced by an overdensity thus ‘squash’ (compress) the two-point correlation along the line-of-sight direction. We denote this two point function in redshift space with  $\xi_s(\mathbf{x}_s)$ . More generally, a subscript ‘s’ on any quantity means it is measured in redshift space.

The squashing of  $\xi_{s,g}(\mathbf{x}_s)$  along the line-of-sight direction affects the redshift space power spectrum  $P_{s,g}(\mathbf{k}_s)$ . Note that the extra label ‘g’ is two emphasis that we now look at the galaxy two-point function and galaxy power spectrum. Note that large  $r_{||}$  correspond to small  $k_{||}$  (and vice versa), and the galaxy power spectrum in redshift space is squashed along the transverse direction. This is illustrated visually by Figure 22.

The deformation of the 2D galaxy power spectrum in redshift space is quantitatively given by the *Kaiser formula*:

---

a factor of  $x$  and  $H(z)$  by a factor of  $x^{-1}$  (this results in the same boost  $x$  in both  $r_{\perp}$  and  $r_{||}$ , which preserves the circularity of the contours). Enforcing circular constraints therefore constrains the product  $H(z)d_A(z)$ .

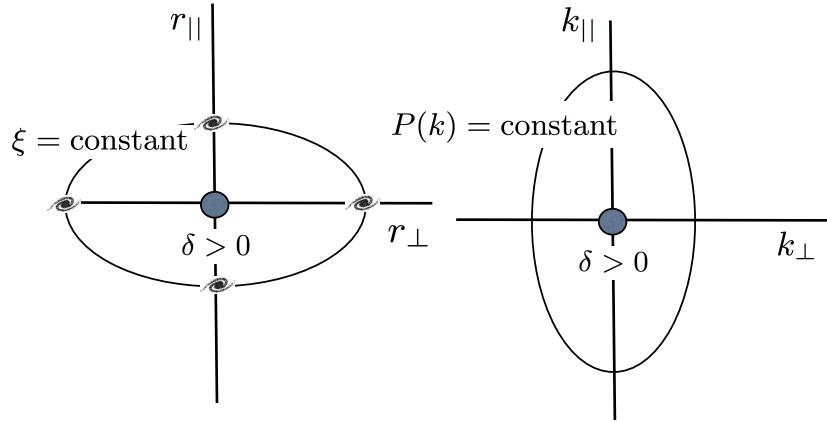


FIG. 22 .

$$P_{s,g}(\mathbf{k}_s) = P_g(k)[1 + \beta\mu_{\mathbf{k}_s}^2]^2 \quad (134)$$

$$\beta \approx \frac{\Omega_{m,0}^{0.6}}{b_E}, \quad \mu_{\mathbf{k}_s} \equiv [\mathbf{k}_s \cdot \mathbf{e}_{||}] / k_s, \quad k_s = |\mathbf{k}_s|,$$

where  $\mathbf{e}_{||}$  denotes a unit-vector in the line-of-sight direction. So we therefore have

$$P_{s,g}(\mathbf{k}_s) = P_g(k)[1 + \beta]^2 \quad \mu_{\mathbf{k}_s} = 1 \quad (135)$$

$$P_{s,g}(\mathbf{k}_s) = P_g(k) \quad \mu_{\mathbf{k}_s} = 0.$$

Importantly,  $\beta$  depends on the cosmological mass density parameter  $\Omega_{m,0}$  and the bias of the galaxies we are studying. The deformation of the 2D power spectrum in redshift therefore encodes cosmological information as well as additional constraints on the bias of the galaxies.

The Kaiser deformation is due to the effects of gravity slowing down the expansion on large scales. On smaller scales, virialization also affects  $\xi(r)$ : the virialized motion of galaxies through a galaxy cluster, gives them peculiar L.O.S motions which can be as large as  $\sim 10^3$  km s $^{-1}$  in random directions. A galaxy cluster would therefore appear smeared out in redshift over  $\sim 10^3$  km s $^{-1}$ . This elongation of massive structures along the L.O.S. gives rise to the ‘Fingers-of-God’ effect, which is shown in Fig 23, which shows observed clustering of galaxies in redshift space. Note that on the largest scales - not affected by Kaiser squashing, or the Fingers-of-God effect, we expect circular contours (as we discussed above).

## XII. WHAT SHAPES THE MATTER POWER SPECTRUM $P(k)$

The power spectrum  $P(k)$  emerged in several of our previous lectures:

- It fully characterised the properties of a Gaussian random (density) field.
- It directly determines  $\sigma(M)$ , which directly determines  $n(M)$  (in Press-Schechter theory, and in general).

Now we will discuss some key properties of  $P(k)$ . Linear theory allowed us to describe the time evolution of a density perturbation  $\delta$  with wavenumber  $k$ . This time evolution was

$$\delta(k, t) = \delta(k, t = 0)D_+(t), \quad (136)$$

## Redshift-Space Distortions

Examples of the observed 2D galaxy correlation function in redshift space.

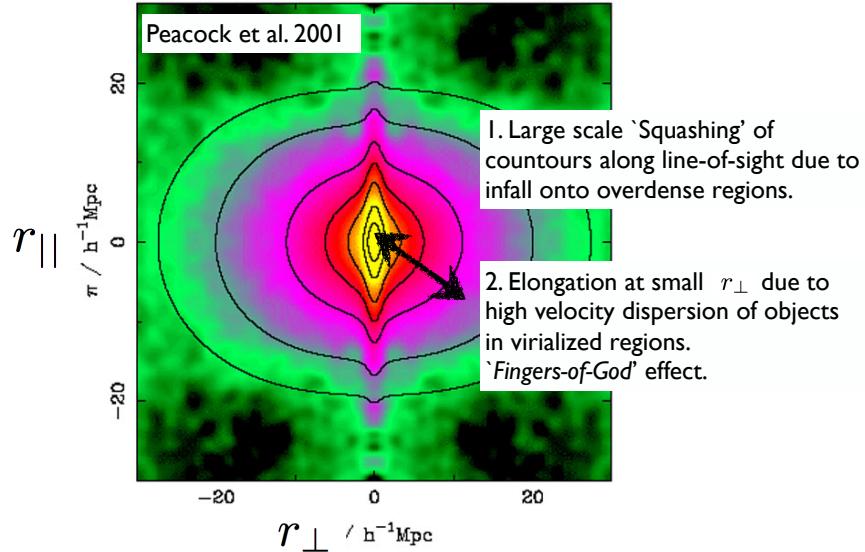


FIG. 23

where we derived that  $D_+(t) \propto t^{2/3} \propto a$  in the Einstein-de Sitter Universe ( $\Omega_m = 1.0, \Omega_\Lambda = 0.0$ ). In **the assignment** you are asked to show that  $P(k) = \langle |\delta(k)|^2 \rangle$ . It is then obvious that

$$P(k, t) = P(k, t=0) D_+^2(t) \equiv P_0(k) D_+^2(t). \quad (137)$$

One of the goals of modern cosmology is to calculate  $P_0(k)$ .

A good guess is that  $P_0(k)$  is a power law:

$$P_0(k) = Ak^n \quad (138)$$

as it has no preferred length scale, which represents that there are no preferred length scales in the very early Universe. We may think of this  $P_0(k)$  as the ‘primordial’ power spectrum as it describes density fluctuations at  $t = 0$ . As we will see next, perturbations with different wave numbers evolved differently in the very early Universe. This modifies the matter power spectrum from the power law form given above. These modifications are encoded in the so-called ‘transfer function’  $T(k)$ . This transfer function  $T(k)$  encodes the information on the evolution of some density perturbation  $\delta(k)$ , and therefore affects the power spectrum as

$$P(k, t) = Ak^n T^2(k) D_+^2(t). \quad (139)$$

Our goal is now to obtain some intuition for  $T(k)$ , and to derive expressions for the behaviour of  $T(k)$  in the limiting cases of very large and very small perturbations. Before proceeding it is useful to discuss the concept of a ‘horizon’, as it plays a key role in shaping the transfer function.

### A. Intermezzo: The Horizon.

There are many horizons in cosmology. We focus on the so called ‘particle horizon’, which corresponds to the ‘maximum proper distance over which there can be causal communication at time  $t$ ’ (Quote lifted from M. Longair’s

book). The particle horizon thus corresponds to the maximum proper distance a photon could have travelled between  $t = 0$  and  $t = t$ . This distance corresponds to

$$r_H(t) = a(t) \int_0^t \frac{cdt'}{a(t')}.$$
 (140)

The integral gives the total comoving distance traversed by the photon. The term  $a(t)$  converts that into a proper distance at time  $t$ . During radiation domination the scale factor  $a(t) \propto t^{1/2} \equiv Ct^{1/2}$ , where  $C$  is a constant. We can solve the integral to give  $r_H = 2ct$ . This is a factor of two larger than the maximum distance a photon could travel in a static medium. The factor of 2 accounts for the fact that space itself is expanding during the photon's flight.

Now consider the evolution of a perturbation of some comoving length  $L$ . The proper (i.e. real physical) length of this perturbation  $L_p = a(t)L$  (where the subscript p denotes proper). The proper size of the perturbation increases as  $L_p \propto a$ . We previously derived that  $a(t) \propto t^{2/3}$ . This derivation assumed that matter provided the dominant contribution to the Universal energy density. We can do a similar derivation under the assumption that radiation dominates, which is the case at earlier times. Under this assumption,  $a(t) \propto t^{1/2}$ . We therefore have that  $L_p \propto t^{1/2}$ , while the particle horizon  $r_H \propto t$ .

Suppose that  $L_p > r_H$  at some early time  $t_1$ , then the different time dependence of  $r_H$  and  $L_p$  imply that  $L_p = r_H$  at some later time  $t_2$ . We call the moment that  $L_p = r_H$  the moment of *horizon entry*. Horizon entry plays the most important role in shaping the transfer function  $T(k)$ , and therefore the power spectrum  $P(k)$ .

## B. The Transfer Function $T(k)$ .

Figure shows the time evolution of a perturbation  $\delta(k)$  (with corresponding comoving wavelength or ‘size’ of the perturbation  $L = 2\pi/k$ ) that enters the horizon during the radiation dominated era at scale factor  $a_{\text{enter}}$ ). There are three key events during the evolution of this perturbation:

- Before the perturbation enters the horizon, at  $a < a_{\text{enter}}$ , it grows  $\delta \propto a^2$ . As I mentioned in the lecture, this follows from general relativistic perturbation theory. This is beyond the scope of this lecture. The following might provide some insight if you are familiar with general relativistic perturbation theory<sup>17</sup>: a density perturbation  $\delta_k$  generate perturbations in the gravitational potential  $\Phi_k$  which correspond to metric perturbation in general relativity (matter warps space time). The equation that described the time evolution of this metric perturbation corresponds to the perturbed Poisson equation from **lecture 2**, which in Fourier space reads  $-k^2\Phi_k = 4\pi G a^2 \bar{\rho} \delta_k$ . If we require that the *metric* perturbation cannot evolve for perturbations outside the horizon, then we must have that  $a^2 \bar{\rho} \delta_k$  is independent of time. We therefore must have  $\delta_k \propto a^2$  during radiation domination ( $\rho_{\text{rad}} \propto a^{-4}$ ), and  $\delta_k \propto a$  during matter domination ( $\rho_{\text{rad}} \propto a^{-3}$ ).
- When the matter density starts to dominate the Universal energy density - this happens at the redshift of matter-radiation equality  $z_{\text{eq}} \sim 24000$  - the dark matter perturbation grows as  $\delta \propto a$  during the matter dominated era. We derived this dependence in **Lectures 2-3** for perturbations that are smaller than the horizon size (as mentioned above, general relativity predicts exactly the same evolution for perturbations larger than the horizon size in the era).
- At redshifts  $z > z_{\text{eq}}$  - i.e.  $a < a_{\text{eq}}$  - the Universal energy density is dominated by radiation. During radiation domination the scale factor grows as  $a \propto t^{1/2}$  (as mentioned above). This different time dependence of the scale factor - combined with the fact that radiation dominates the Universal (mass-)energy density - gives rise to a drastically different predicted time evolution for  $\delta$ . As we will see next,  $\delta$  barely grows at all during radiation dominance. This ‘stalling’ of the growth of density perturbation in the radiation-dominated era is known as the Meszaros effect. The fluctuations are said to be frozen in the background. Mathematically, the Meszaros effect is easy to understand. Recall that the density evolution of a perturbation  $\delta$  was given by the following differential equation:

---

<sup>17</sup> I am not.

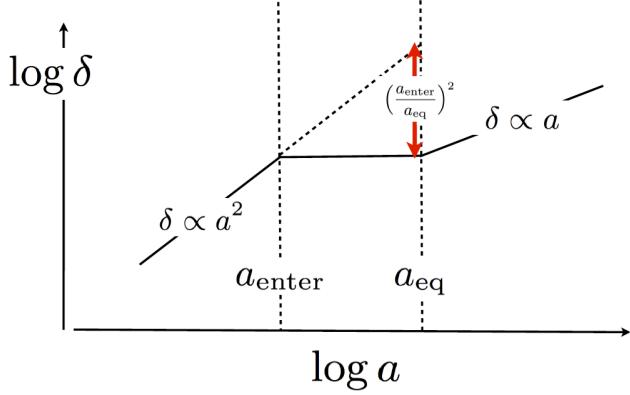


FIG. 24 This Figure shows (schematically) the time evolution of an overdensity  $\delta$  on some scale  $L$  that enters the horizon at  $a_{\text{enter}} \ll a_{\text{eq}}$ . The time evolution goes through three different phases: (i)  $\delta \propto a^2$  before horizon entry which follows from general relativistic perturbation theory; (ii)  $\delta = \text{constant}$  after horizon entry, and up until  $a_{\text{eq}}$ . This stalling of the growth of the perturbation is known as the Meszaris effect; (iii) when matter starts to dominate the Universal energy density  $\delta \propto a$  as we derived in previous lecture. This Figure illustrates that the Meszaris effect suppresses the growth of this perturbation by a factor of  $(a_{\text{enter}}/a_{\text{eq}})^2$  compared to uninhibited growth.

$$\ddot{\delta}_m + 2\frac{\dot{a}}{a}\dot{\delta}_m = 4\pi G\rho_m\delta_m. \quad (141)$$

Divide both sides by  $H^2 = \frac{8\pi G\rho_{\text{tot}}}{3}$ , where  $\rho_{\text{tot}} = \rho_{\text{rad}} + \rho_m$ . Using  $H = \frac{\dot{a}}{a}$  we find

$$\frac{\ddot{\delta}_m}{H^2} + \frac{2}{H}\dot{\delta}_m = \frac{3\rho_m\delta_m}{2[\rho_m + \rho_{\text{rad}}]}. \quad (142)$$

If we now use that deep in the radiation dominated era  $\rho_{\text{rad}} \gg \rho_m$ , then the term on the RHS can be ignored. This is because we are multiplying a small number  $\delta$  with another small number, and can see this term effectively as a second order term. The differential equation then simplifies to

$$\frac{\ddot{\delta}_m}{H} + 2\dot{\delta}_m = 0. \quad (143)$$

If we further use that  $H = \dot{a}/a = 1/[2t]$  then we are left with

$$\ddot{\delta}_m + \frac{\dot{\delta}_m}{t} = 0 \Rightarrow \delta_m = A + B\log t = A + C\log a. \quad (144)$$

The perturbation thus only grows logarithmically with the scale factor. This growth is represented by the horizontal line in Figure 1.

Figure 1 shows that for the perturbation that entered the horizon during radiation dominance at  $a_{\text{enter}}$  the growth was suppressed by a factor of  $T(k) = \left(\frac{a_{\text{enter}}}{a_{\text{eq}}}\right)^2$ . This suggests that  $T(k) \rightarrow 1$  if  $a_{\text{enter}} \rightarrow a_{\text{eq}}$ . Indeed, for perturbations that enter the horizon during matter domination we do not have any inhibition of growth:  $\delta \propto a^2$  during radiation dominance, and  $\delta \propto a$  during matter dominance (see previous discussion). This discussion clearly suggest that there is a particular scale of interest, namely that for which  $a_{\text{enter}} = a_{\text{eq}}$ . This corresponds to the *smallest* scale for which there is no suppression of growth by the Meszaros effect (for larger scales, perturbations enter the horizon after matter-radiation equality and there is no Meszaros suppression either). The comoving size of this perturbation is therefore equal to the horizon scale at matter-radiation equality, over the scale factor (because remember that  $r_H$  was a proper distance!):

$$L_0 = r_H(a_{\text{eq}})/a_{\text{eq}} = \int_0^{t_{\text{eq}}} \frac{cdt}{a(t)} = \dots \underset{\text{assignment 5}}{=} \frac{c}{H_0} \frac{1}{\sqrt{2\Omega_{\text{m},0} z_{\text{eq}}}} \approx 80 \text{ cMpc}, \quad (145)$$

where  $\Omega_{\text{m},0}$  denotes the present-day mass density parameter,  $H_0$  is the present-day Hubble constant, and ‘cMpc’ denotes comoving Mpc (just to emphasize that  $L_0$  is a comoving quantity). The corresponding wavenumber is  $k_0 = 0.1 \text{ cMpc}^{-1}$ .

Finally, we would like to express  $T(k)$  as a function of  $k$ . We found that the suppression  $T(k)$  for perturbations entering at  $a < a_{\text{eq}}$  was given by  $T(k) = \left(\frac{a_{\text{enter}}}{a_{\text{eq}}}\right)^2$ . We would like to express  $T(k)$  as a function of  $k$ , rather than  $a(t)$ . We can do via the condition for horizon entry which states that  $La(t_{\text{enter}}) = 2ct_{\text{enter}}$ , where  $t_{\text{enter}}$  denotes the age of the Universe when  $L$  enters the horizon. Since  $a(t) = kt^{1/2}$  ( $k$  is just a constant), we can recast the horizon-entry condition as  $L \propto t_{\text{enter}}^{1/2} \propto a_{\text{enter}}$  (were  $a_{\text{enter}} \equiv a(t_{\text{enter}})$ ). In other words

$$\frac{a_{\text{enter}}}{a_{\text{eq}}} = \frac{L}{L_0} = \frac{k_0}{k}, \quad (146)$$

where we used that  $k = 1/L$ . We therefore have that

$$T(k) \propto \begin{cases} 1 & k \ll k_0 \\ k^{-2} & k \gg k_0, \end{cases} \quad (147)$$

Figure 25 shows the full form of the transfer function, including intermediate scales.

### XIII. THE MATTER POWER SPECTRUM $P(k)$

#### A. The Transfer Function $T(k)$

In the previous lecture we found that

$$T(k) \propto \begin{cases} 1 & k \ll k_0 \\ k^{-2} & k \gg k_0, \end{cases} \quad (148)$$

where  $L_0 = 1/k_0$  denotes the (comoving) particle horizon size at matter-radiation equality.

#### B. The Power Spectrum $P(k)$ .

The power spectrum is given by  $P_0(k) = Ak^n T^2(k)$ . We have specified  $T(k)$ . The slope of the ‘primordial’ power spectrum has been inferred (from the Cosmic Microwave Background) to be close to 1 (see Fig 26). This value is

## Transfer Function

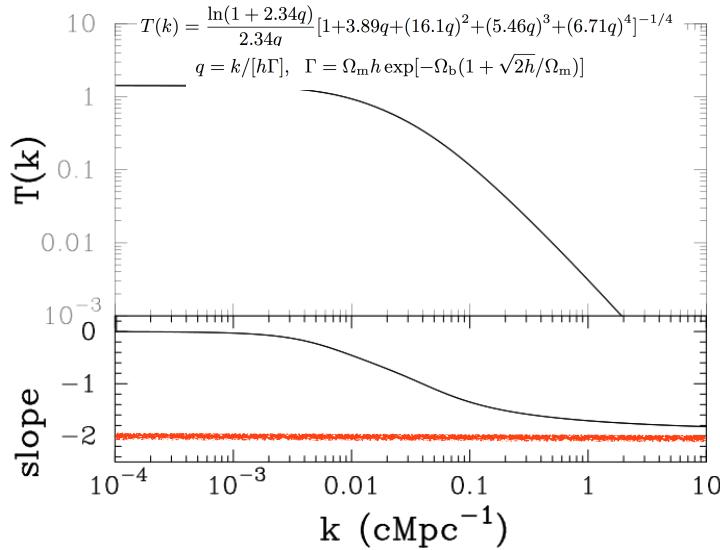


FIG. 25 .

predicted naturally by inflation theories. The power spectrum therefore scales as

$$P(k) \propto \begin{cases} k & k \ll k_0 \\ k^{-3} & k \gg k_0, \end{cases} \quad (149)$$

with a turn-over at  $k = k_0$ . Figure 26 shows recent observational constraints on  $P(k)$ . It is worth stressing that the observed turn-over in the power spectrum from  $n \sim 1$  on large scales to  $n \sim -3$  on small scales is a direct consequence of general relativity which predicts that  $\delta \propto a^2$  on superhorizon scales.

Several additional comments on  $P(k)$

- The normalisation constant  $A$  is obtained by matching to observations of the Cosmic Microwave Background.
- Baryons affect precise shape of  $P(k)$  (see paper by Eisenstein & Hu 1998). One example of how baryons affect the mass power spectrum was given in our discussion of the acoustic peak in the two-point correlation function (the single acoustic peak in  $\xi(r)$  corresponds to a series of oscillations in the power spectrum).
- The case  $n = 1$  corresponds to a special case which yields scale-invariant fluctuations. This is discussed next.

### C. Why $n = 1$ corresponds to Scale Invariance.

We derived the relation between  $P(k)$  and the variance in the mass density field averaged over some mass-scale  $M$  in previous lectures. The RMS (root mean square) amplitude of fluctuations smoothed over mass-scale  $M$  is given by

$$\sqrt{\sigma^2(M)} \equiv \Delta(M) \propto M^{-(n+3/6)} \underset{n=1}{\propto} M^{-2/3}. \quad (150)$$

That is, the RMS amplitude of density fluctuations as a function of mass scale  $M$  decreases with mass. However, the higher the mass, the later horizon entry and the less suppression by the Meszaros effect. The Meszaros effect thus tends to equalise the RMS amplitude of density fluctuations as a function of mass  $M$ . We previously noted that  $n \rightarrow -3$  for small scales, which implies that the variance  $\Delta M$  should not depend on  $M$  on these scales.

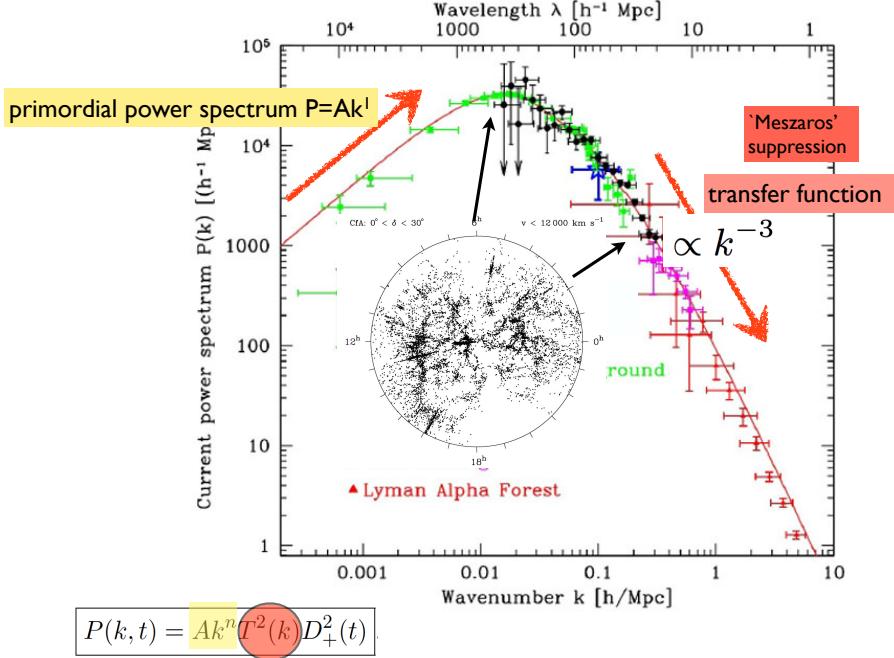


FIG. 26 .

Another way to look at this is by considering the time evolution of  $\Delta M$  prior to horizon entry,  $a_{\text{enter}} < a_{\text{eq}}$  (i.e. the mass scale  $M$  enters the horizon before matter radiation equality). For general relativity dictates that  $a < a_{\text{enter}} \delta \propto a^2$ , and so

$$\Delta(M) \propto a^2 M^{-2/3}. \quad (151)$$

We can relate the mass  $M$  of a perturbation to the scale factor at which it entered the horizon as

$$M = M_h(t) \propto \rho_m r_H^3 \propto a^{-3} t^3 \propto t^{-3/2} t^3 \propto t^{3/2}, \quad (152)$$

where we used that the mass density  $\rho_m \propto a^{-3}$ , that  $a(t) \propto t^{1/2}$  during radiation domination, and that the proper horizon scale scales as  $r_H \propto t$ . Substituting this back into Eq 4 we have

$$\Delta(M) \propto a^2 M^{-2/3} \propto a^2 t^{-1} = \text{constant}. \quad (153)$$

That means that fluctuations that enter the horizon<sup>18</sup> at  $a = a_{\text{enter}} < a_{\text{eq}}$  therefore have the *same RMS amplitude* at horizon entry. The subsequent growth of these perturbations is stalled until  $a_{\text{eq}}$ , after which they all grow as  $\delta \propto a$ . This identical growth ensures that the RMS amplitude of these fluctuations remains independent of scale at all times. Although I have not shown it in the lecture, you can do the same analysis for perturbations that enter the horizon at  $a > a_{\text{enter}}$  and get the same result<sup>19</sup>: namely that  $\Delta(M) = \text{constant}$ ! This shows that *all* fluctuations enter the horizon with the same RMS amplitude. This remarkable scale invariance is a special property of the power spectrum with  $n = 1$ .

<sup>18</sup> Convince yourself that this corresponds to fluctuations on mass-scales that are relevant for astrophysical objects (galaxies, groups of galaxies, clusters of galaxies).

<sup>19</sup> As we noted in the lecture, we made several approximations to arrive at this precise scale invariance. In reality,  $\sigma(M)$ , is not exactly a constant.

#### D. Supplemental Reading Material I: Pre-Midterm

The following texts/books can be helpful:

- Expressions for virial radii, temperatures, and circular velocities for cosmologies other than Einstein-de-Sitter, see <http://adsabs.harvard.edu/abs/2001PhR...349..125B>, or <http://arxiv.org/pdf/astro-ph/0010468v3.pdf> (page 16).
- For the discussion on the relevance of the cooling times & collapse times for galaxy formation: The ‘classical’ papers by Rees & Ostriker 1977 (<http://adsabs.harvard.edu/abs/1977MNRAS.179..541R>), and White & Rees 1978 (<http://adsabs.harvard.edu/abs/1978MNRAS.183..341W>). Most textbooks will good discussion on this (I liked using Peacock’s cosmological physics: Chapter 17.3). Peacock gives the expression for galaxy mass in fundamental constants.
- On Press-Schechter Theory, random walks shown in Figure 17 and how to get the fudge factor of 2 I consulted books by J. Peacock (Cosmological Physics), and M. Longair (Galaxy Formation). More extended (& formal) descriptions can be found in the book by Mo, Van den Bosch & White. Different books give different descriptions & derivations. If anything is unclear, I would recommend trying different books until you find one that works best for you, or come talk to me.
- The non-technical description if the BAO ([EisensteinAcousticpeak.pdf](#)).
- Helpful additional discussion and (working) movies that support the notes by D. Eisenstein can be found on <http://scienceblogs.com/startswithabang/2008/04/25/cosmic-sound-waves-rule/>.
- The discussion/derivation of Eulerian & Lagrangian bias was taken from J. Peacock. ‘Cosmological Physics’. Chapter 17. Feel free to consult your preferred books on alternative views/discussions (I am sure that Mo, Van den Bosch & White will be extremely useful as well).
- For the discussion of Transfer functions I used the book by Peter Schneider ‘*Extragalactic Astronomy and Cosmology: An Introduction*’.
- Useful discussions on horizon scales in cosmology, and the difficulty of modelling super horizon scales are given in Longair ‘*Galaxy Formation*’. Specifically Chapter 12.2 + 12.3.
- The lectures by Frank van den Bosch have some very nice & clear slides. See [http://www.astro.yale.edu/vdbosch/astro610\\_lecture4.pdf](http://www.astro.yale.edu/vdbosch/astro610_lecture4.pdf) for a very brief discussion why  $\delta \propto a^2$  before horizon entry. His site also has references to his book (Mo, Van den Bosch & White) which provides much more details to these lectures. Can be very useful.

## XIV. STRUCTURE OF DARK MATTER HALOS & CUSP-CORE PROBLEM

### A. The Isothermal Density Profile

We would like to understand the structure of dark matter halos. In the standard cosmological model dark matter only interacts gravitationally, is collisionless and (hence) has no pressure. Because dark matter particles do not collide, they do not behave as an ideal fluid. Recall that ideal fluids were described completely by their density and pressure. In ideal fluids the velocity distribution of particles is completely isotropic, and the pressure provides a measure of the mean speed of the gas particles. In theory a dark matter halo can contain particles with more complex velocity distributions (for example, an in falling low mass halo can survive as a separate sub halo within a larger halo. This cannot happen for an ideal fluid).

If we want to derive the density profile of a dark matter halo then we cannot use the fluid equations we used before. Instead, we need to introduce the so-called ‘distribution function’,  $f(\mathbf{x}, \mathbf{v})$ , which corresponds to the (mass) density of particles in phase-space. Phase-space is the space spanned by 3D physical space and 3D velocity space, i.e. a point in 6D phase space has 6 coordinates  $(x, y, z, v_x, v_y, v_z)$ . The distribution function relates to more commonly encountered quantities density  $[\rho(\mathbf{x})]$  and the velocity distribution  $[f(\mathbf{v})]$  as

$$\begin{aligned}\rho(\mathbf{x}) &= \int d^3\mathbf{v} f(\mathbf{x}, \mathbf{v}) \\ f(\mathbf{v}) &= \int d^3\mathbf{x} f(\mathbf{x}, \mathbf{v}).\end{aligned}\tag{154}$$

Distribution functions help us accurately describe collision less systems such as stars and dark matter particles. At first glance, going to 6D to describe a system does not look very appealing at all. Fortunately, Boltzmann’s brilliance comes to the rescue here. Boltzmann proposed that for systems in ‘thermodynamic equilibrium’ the phase function has the following simple form:

$$f(\mathbf{x}, \mathbf{v}) \propto \exp\left(\frac{-E}{k_b T}\right),\tag{155}$$

where  $E$  denotes the total energy of a particle,  $k_b$  is the Boltzmann constant, and  $T$  denotes the temperature of the system (note that ‘temperature’ for a system of stars clearly does not mean ordinary temperature. Instead it provides a measure of the velocity dispersion of the stars.). The energy per dark matter particle is given by  $E = E(\mathbf{v}, \mathbf{x}) = \frac{1}{2}m_{\text{DM}}|\mathbf{v}|^2 + m_{\text{DM}}\Phi(\mathbf{x})$ , where  $m_{\text{DM}}$  denotes the mass of a dark matter particle, and  $\Phi$  denotes the gravitational potential (note that this is the ordinary gravitational potential, and not the perturbed one that we worked with in perturbation theory).

The dark matter density at  $\mathbf{x}$  is given by

$$\rho(\mathbf{x}) \propto \int d^3\mathbf{v} f(\mathbf{x}, \mathbf{v}) = \mathcal{N} \int_0^\infty 4\pi v^2 dv \exp\left(-\frac{m_{\text{DM}}v^2}{2k_b T}\right) \exp\left(\frac{-m_{\text{DM}}\Phi(\mathbf{x})}{k_b T}\right)\tag{156}$$

, where we have adopted the notation  $|\mathbf{v}| = v$ . The potential energy of a dark matter particle only depends on its position, and the term  $\exp\left(\frac{-m_{\text{DM}}\Phi(\mathbf{x})}{k_b T}\right)$  can be taken outside of the integral. The integral over  $v$  converges, and can be replaced with a number  $C_2(\mathbf{x})$ . Note that the  $\mathbf{x}$ -dependence of  $C_2$  enters entirely through  $T = T(\mathbf{x})$ . We define  $C_1(\mathbf{x}) = C_2(\mathbf{x})\mathcal{N}$ , and we are left with

$$\rho(\mathbf{x}) = C_1(\mathbf{x}) \exp\left(\frac{-m_{\text{DM}}\Phi(\mathbf{x})}{k_b T(\mathbf{x})}\right),\tag{157}$$

where I make the possible  $\mathbf{x}$ -dependence of  $T$  explicit. We can invert this equation as

$$\Phi(\mathbf{x}) = -\frac{k_b T(\mathbf{x})}{m_{\text{DM}}} \ln\left(\frac{\rho(\mathbf{x})}{C_1(\mathbf{x})}\right).\tag{158}$$

This is useful because there is a second equation that relates the gravitational potential and the density, namely the Poisson equation. The Poisson equation reads

$$\nabla^2 \Phi = 4\pi G\rho \rightarrow \frac{1}{r^2} \frac{d}{dr} r^2 \frac{d}{dr} \Phi(r) = 4\pi G\rho(r), \quad (159)$$

where we expressed the  $\nabla^2$  operator in spherical coordinates and, and we assumed spherical symmetry in the density and gravitational potential. The assumption of spherical symmetry translates to  $\Phi(\mathbf{x}) = \Phi(r)$ ,  $T(\mathbf{x}) = T(r)$ , and  $\rho(\mathbf{x}) = \rho(r)$ , where  $r = |\mathbf{x}|$ . We now substitute Eq 158 for  $\Phi(\mathbf{x})$ . We then we obtain a second order differential equation for  $\rho(r)$ , which we can simplify, if we further assume that  $T(r) = T$  (which implies that  $C_1(r) = C_1$ ):

$$\frac{k_b T}{m_{DM} r^2} \frac{d}{dr} r^2 \frac{d}{dr} \ln\left(\frac{\rho}{C_1}\right) = \frac{k_b T}{m_{DM} r^2} \frac{d}{dr} r^2 \frac{d}{dr} \ln \rho = 4\pi G \rho(r), \quad (160)$$

where we used that  $\frac{d}{dr} \ln\left(\frac{\rho}{C_1}\right) = \frac{d}{dr} (\ln \rho - \ln C_1) = \frac{d}{dr} (\ln \rho)$ . We now have a second order differential equation. It is straightforward to show (by substituting it into the differential equation) that the following provides a solution to this equation:

$$\boxed{\rho(r) = \frac{A}{r^2}}, \quad A = \frac{k_b T}{2\pi G m_{DM}}. \quad (161)$$

This density profile that falls off as  $r^{-2}$  is known as an *isothermal density profile*, which is because we assumed a constant  $T$  when solving for  $\rho(r)$ . If we perturb a smooth distribution of dark matter particles, further assume that these dark matter particles have the same velocity distribution (i.e. ‘temperature’) everywhere in space<sup>20</sup>, then this is the profile that gravitational collapse will lead to.

It may help to provide some intuition by comparing the equilibrium solution of an ideal gas in pressure equilibrium with the gravitational potential generated by the same gas. For gas in hydrostatic equilibrium with gravity we have

$$\frac{dP}{dr} = -\frac{GM(< r)\rho}{r^2}. \quad (163)$$

For gas with the same temperature - isothermal gas - we have  $P = nk_b T_0 = \frac{\rho}{m_p} k_b T_0$ . Note that  $M(< r) = 4\pi \int_0^r x^2 dx \rho(x)$ . I will leave it as an exercise to show that this is the same differential equation, and that isothermal gas hydrostatic equilibrium therefore would settle into the same density-profile.

## B. Simulated Dark Matter Density Profiles: Navarro-Frenk-White (NFW) Profiles

Full cosmological N-body simulations (i.e. simulations that evolve the forward in a density field that is initially characterized by the linear power spectrum that we described earlier) show a somewhat different picture: Figure 27 shows a close-up view of a massive simulated dark matter halo (which cover most of the image), in which high density regions are shown as bright, white spots. This figure shows that there is a lot of ‘substructure’ within the massive halos, which can be identified in the simulation as lower-mass dark matter halos, which collapsed at earlier times (and which are therefore denser). This is consistent with our discussion of the evolution of the dark matter halo mass function, in which low-mass dark matter halos form first, and in which increasingly massive halos collapse

<sup>20</sup> We have practically shown already what this distribution is. Recall that  $f(\mathbf{v}) d^3 \mathbf{v} = d^3 \mathbf{v} \int d^3 \mathbf{x} f(\mathbf{x}, \mathbf{v})$ . Under the assumption of thermodynamic equilibrium we found that

$$f(\mathbf{v}) dv \int d^3 \mathbf{x} f(\mathbf{x}, \mathbf{v}) = 4\pi v^2 dv \exp\left(-\frac{m_{DM} v^2}{2k_b T}\right) \int d^3 \mathbf{x} \exp\left(\frac{-m_{DM} \Phi(\mathbf{x})}{k_b T}\right). \quad (162)$$

Thermodynamic equilibrium therefore implies that the velocity distribution of dark matter particles obeys a Maxwell-Boltzmann distribution with temperature  $T$ .

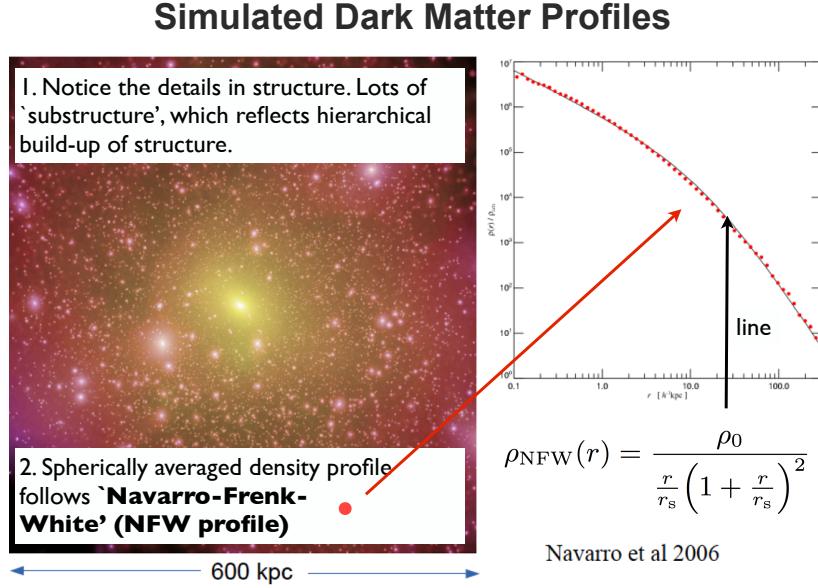


FIG. 27

as time progresses.

In spite of this complexity, dark matter halos all seem to have very similar structures, irrespective of their mass and/or collapse time. If we look at the spherically averaged density profile of simulated dark matter halos, then these have density profiles that have  $\frac{d \log \rho}{d \log r} = -1$  at small  $r$ , and which evolve to  $\frac{d \log \rho}{d \log r} = -3$  at large  $r$ . For comparison, we previously derived analytically that a self-gravitating collection of dark matter particles with the same velocity distribution settles into an ‘isothermal’ density distribution in which  $\rho \propto r^{-2}$ , i.e.  $\frac{d \log \rho}{d \log r} = -2$ . Simulated dark matter halos seem to be universally described by a density profile of the form

$$\rho_{\text{NFW}}(r) = \frac{\rho_0}{\frac{r}{r_s} \left(1 + \frac{r}{r_s}\right)^2}, \quad (164)$$

where ‘NFW’ stands for Navarro, Frenk & White who first discovered this. The ‘NFW-profile’ of dark matter halos is characterized by two parameters:  $\rho_0$ , which sets the overall normalization of the density of the halo (and which is therefore closely related to the collapse time of the halo), and  $r_s$  which corresponds to the ‘core’ radius. The core radius determines where the transition from  $\frac{d \log \rho}{d \log r} = -1$  to  $\frac{d \log \rho}{d \log r} = -3$  occurs. The core radius depends on halo mass, but also on collapse redshift. The NFW profile applies over a remarkable range of scales: it describes the most massive, simulated dark matter halos to form today, but also low-mass dark matter halos that formed early on. In spite of the simple analytic form of the NFW profile, its origin is still debated. It is helpful to explicitly write down the (previously mentioned) limiting behavior of the NFW profile, namely that:

$$\rho_{\text{NFW}}(r) \propto \begin{cases} r^{-1} & r \ll r_s \\ r^{-3} & r \gg r_s \end{cases} \quad (165)$$

Because of its  $r^{-1}$  dependence at small radii, the density of the NFW profile diverges at  $r = 0$ . NFW profiles are

therefore said to be ‘cuspy’<sup>21</sup>. Simulated dark matter profiles are therefore ‘cuspy’. As we will see below, this is in apparent conflict with observational constraints on inferred dark matter density profiles.

### C. Observational Constraints on Dark Matter Density Profiles: Rotation Curves

Measuring rotation curves of galaxies has been a powerful way to constrain their masses, and their density profiles. The main idea is that gravity keeps gas in circular orbits around the center of the gravitational potential. In this case, the rotational velocity  $v_{\text{rot}}(r)$  at some distance  $r$  from the center relates to the gravitational potential  $\Phi(r)$  as

## Measuring Rotation Curves in Galaxies

Example of 2D velocity field of a galaxy M33 inferred from 21-cm observations.

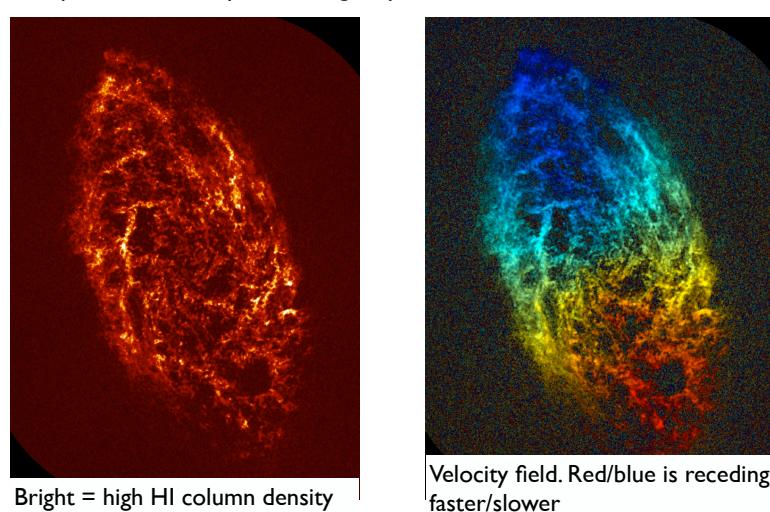


FIG. 28 .

$$v_{\text{rot}}^2(r) = r \left| \frac{\partial \Phi(r)}{\partial r} \right| = \frac{GM(< r)}{r}, \quad M(< r) = 4\pi \int_0^r x^2 dx \rho(x), \quad (166)$$

where  $M(< r)$  is the total enclosed mass within radius  $r$ . Eq 166 shows that if you can measure  $v_{\text{rot}}(t)$ , you directly constraint  $\rho(r)$ . Rotation curves of galaxies can be measured if we can observed a spectral line associated with gas inside galaxies. Gas has finite pressure, which circularizes orbits, and makes the assumption of circular orbits most reliable. The basic idea behind measuring rotation curves of galaxies is that motion of gas through the galaxies imparts an additional<sup>22</sup> Doppler shift on the line. For example, if we see a disk galaxy from its side (‘edge-on’), one side of the galaxy will be redshifted more than its center, which in turn is redshifted more than the other side of the disk (**Draw this if unclear**). Hydrogen is the most abundant element in our Universe, and spectral lines associated with atomic hydrogen provide ideal tracers of gas kinematics. In the lecture I showed one example of an observation which used the Balmer- $\alpha$  (also known as H $\alpha$ )<sup>23</sup> line to observe how the H $\alpha$  redshift varied across the galaxy due to gas motions. A more traditional line to measure rotation curves of galaxies is the 21-cm line<sup>24</sup> The 21-cm transition

<sup>21</sup> From wikipedia: From Latin *cuspis* ?(‘a point, spear, pointed end’); (geometry) A point of a curve where the curve is continuous but has no derivative, but such that it has a derivative at every nearby point.

<sup>22</sup> ‘Additional’ to the cosmological redshift due to the expansion of the Universe.

<sup>23</sup> This corresponds to the transition from the second to first excited state, and which has a wavelength of  $\lambda \sim 6536 \text{ \AA}$ .

<sup>24</sup> This line corresponds to a transition between two hyperfine levels of hydrogen in the ground state. The two hyperfine levels correspond to states of the Hydrogen atom in which the electron and proton spin are anti-parallel and parallel. The 21-cm transition is also referred

falls in the radio and must be therefore be detected with radio telescopes (more specifically, radio interferometers such as Westerbork, the Very Large Array, the Australia Compact Array, to name a few).

## Measuring Rotation Curves in Galaxies

Sample of rotation curves III: showing diversity of galaxy rotation curves.

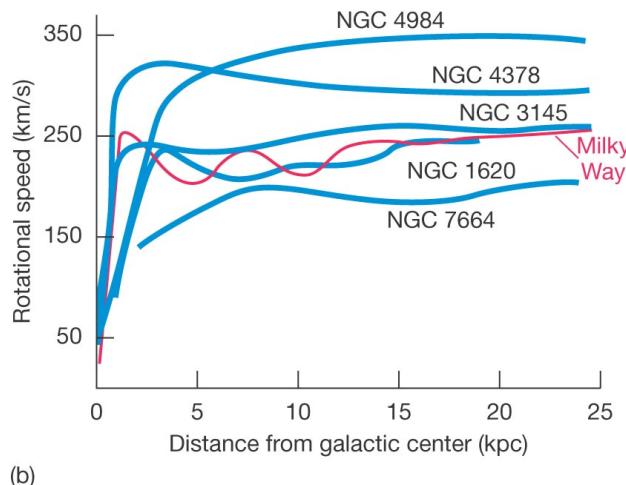


FIG. 29 .

Figure 28 shows an example of a 21-cm observation of a nearby galaxy. The *left image* shows the 21-cm intensity, which can be converted to an HI column density (the projected number density of atoms per unit area). The *right image* shows the 2-D velocity field. Blue (red) indicates that the gas is coming towards us (away from us) relative to the center of the galaxy. The bottom half of the galaxy is moving away from us, while the top half is moving towards us relative to the center of the galaxy, which implies that the galaxy is rotating. To infer a *rotation curve* [i.e.  $v_{\text{rot}}(r)$ ] from the 2-D velocity field is not-trivial, and requires constructing models of a series of concentric rings, each with their own rotation velocity, and their own orientation. In so-called 'tilted-ring' models, one varies rotational velocity and ring orientation for each ring until one reproduces the observed 2D velocity field. In the lecture we showed how good data can constrain, besides the rotation curves, whether galaxy disks are flat like dinner plates, or whether they are 'warped' or not.

Figure 29 shows examples of inferred rotation curves in several nearby galaxies. Most rotation curves show a rapid increase, followed by a range of radii in which it stays relatively constant. It is helpful to point out that

$$\rho(r) \propto r^{-1} \Rightarrow v_{\text{rot}} \propto r^{1/2}, \quad \rho(r) \propto r^{-2} \Rightarrow v_{\text{rot}} = \text{constant}, \quad \rho(r) \propto r^0 \Rightarrow v_{\text{rot}} \propto r. \quad (167)$$

NFW profiles have 'cusps' in their centers, which translates to  $v_{\text{rot}} \propto r^{1/2}$ . For comparison, isothermal density profiles that we discussed previously would have  $v_{\text{rot}} = \text{constant}$ , while 'cored' density profiles - in which the density stays constant out to some radius -  $v_{\text{rot}} \propto r$ . It is also interesting to contrast this with a Keplerian rotation curve, which corresponds to a rotation curve when

$$\rho(r) = 0 \quad r > r_b \Rightarrow v_{\text{rot}} = r^{-1/2} \text{ at } r > r_b. \quad (168)$$

The observed flattening of rotation curves at larger radii seems to imply that the density profile approaches

---

to as the 'spin-flip' transition.

$\rho(r) \propto r^{-2}$ , which does not agree with NFW, but does corresponds nicely to the isothermal density profile that we derived in § XIV. Interestingly, the flattening also extends well beyond the observed extent of the stellar component of the galaxy. That is, if the galaxy would ‘end’ where we stop seeing stars, we would have expected the rotation curve to fall off like  $v_{\text{rot}} \propto r^{-1/2}$ . What we see instead is a roughly constant  $v_{\text{rot}}$ , which provides more support that the galaxy resides in a more extended dark matter halo (whose density profile seems close to isothermal!).

## Inferring Dark Matter Contribution & Distribution

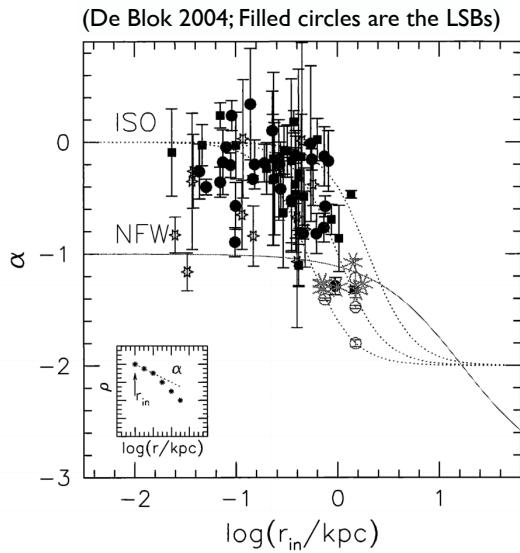


FIG. 30 .

The comparison with NFW can be made more quantitative by comparing the slope  $\alpha \equiv \frac{d \log \rho}{d \log r}$  at small radii as inferred from observations to the slope predicted by simulations. Figure 30 shows  $\alpha$  as a function of radius  $r$  for a suite of gas-dominated dwarf galaxies and so-called ‘low-surface brightness galaxies’. Gas dominated dwarf galaxies are low in stellar content, and it is thought that most of the ordinary baryonic component is in gas (which we can detect through its 21-cm line). Since the 21-cm observations give us direct constraints on the the gas density, it becomes possible to reliably constrain the dark matter density profile from observed rotation curves, and compare these to NFW profiles. Similarly, so called ‘*low surface brightness galaxies*’ (LSB) also have a low stellar content, they appear to reside in calm environments of the Universe (which would give the dark matter the time to settle into an equilibrium distribution). Constraints on  $\alpha$  from LSBs (*filled circles*) also prefer  $\alpha = 0$  over  $\alpha = -1$ .

An NFW profile would have  $\alpha = -1$  for small radii, and  $\alpha \rightarrow -3$  for larger radii. Figure 30 shows that observations indicate that  $\alpha$  is much closer to  $\alpha \sim 0$  at small radii. That is, observations indicate that dark matter density profiles are ‘cored’. This contrast between simulated ‘cusp’ density profiles, and observed ‘cored’ profiles, is known as the ‘cusp-core’ problem. This still represents an outstanding problem in theoretical cosmology. In the lecture we discussed some hypothesised solutions to this problem, one of which invokes that NFW profiles are obtained from dark matter simulations only, while in reality the baryons do interact with the dark matter gravitationally. We discussed one recently proposed idea how ‘feedback’ from star formation (e.g.. via supernova explosions) may drive gas outwards from the center of the gravitational well, which would temporarily make the potential well less deep, which would make it more difficult to gravitationally bind dark matter in a cuspy profile. This represents the first time we invoke ‘feedback’ to solve an apparent problem for our cosmological model!

## XV. STRUCTURE OF DARK MATTER HALOS & MISSING SATELLITE PROBLEM

We previously discussed differences between simulated dark matter density profiles and observational constraint on density profiles near the center of the gravitational potential of galaxies/dark matter halos. This difference is known as the ‘cusp-core’ problem. Now we focus on another problem, the so-called ‘missing satellite problem’ (sometimes abbreviated to MSP’). Figure 27 shows the complex substructure that exists inside simulated dark matter halos, which mostly consists of denser, lower-mass dark matter halos which collapsed at earlier cosmic times, and which reflects the hierarchical build-up of structure in our Universe (low-mass dark matter halos are the first ones to collapse, more massive halos collapse later, ‘absorbing’ a fraction of the older, low mass population).

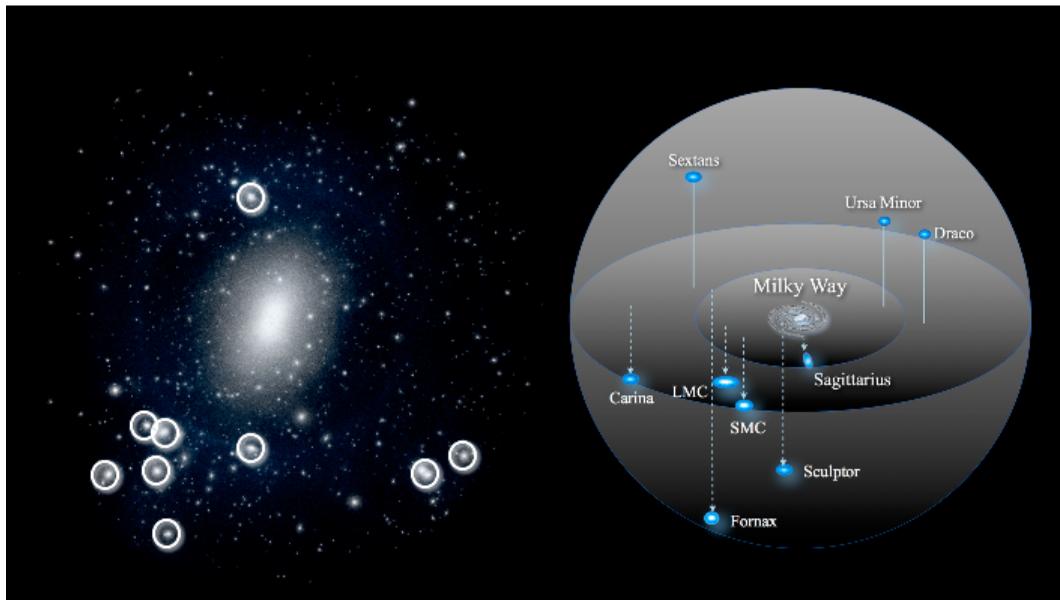


FIG. 31

The *left panel* of Figure 31 again shows a snap shot from a dark matter simulation (brightness again measures dark matter density). Bright dots correspond to dark matter substructures). The *right panel* shows the distribution of satellite galaxies around our Milky Way (MW) galaxy. This figure illustrates visually that the number of known satellites our own MW galaxy is much smaller than the number of the dark matter substructures orbiting our MW halo according to cosmological simulations of dark matter halos with a mass corresponding to that hosting our MW (this discrepancy will be made more quantitative later, in Fig 34). There are a number of ways out of this problem which invoke ‘gastrophysics’, i.e. physical processes which affect the observable baryons but not the dark matter. We will discuss two solutions: (*i*) ‘photoionization’ feedback on the formation of low-mass galaxies (see § XV.A), and (*ii*) ‘supernova feedback’ on galaxy evolution. This discussion immediately sheds some light on the most fundamental challenge to understanding galaxy formation: once galaxies start forming, they affect their environment (and the physical state of their own gas) that makes it very difficult to predict their subsequent evolution from first principles. Indeed, understanding ‘feedback’ mechanisms represents one of the main challenges to our theoretical understanding of galaxy formation and evolution.

### A. Feedback to the Rescue: Photoionization Feedback

In Figure 10 we first introduced the cooling function. We discussed how galaxy formation requires that gas is able to cool and collapse down to the center of the gravitational potential generated predominantly by the dark

matter halo. The cooling curve shows that gas cooling is inefficient at  $T \lesssim 10^4$  K. This implies that galaxy formation is inefficient in galaxies with halo masses with virial temperatures below  $T_{\text{vir}} < 10^4$  K. This statement is correct, but only partially: (i) first Figure 10 does not show the contribution from cooling by molecular hydrogen ( $H_2$ ), which does allow for some cooling and eventually star formation inside halos with  $T_{\text{vir}} < 10^4$  K; (ii) in realistic astrophysical environments, cooling is suppressed as a result of the presence of extragalactic radiation backgrounds (radiation generated by all galaxies combined surrounding a certain collapsing dark matter halo. We will return to radiation backgrounds in § XVIII.B). Specifically, the presence of ionizing radiation removes neutral atoms, and thus suppresses cooling channels such as the 'bound-bound' and 'bound-free' channels. We will show here that photoionization feedback provides quite a natural explanation for the MSP.

### Gas Cooling/Heating Curve in Presence Ionizing Radiation.

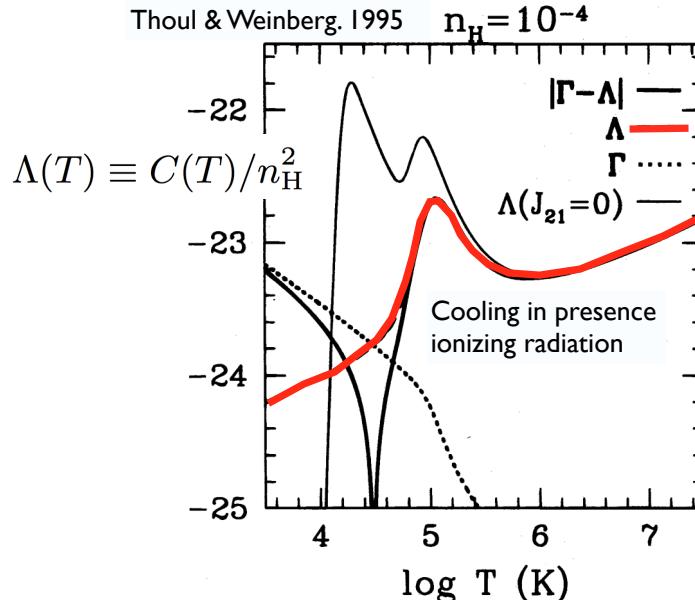


FIG. 32

The impact of an ionizing background on the cooling efficiency of the gas is shown in Fig 32. The *black solid line* now shows cooling function  $\Lambda(T)$  that we showed previously. The *red solid line* shows the reduced cooling efficiency as a result of an ionizing background. The cooling efficiency is reduced by two orders of magnitude at  $T \sim 10^4$  K due to the suppression of neutral hydrogen atoms, where we should stress that quantitatively, the precise reduction does depend on density (which is why we specified the value for  $n_H$  in Fig 32) and intensity of the ionizing background (also see the calculations below). The presence of ionizing radiation does not just affect the cooling efficiency. Photoionization knocks off an electron of H-atoms, and therefore injects thermal energy (i.e. heat) into the gas. This heating rate,

which does *not*<sup>25</sup> on the intensity of the ionizing background, is shown as the *dotted line*<sup>26</sup>. The heating efficiency decreases with  $T$  because collisions reduce the ‘neutrality’ of the medium at higher temperatures. The *thick-solid line* shows the net cooling efficiency (cooling - minus- heating efficiencies). The net cooling efficiency is *negative* for  $\log T \lesssim 4.5$ , i.e. the ionizing background heats the gas, while the net cooling efficiency is positive for  $\log T \gtrsim 4.5$ . *This implies that in the presence of an ionizing background, gas is typically at a temperature of a  $\log T \sim 4.0 - 4.5$ .* As we will see in lectures on the Ly $\alpha$  forest, observations of spectra of distant quasars indeed show that the low-density intergalactic medium is at this temperature. The Jeans mass in gas at  $T = 10^4 K$  at mean density of the Universe is

$$M_{\text{Jeans}} \sim 10^{10} \left( \frac{T}{10^4 \text{ K}} \right)^{3/2} \left( \frac{1+z}{4} \right)^{-3/2} M_{\odot}, \quad (172)$$

which would correspond to a circular velocity of  $v_{\text{circ}} \approx 50([1+z]/4)^{1/2} \text{ km s}^{-1}$  (see Eq 82). This scale corresponds closely to the scale where the missing satellites problem is expected to kick in.

## B. Feedback to the Rescue: (A simple model for) Supernova Feedback

There are other feedback mechanisms operating. Observations indicate that the formation of stars is associated with ‘violent’ feedback. A visually stunning example of this is shown in Figure 33, which shows nearby galaxy M82 that is vigorously forming stars, and that is driving gas out of the disk of the galaxy in two cones perpendicular of the disk. The tips of both cones are centered on the region with the most intense star formation. The outflowing gas contains a complex mixture of hot X-ray emitting gas, and cooler H $\alpha$  emitting gas (*red*). Practically all star forming galaxies show evidence of outflowing gas in their spectra, so this phenomenon appears ubiquitous (though not likely not as spectacular as in M82).

<sup>25</sup> The heating rate per unit volume is

$$H = n_{\text{HI}} \mathcal{H} = n_{\text{HI}} \Gamma \langle E \rangle \quad (169)$$

, where  $\mathcal{H}$  is the heating rate per neutral hydrogen atom. Furthermore,  $\Gamma$  denotes the photoionisation rate per neutral hydrogen atom, which is the rate at which a hydrogen atom absorbs photons. The term  $\langle E \rangle$  denotes the average energy that is deposited into the gas per photoionisation event. The heating rate & photoionisation rate *per neutral hydrogen atom* increase in proportion to the intensity in the ionising background. The higher the intensity of the radiation field, the higher the rate at which an individual atom absorbs this radiation. To get the heating rate per unit volume we need to compute  $n_{\text{HI}}$  which depends on the ionising background intensity. The *total* number density of hydrogen nuclei (i.e. in neutral and ionised state) is given by  $n_{\text{H}} = n_{\text{HI}} + n_{\text{HII}} = n_{\text{HI}} + n_{\text{P}}$ . We further introduce the neutral fraction  $x_{\text{HI}} \equiv n_{\text{HI}}/n_{\text{H}}$ . The neutral fraction  $x_{\text{HI}}$  is set by an equilibrium between photoionisation and recombination. As mentioned above, the total **photoionisation rate per unit volume** is  $n_{\text{HI}} \Gamma$ . The photoionisation rate *per atom* is again proportional to the intensity in the ionising background. The total **recombination rate per unit volume** is equal to  $n_e n_p R(T)$ , where  $R(T)$  denotes the recombination coefficient. This coefficient  $R(T)$  is a quantity that can be obtained from quantum physics. If we balance photoionisation and recombination

$$n_e n_p R(T) = n_{\text{HI}} \Gamma \Rightarrow (1 - x_{\text{HI}})^2 n_{\text{H}}^2 R(T) = x_{\text{HI}} n_{\text{H}} \Gamma \Rightarrow (1 - x_{\text{HI}})^2 n_{\text{H}} R(T) = x_{\text{HI}} \Gamma$$

. This is a quadratic equation for  $x_{\text{HI}}$  which we can solve for. Under the assumption that the medium is mostly ionised ( $x_{\text{HI}} \ll 1$ ) - which is a reasonable approximation in typical conditions in the intergalactic medium - we can approximate  $(1 - x_{\text{HI}})^2 \sim 1$ , and

$$x_{\text{HI}} = \frac{n_{\text{H}} R(T)}{\Gamma}. \quad (170)$$

This expression states that the neutral fraction increases linearly with density (as recombination becomes more efficient), and inversely with the ionising background intensity. Thus we can **finally** write for the volume heating rate

$$H = x_{\text{HI}} n_{\text{H}} \Gamma \langle E \rangle = n_{\text{H}}^2 \langle E \rangle R(T). \quad (171)$$

That is, the heating rate per unit volume does not depend on the intensity of the ionising background!

<sup>26</sup> The *dotted line* shows  $H/n_{\text{H}}^2$ , in which  $H$  denotes the cooling rate per unit volume (see footnote 24). This is similar to the fact that  $C(T)$  denotes the cooling rate per unit volume, and because we choose to plot cooling curve  $\Lambda(T)/n_{\text{H}}^2$ . We can call  $H/n_{\text{H}}^2$  the ‘heating curve’. Using Eq 171 it becomes clear that the heating curve  $H/n_{\text{H}}^2 = \langle E \rangle R(T)$ , and the  $T$ -dependence of the cooling curve reflect the  $T$ -dependence of the recombination coefficient  $T$ .

## Feedback Mechanism II: Stellar Feedback.



**M82**, nearby dwarf galaxy with a recent star burst. Red indicates H-alpha emission from outflowing gas. Outflow driven by supernova explosions, and/or radiation pressure.

FIG. 33

What is driving these outflows is not well understood. A variety of physical explanations include: ‘kinetic’ feedback by supernova ejecta (i.e. a collection of supernova explosions can simultaneously sweep up gas), radiation pressure off dust grains, and/or ‘cosmic ray’ pressure (relativistic particles). Understanding the details of ‘stellar feedback’ lies at the core of understanding galaxy formation. We will present a simple argument here why feedback preferentially removes gas from low-mass dark matter halos (and thus make these halos appear ‘dark’). The idea is that a dark matter halo of mass  $M$  has a total baryonic content  $M_b = M \frac{\Omega_b}{\Omega_m}$  (where  $\Omega_m$  denotes the ordinary density parameter of all matter, while  $\Omega_b$  denotes the density parameter for ‘just’ baryons.). Suppose that a fraction  $f_*$  of all these baryons is turned into stars, while fraction  $(1 - f_*)$  remains in gaseous phase. The total gravitational binding energy of the baryonic gas is

$$U_b = -\frac{GMM_b(1 - f_*)}{R} = -\frac{\Omega_b}{\Omega_m} \frac{GM^2(1 - f_*)}{R}, \quad (173)$$

where we  $R$  denotes the virial radius (this radius is not really correct. In reality the gas will have assembled deeper in the potential well at a smaller radius. For simplicity, we ignore this).

The (maximum) total energy that can be deposited into this gas by supernova is equal to

$$E_{\text{sn}} = \epsilon_{\text{sn}} \eta M_* = \epsilon_{\text{sn}} \eta f_* \frac{\Omega_b}{\Omega_m} M, \quad (174)$$

where  $\epsilon_{\text{sn}}$  denotes the total energy released per supernova explosion,  $\eta = 0.01$  is the number of supernovae per unit solar mass of gas that forms stars. The total energy deposited by supernovae equals the binding energy of the remaining gas when

$$\frac{\Omega_b}{\Omega_m} \frac{GM^2(1 - f_*)}{R} = \epsilon_{\text{sn}} \eta f_* \frac{\Omega_b}{\Omega_m} M \Rightarrow \frac{f_*}{1 - f_*} = \frac{GM}{R\epsilon_{\text{sn}}\eta} = \frac{v_{\text{circ}}^2}{\epsilon_{\text{sn}}\eta}. \quad (175)$$

## Missing Satellite Problem in Models that Include Feedback

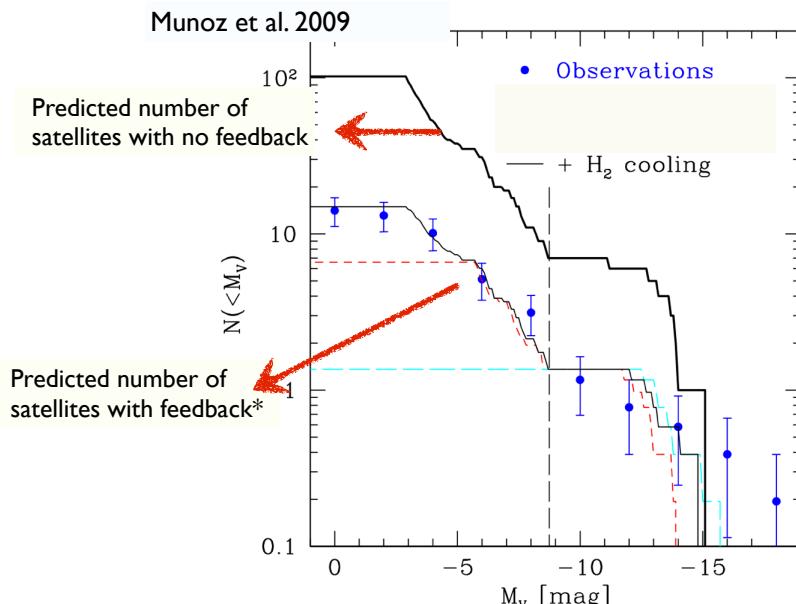


FIG. 34

\* Note: feedback tuned to reproduce observations.

For  $f_* \ll 1$ , we therefore have  $f_* \propto v_{\text{circ}}^2 \propto M^{2/3}$ . This means that the fraction of baryons that is needed to unbind/blow out all gas from the dark matter potential halo increases with mass. In other words, it is easier to evacuate all gas from the lower mass halos with supernova explosions.

Figure 34 shows how feedback can solve the MSP: the figure shows the observed (*blue data points*) and predicted number of satellites around the MW as a function of their absolute magnitude (luminosity increases from left-to-right) in a cosmological hydrodynamical simulation. The *upper line* shows the predicted number of satellites with feedback switched off, while the *lower line* shows the predictions if we switch on photoionization, supernova (and also suppress  $H_2$  cooling, which we have not discussed, but which works similar as photoionization feedback) feedback. The main point of this plot is that by invoking feedback, we can solve the MSP. The main issue is that since we do not really understand what is driving - say - stellar feedback, we have some freedom in tuning the feedback such that the MSP goes away. As we will discuss next,

### C. Has the MSP morphed into the ‘Too Big to Fail’ Problem?

As we discussed in the lectures, it is observationally difficult to find faint satellites. It has only recently been possible to take good quality spectra of some of the MW satellites. With these observations, it became possible to measure maximum rotation velocities of the satellites,  $v_{\text{max}}$ . This maximum rotation velocity,  $v_{\text{max}}$ , is shown as a function of satellite luminosity (phrased in terms of their absolute V-band magnitude,  $M_V$ ) in the *right panel* of Figure 35 as *black points*.

For comparison, the *purple filled circles* show maximum rotation velocities of the most massive dark matter satellites in a simulation in which feedback resolved the MSP (see the *left panel* which shows that the number of observed and satellites around the MW are in good agreement, just as in Fig 34). Fig 35 illustrate that models that resolve the MSP seem to get the physical properties of the satellites wrong. Perhaps most troubling is that the simulated dark matter halos have maximum rotation velocities, which provide a close measure of their circular velocities, that are larger than observed. In other words, the simulated satellites appear more massive. A caveat is that we are comparing the most massive simulated satellite dark matter halos with the most luminous observed satellites. This comparison is only truly ‘fair’ if the most luminous satellites reside in the most massive halos. This is not necessarily the case, especially given that in reality, a halo of a fixed dark matter halo mass should have galaxies with a wide range of luminosities associated with them. Note however, in order to make the observations and simulations agree, we would want to preferentially suppress star formation in halos with  $v_{\text{circ}} \gtrsim 25 \text{ km s}^{-1}$ , while not suppressing it in halos with

## `Too Big To Fail' Problem.

Boylan-Kolchin et al. 2011/2012

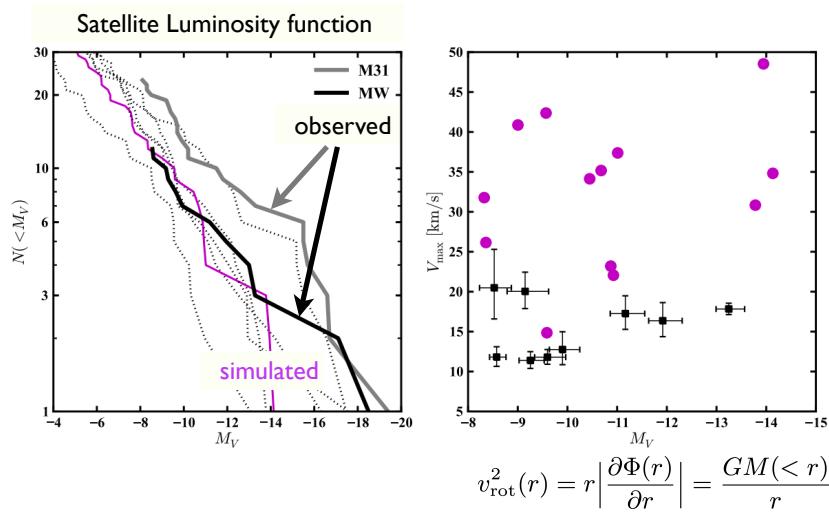


FIG. 35

$v_{\text{circ}}$  less than that. Even if we can explain the satellites with the less massive satellites halos with  $v_{\text{circ}} \sim 10 - 20 \text{ km s}^{-1}$ , it becomes challenging to explain why star formation was so much less efficient inside the more massive satellites with conventional feedback. These more massive satellites are more stable against feedback, hence ‘too-big to fail’.

In the lectures, we discussed how the too-big-to-fail problem may be related to the cusp-core problem. The main idea is that observed simulated dark matter halos are ‘cuspy’ while observations indicate that density profiles are ‘cored’. Observed rotation curves may therefore rise more slowly, and peak further out. At fixed radius, simulated and observed rotation curves would therefore be different in the sense that at a fixed radius the simulated rotation velocity would be higher. This might explain the discrepancies between observed and simulated  $v_{\text{max}}$ .

## XVI. IS THE SMALL-SCALE CRISIS IN CDM RELATED TO THE NATURE OF DARK MATTER?

The combination of the ‘cusp-core problem’ and the MSP (or the TBTF) problem is known as the ‘small scale crisis’ in cold dark matter cosmologies. Naturally, people have looked at looking at this problem at the dark matter level.

### A. Cold vs Warm Dark Matter

The kinetic energy of a particle is  $K \sim k_B T$  (in thermodynamic equilibrium). As the Universe cools adiabatically, the temperature of the Universe decreases. A particle of mass  $m$  is said to become non-relativistic when

$$k_B T \sim mc^2, \quad (176)$$

where  $m$  denotes the mass of the particle,  $c$  the speed of light, and  $k_B$  denotes the Boltzmann constant. In lecture 1 we discussed that for relativistic particles we have that  $T \propto a^{-1}$ , while for non-relativistic particles we have  $T \propto a^{-2}$  (see Eq 9 and discussion below that). When a particle becomes non-relativistic, it suddenly cools much faster than the relativistic particles. In all our previous discussions, we implicitly assumed that dark matter was ‘cold’. This means that it was composed of massive particles. The main way this implicit assumption entered our analysis was when we assumed that dark matter density fluctuations were able to grow under the influence of gravity. This is only the case when particles are ‘non-relativistic’.

Relativistic particles propagate at the speed of light, and cannot be confined to gravitational potential wells. Therefore the ‘free-streaming’ of relativistic particles washes out any structure on scales smaller than the particle horizon of the Universe when they first become non-relativistic (this is the maximum distance these particles

can travel). Denote this time with  $t_{\text{nr}}$ , where the subscript ‘nr’ stands for ‘non-relativistic’. The scale factor of the Universe at time  $t_{\text{nr}}$  corresponds to  $a_{\text{nr}}$ . If  $a_{\text{nr}} \ll a_{\text{eq}}$  (where  $a_{\text{eq}}$  denotes the scale factor of the Universe at matter-radiation equality), then the particle horizon  $r_H = 2ct$  (also see § XII.A).

Recall that  $r_H = 2ct$  is a proper (or physical) quantity. We would like to compare this scale to the horizon size at matter-radiation equality (we denoted this with  $L_0$  in § XII.B), because it leaves such a prominent feature in the matter power spectrum. We know from § XII.A that the *comoving* particle horizon scales as  $\propto \int dt/a(t) \propto \int dt t^{-1/2} \propto t^{1/2} \propto a$ . Let us denote the *comoving* particle horizon at  $a_{\text{nr}}$  with  $r_{H,\text{nr}}^c$ , and the *comoving* particle horizon at  $a_{\text{eq}}$  with  $L_0^c$ . We then have

$$\frac{r_{H,\text{nr}}^c}{L_0^c} = \frac{a_{\text{nr}}}{a_{\text{eq}}} = \frac{T_{\text{eq}}}{T_{\text{nr}}} = \frac{k_B T_{\text{eq}}}{k_B T_{\text{nr}}}, \quad (177)$$

where in the second last equality we used that  $T \propto a^{-1}$  (recall that matter and radiation are coupled tightly prior to when the Universe recombined, and that the radiation temperature always decreases as  $T \propto a^{-1}$ ).

Now, matter-radiation equality occurs prior to recombination at a time when  $k_B T \sim 1 \text{ eV} \equiv k_B T_{\text{eq}}$  (**Verify this yourself!**). Therefore, the comoving particle horizon size at  $a_{\text{nr}}$  equals

$$r_H(a_{\text{nr}}) \sim r_H(a_{\text{eq}}) \frac{1 \text{ eV}}{k_B T_{\text{nr}}} \sim 10^{-3} r_H(a_{\text{eq}}) \frac{\text{keV}}{k_B T_{\text{nr}}} \sim 10^{-3} r_H(a_{\text{eq}}) \frac{\text{keV}}{m_X c^2}, \quad (178)$$

where in the last equality I used Eq 176.

For example, for a particle mass  $m_X c^2 = 1 \text{ keV}$ , the horizon size when the particle becomes non-relativistic is  $\sim 10^{-3} r_H(a_{\text{eq}})$ . Structures smaller than this scale would have been washed out completely by the free streaming of the particles. We therefore expect structure formation to be suppressed at scales  $r < 10^{-3} r_H(a_{\text{eq}})$ . Another way of phrasing this is that the structure is washed out at wave numbers larger than  $k > k_{\text{fs}} = 10^3 (m_X c^2 / \text{keV}) k_0$ , where  $k_0 \sim 0.01 \text{ cMpc}$ , which corresponds to the wavenumber  $k_0$  at which the power spectrum turns over due to the Meszaros effect. We therefore expect dark matter particles with mass  $m_X$  to suppress the power spectrum at wave numbers larger than

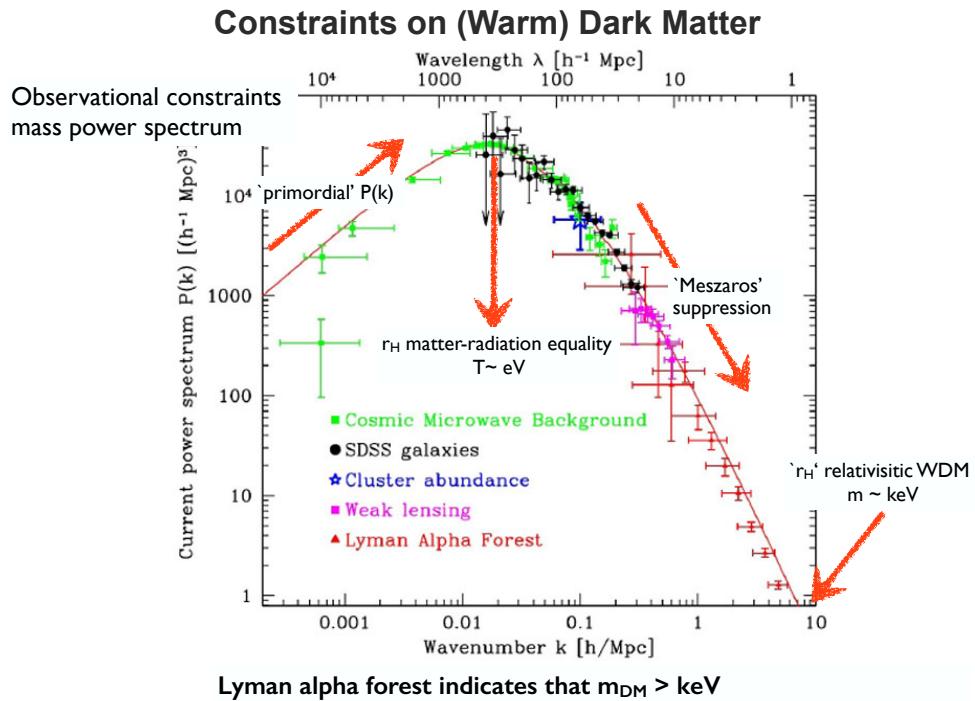


FIG. 36

$$k_{\text{fs}} \sim 10(m_X/\text{keV}) \text{ cMpc}^{-1}. \quad (179)$$

Figure 36 shows that at these wave numbers the mass power spectrum is best constrained by Ly $\alpha$  forest data. We will discuss the Ly $\alpha$  forest in § XVII. Figure 36 shows that there is no evidence for a suppression in the power-spectrum - additional to that expected from the Meszaros effect - down to the smallest scales that can be probed with Ly $\alpha$  forest observations. This provides as with a *lower limit* on the dark matter mass at of  $m_X \gtrsim 2.4$  keV. For smaller masses, the turn-over would have occurred at smaller  $k$ .

What is interesting about WDM (warm dark matter) cosmologies is that the suppression of power on small scales, and consequently suppresses the number of low-mass dark matter halos, and naturally addresses the MSP.

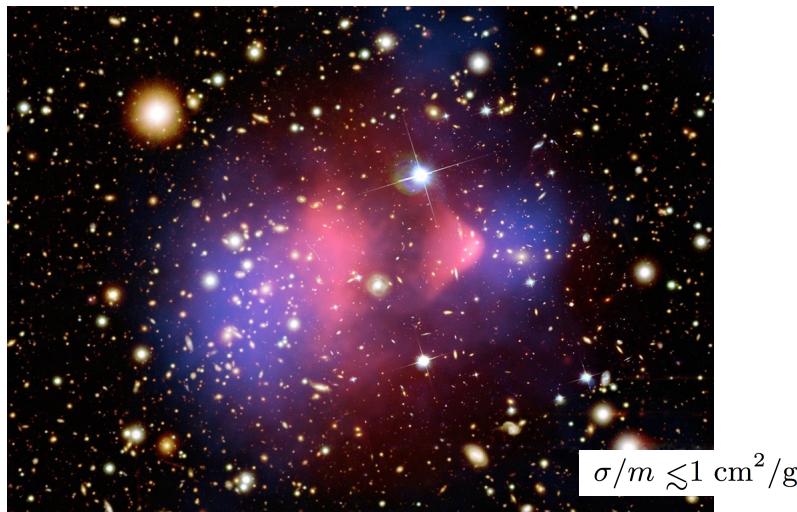
## B. Constraints of Other Properties of Dark Matter

In the lectures, we discussed other observational constraints on dark matter following the pedagogical review by A. Peter in arXiv:1201.3942 (Peter, 2012). I refer you to this review for a more extended analysis.

- If dark matter particles were charged, then they would couple to both photons and dark matter. This affects the small-scale structure in the CMB in ways that can be computed.

## Constraints on other Properties of Dark Matter

“Bullet Cluster”: two merging clusters.



Pink: hot X-ray emitting gas. Blue: dark matter in the cluster, determined from measuring the lensing signal (lecture~20) from the visible-light images of the galaxies.

FIG. 37

- We have always assumed dark matter to be collisionless (and therefore have zero pressure). As we discussed in the lecture notes, direct constraints on how ‘collisionless’ dark matter really is come from the so-called ‘Bullet cluster’, which consists of a merger of two galaxy clusters (see Fig 37). The collisional hot baryonic fluid is traced by X-ray emission (shown in red), while the distribution of the dark matter can be constrained via gravitational lensing (the mass distribution is shown in blue). Comparing the distribution of the dark matter to that of the X-ray emitting gas, it becomes clear that the dark matter is displaced w.r.t to the baryons: the collisional nature of the baryons ‘stops’ the gas when the two galaxy clusters first collide. The dark matter halos on the other hand, can stream through each other (recall the video I showed when we were discussing virialization: the dark matter particles often shot through the center of the gravitational potential well without physically colliding

with any other dark matter particle). Quantitatively, the observed off-set between the baryons and the dark matter constrains the cross-section for *self-interaction* between dark matter particles to be  $\sigma/m_{\text{dm}} \lesssim 1\text{cm}^2/\text{g}$ . The reason we get constraints in these units is that observations constrain the collision rate  $\Gamma_{\text{coll}} (\text{s}^{-1})$ , which can be related to the collision cross-section as  $\Gamma_{\text{coll}} = n_X \sigma \langle v \rangle$ . Here,  $n_X = \rho_X/m_{\text{dm}}$  denotes the number density of DM particles (where  $\rho_X$  denotes the DM density) in  $\langle v \rangle$  denotes the average (or r.m.s) DM particle velocity. Since we know  $\rho_X$  and  $\langle v \rangle \sim v_{\text{circ}}$ , constraints on  $\Gamma_{\text{coll}}$  directly constrain  $\sigma/m_{\text{dm}}$ . Recent studies have simulated dark matter density profiles with self-interacting dark matter, which satisfies this constraint, and showed that the cusp-core problem & MSP in so-called ‘SIDM’-models is less severe than for CDM ([Vogelsberger et al., 2012](#), though not eliminated completely).

### C. Candidate Dark Matter Particles

- **WIMPs:** It is thought that new particles should exist around a mass-scale around 200 GeV, which correspond to the energy scale at which the weak and electromagnetic interaction merges into the ‘electroweak interaction’. In the very early Universe, kinetic energy of individual particles in the primordial soup was large enough that WIMPs were continuously created and annihilated into quarks + antiquarks in the early Universe. When density and temperature dropped sufficiently, creation and annihilation became increasingly rare, until eventually the comoving number density of WIMPs  $n_X$  ‘froze’ out. The predicted mass density in this relic density of particles - for the standard assumptions for the mass and annihilation coupling strength - comes such that

$$\Omega_X \sim 0.1. \quad (180)$$

The fact that particle physics considerations alone, can give the correct order of magnitude for the dark matter density is referred to as ‘WIMP Miracle’, which is why WIMPs have been among the most popular candidates for dark matter.

- **Axions.** A hypothetical particle that was introduced to resolve the strong CP problem in quantum chromodynamics.
- **Gravitinos.** The supersymmetric partner of the graviton. Not as popular as WIMPs because hard to detect, and tuning is required to get matter density correct.
- **Sterile Neutrinos:** these are neutrinos that do not act electro-weakly. Introduced to generate mass for ‘active neutrinos, explain neutrino experiment anomalies,...
- **Macro dark matter.** General class of dark matter candidates which contains particles which are much more massive, spanning the range  $10^{-12} - 10^{34}\text{g}$ ! An example we discussed in the lectures are ‘nuclearites’, which are intermediate between the most massive nuclei ( $A \sim 260$ ) and full neutron stars ( $A \sim 10^{57}$ ). Interestingly, it is observationally difficult to rule out a large portion of the parameters space.

Fun thing to think about: the dark sector does not necessarily have to consist of one particle. It could be as rich in phenomenology(or even richer) as the the standard model. This phenomenology would be largely hidden from us, and as a result this is sometimes referred to as ‘hidden sectors’ in the dark matter, which may contain for example something like ‘dark photons (to detect these ‘dark photons’ you would need ’dark telescopes’).

### D. Searches & Claimed Detections for Dark Matter

In the lectures we discussed several searches and claimed detections of dark matter. These include

- **Colliders.** Because dark matter particles are mostly neutral and interact only weakly with particles in the standard model, they are difficult to pick-up in collider experiments. Missing energy in high-energy collisions can be accounted for by new particles. So far, no evidence for new particles beyond the standard model have been found at the energies where WIMPs could have existed. This should demote the status of the WIMP as the most popular DM candidate.
- **Direct detection via nuclear recoil.** A (rare) collision between a dark matter particle (e.g. a WIMP) and an atomic nucleus, can leave the nucleus in an excited state, which can then transition back to the ground state by emitting photons. Experiments include DAMA/LIBRA, CRESST, CoGeNT, XENON100, CDMS-II, COUPP. The XENON experiment uses liquid xenon, because this is so dense that normal neutrons cannot enter. Moreover, the liquid is transparent to photons emitted by excited nuclei, which makes it easier to detect their radiation.

The DAMA/LIBRA experiment claims a detection (the annular modulation of their detection may be due to the relative velocity of the Earth through the halo of dark matter halo), though this detection is inconsistent with other experiments (such as e.g. XENON100).

- We discussed a few other possible detections, though none are convincing. These include (*i*) the Fermi ‘haze’ centered on the center of our MW galaxy, (*ii*) the unknown 3.5 keV line detected in clusters of galaxies. This line is real, but is likely associated with known elements.

## XVII. LY $\alpha$ FOREST

We have encountered the Ly $\alpha$  forest previously when discussing observational constraints on the matter power spectrum  $P(k)$  (see Fig 26), and on warm dark matter (see § XVI.A). We will now discuss why the Ly $\alpha$  forest is such a powerful probe of cosmology. First, we need to describe what it actually is.

### The Lyman Alpha Forest

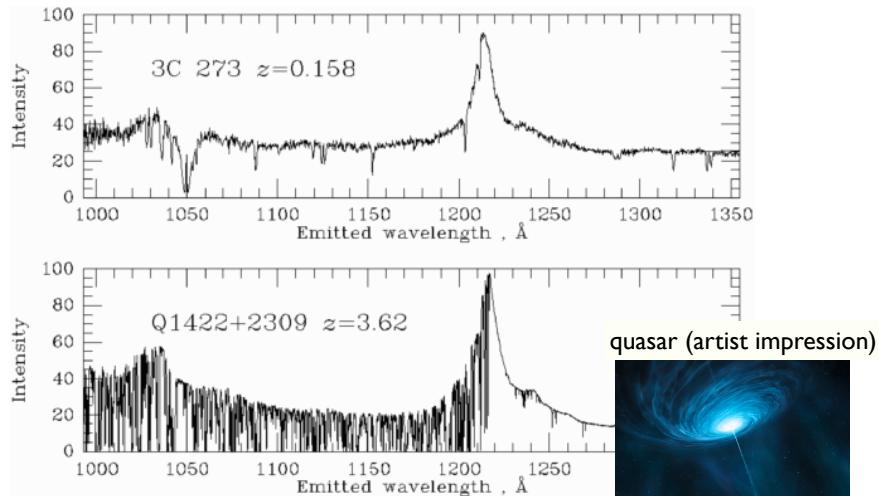


FIG. 38

Figure 38 shows two spectra of distant quasars. Quasars are ‘quasi-stellar’ radio sources, are very luminous, compact sources (they are compact sources, and therefore resemble stars in images), which are powered by accretion of gas onto a supermassive black hole. We will talk about this more in later lectures. The *top* [*lower*] panel shows the spectrum of a quasar at  $z = 0.158$  [ $z = 3.62$ ]. Both spectra are shifted back into *their rest-frame*, to be able to properly compare them. The rest-frame spectra differ from the observed spectra by a factor of  $1 + z$ . Both spectra show a broad Ly $\alpha$  emission line at  $\lambda_{\alpha} \sim 1216$  Å. The main difference between the two spectra is that for  $\lambda < \lambda_{\alpha}$  the spectrum of the high- $z$  quasar looks much more noisy. However, the ‘noise’ that we see in the spectrum of the high- $z$  quasar at  $\lambda < \lambda_{\alpha}$  is not noise at all! The rapid fluctuation in the spectrum is real, and due to absorption by HI gas in between us and the quasar. The collection of absorption lines that we see at  $\lambda < \lambda_{\alpha}$  is what is known as the Ly $\alpha$  forest, and consists of (mostly) Ly $\alpha$  absorption by intervening hydrogen. Observations indicate clearly that as we go to higher redshift, the number of absorption lines in the spectrum increases dramatically. The goal of this lecture is to quantitatively understand the origin of this absorption, and what it teaches us about the physical state of the intergalactic medium, and even the distribution of dark matter through the Universe. We will calculate the expected opacity of the Universe to photons emitted at  $\lambda < \lambda_{\alpha}$ , and will compare to observations.

#### A. The Mean Number Density of Hydrogen Nuclei/Atoms in the Universe

It is helpful to start by deriving the (average) number density of hydrogen atoms/nuclei in our Universe. We know that the expansion of the Universe dilutes the mean density of baryons,  $\bar{\rho}_b$ , as  $\bar{\rho}_{b,0}a^{-3} = \bar{\rho}_{b,0}(1+z)^3$ . We also know

that the mean baryon density is given by

$$\Omega_b \equiv \frac{\bar{\rho}_b}{\rho_{\text{crit}}} \Rightarrow \bar{\rho}_b = \Omega_b \rho_{\text{crit}} = \Omega_b h^2 \times 1.88 \times 10^{-29} \text{ g cm}^{-3}, \quad (181)$$

where I used that  $\rho_{\text{crit}} = 1.88 \times 10^{-29} h^2 \text{ g cm}^{-3}$ . The mean number density of hydrogen nuclei as a function of  $z$  is therefore

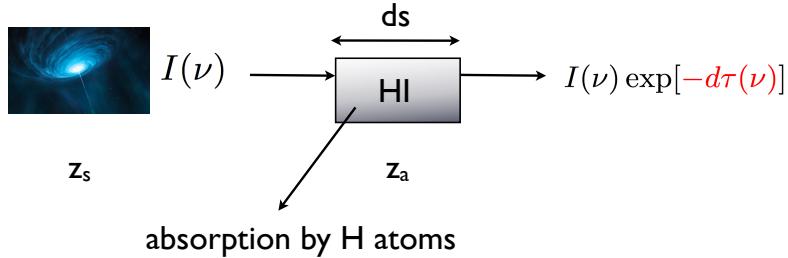
$$\bar{n}_H = f_H \frac{\bar{\rho}_b}{m_p} = (1 - Y_{He}) \frac{\bar{\rho}_{b,0}(1+z)^3}{m_p} = (1 - Y_{He}) \rho_{\text{crit}} \Omega_b (1+z)^3 / m_p, \quad (182)$$

where  $f_H$  denotes the mass-fraction of baryons that is hydrogen. Here,  $f_H = 1 - Y_{He} = 0.76$ , in which  $Y_{He} = 0.24$  denotes the max fraction in helium. Note that  $f_H + Y_{He} = 1.0$ , i.e. all baryons are either hydrogen or Helium. Substituting numbers we get

$$\boxed{\bar{n}_H \approx 2.0 \times 10^{-7} (1+z)^3 \text{ cm}^{-3}}. \quad (183)$$

We will use this number in our calculation of the opacity of the Universe.

## B. The Ly $\alpha$ Forest



I have adopted the notation used in AST4310. The radiative transfer equation reads (page 43 Rutten)

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu, \quad (184)$$

where - in the case of Ly $\alpha$  -  $j_\nu$  denotes the Ly $\alpha$  line emission coefficient, and  $\alpha_\nu$  denotes the line extinction coefficient (see Chapter 5.4 in Rutten), and is given by

$$\alpha_\nu = \frac{h\nu_0}{4\pi} n_l B_{lu} \varphi(\nu - \nu_0) = \frac{\pi e^2}{m_e c} n_l f_{lu} \varphi(\nu - \nu_0), \quad (185)$$

where  $\nu_0 = 2.46 \times 10^{15} \text{ Hz}$  denotes the frequency of the Ly $\alpha$  transition,  $B_{lu}$  denotes the Einstein-B coefficient of the Ly $\alpha$  transition,  $\varphi(\nu)$  denotes the extinction profile function (which is normalised to  $\int d\nu \varphi(\nu) = 1$ ),  $n_l$  is the number density of atoms in the lower state of the Ly $\alpha$  connection, i.e. the ground state. By far the majority of hydrogen atoms are in their ground state, and we can write  $n_l = n_{\text{HI}}$ , where  $n_{\text{HI}}$  denotes the number density of hydrogen atoms. In the equation on the right hand side,  $f_{lu} = 0.416$  denotes the oscillator strength,  $m_e$  denotes the electron mass,  $e$  denotes the electron charge,  $c$  denotes the speed of light. We adopt a shorter notation for  $\alpha_\nu$  and write

$$\alpha_\nu = C_1 n_{\text{HI}} \varphi(\nu - \nu_0), \quad (186)$$

where we defined  $C_1 \equiv \frac{\pi e^2}{m_e c} f_{lu}$  (units  $\text{cm}^2 \text{ Hz}$ ).

We are interested in how the intergalactic medium affects the spectrum that we observe from distant sources. In practise, we can ignore the term  $j_\nu$ . That means that in practise the intergalactic medium only removes photons from the spectra of distant sources and does not add any (another way of seeing this is that emission from the intergalactic medium is extremely faint, much fainter than the galaxies we are detecting). The radiative transfer equation then simplifies to

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu, \quad (187)$$

which can be solved to give

$$I_\nu(s) = I_\nu(s=0) e^{-\alpha_\nu s}, \quad (188)$$

if  $\alpha_\nu$  is a constant. If  $\alpha_\nu$  is not a constant, then

$$I_\nu(s) = I_\nu(s=0) e^{-\tau_\nu}, \quad \tau_\nu \equiv \int_0^\infty \alpha_{\nu,s} ds = \int_{\text{source}}^{\text{us}} \alpha_{\nu,s} ds \quad (189)$$

Note that we have to integrate  $s$  all the way from the distant source/galaxy that we are observing to us observers. We have made that more explicit in the final step. We would like to simplify this expression. We can do that by rewriting the integral over  $ds$  as an integral over  $d\nu$ . We know that as photons propagate a differential distance  $ds$  the two ends of that differential length expand away from each with speed  $dv = H(z)ds$  where  $H(z)$  is the Hubble parameter at redshift  $z$ . We can then use that

$$\frac{d\nu}{\nu} = -\frac{dv}{c} = -\frac{H(z_s)ds}{c} \Rightarrow ds = -\frac{cd\nu}{H(z_s)\nu}, \quad (190)$$

xxx If we substitute this into the integral for  $\tau_\nu$  we get

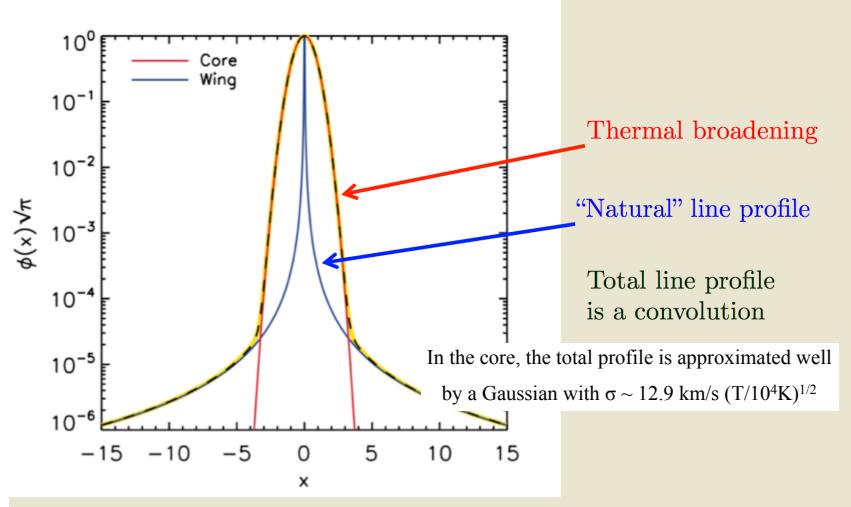
$$\tau_{\text{IGM}}(\nu_e) = - \int_{\nu_e}^{\nu_{\text{obs}}} \frac{cd\nu}{H(s)\nu} C_1 n_{\text{HI}}(s) \varphi(\nu[s]) = \int_{\nu_{\text{obs}}}^{\nu_e} \frac{cd\nu}{H(s)\nu} C_1 n_{\text{HI}}(s) \varphi(\nu[s]), \quad (191)$$

where now we have to integrate over frequency from the source (where the radiation has frequency  $\nu_e$ ) to us (where the radiation has frequency  $\nu_{\text{obs}}$ ). If have added the subscript IGM to  $\tau$  to emphasis that we are looking at the  $\tau$  of the IGM. Note that as photons propagate along  $s$  the Universe continues to expand, so each value of  $s$  corresponds to a redshift  $z$ , with its own Hubble parameter  $H(z)$  and number density  $n_{\text{HI}}(z)$ . In the last step we used that  $\int_a^b = -\int_b^a$ .

Suppose we look at a distant galaxy at some redshift  $z_e$  which emitted radiation with frequency  $\nu_e$ . There are a few interesting regimes for  $\nu_e$  to consider:

- If  $\nu_e < \nu_0$  (i.e.  $\lambda > \lambda_\alpha$ ), then the radiation was initially less energetic than Ly $\alpha$ . The Hubble expansion will only redshift it to lower frequency. Because  $\varphi(\nu)$  is so sharply peaked at  $\nu_0$ ,  $\varphi(\nu \ll \nu_0) \approx 0$ , and the IGM is transparent. This is consistent with observations: the intergalactic medium only affects  $\lambda < \lambda_\alpha$  (see Fig 38).
- If  $\nu_e > \nu_0$ , then the radiation was initially more energetic than Ly $\alpha$ . Now the radiation will redshift until its frequency coincides exactly with  $\nu_0$ . At this moment the probability that the photon is scattered is maximised. We discuss this more below

Because  $\varphi(\nu)$  is extremely narrow (see Fig XVII.B), with a width  $\Delta\nu/\nu_0 \sim 10^{-4}$ , we can simplify the integral. Because we effectively get only non-zero contribution to the integral when  $\nu \approx \nu_0$  (to within  $\Delta\nu/\nu_0 \sim 10^{-4}$ ) we get non-zero contribution to the integral over a very narrow range of redshifts,  $\Delta z \sim \Delta\nu/\nu_0 \sim 10^{-4}$ . Over such a small redshift interval we can safely assume that  $n_{\text{HI}}$  and  $H$  are a constant, with values  $n_{\text{HI}} = n_{\text{HI}}(z_a)$  and  $H = H(z_a)$ ,



47

where  $z_a$  corresponds to the redshift at which photons emitted at  $\nu_e > \nu_0$  redshift into resonance. Since  $\nu \propto (1+z)$  we have

$$1 + z_a = (1 + z_e) \frac{\nu_0}{\nu_e}. \quad (192)$$

So once we know  $z_e$  and  $\nu_e$  we immediately know  $z_a$ . We then (safely) assume that  $n_{\text{HI}}(z_a)$  and  $H(z_a)$  are constants, and that we can take them outside of the integral, i.e

$$\tau_{\text{IGM}}(\nu_e) = \frac{cC_1 n_{\text{HI}}(z_a)}{H(z_a)} \int_{\nu_{\text{obs}}}^{\nu_e} \frac{d\nu}{\nu} \sigma_{\alpha}(\nu) \approx \frac{C_1 c n_{\text{HI}}(z_a)}{H(z_a) \nu_0} \int_{\nu_{\text{obs}}}^{\nu_e} d\nu \varphi(\nu) = \frac{C_1 c n_{\text{HI}}(z_a)}{H(z_a) \nu_0}. \quad (193)$$

Substituting the numerical values and the mean density of hydrogen in the Universe (Eq 193 above), we obtain

$$\boxed{\tau_{\text{IGM}}(\nu_e) = 1.7 \times 10^5 \left( \frac{1 + z_a}{4} \right)^{1.5}}. \quad (194)$$

The spectrum of the quasar at  $\lambda < \lambda_{\alpha}$  (i.e.  $\nu_e > \nu_{\alpha}$ ) is suppressed by a factor of  $e^{-\tau_{\text{IGM}}(\nu_e)}$  (see Eq 189). If all the hydrogen in the Universe were neutral, then the spectrum of the quasar should be completely ‘black’, i.e. no radiation whatsoever should have been able to leak through at  $\lambda < \lambda_{\alpha}$ . Figure 38 shows clearly this is not the case observationally. This can only mean that the hydrogen must not have been neutral, but ionised! We will talk about this next.

### C. Optical Depth of the Photoionized IGM to Photons with $\nu > \nu_0$

We previously discussed photoionisation equilibrium (see Eq 170 and discussion there) when discussing an ionizing radiation field as a ‘feedback’ process. Photoionisation equilibrium implies that photoionization balances recombination:

$$n_{\text{HI}}\Gamma = n_e n_p \alpha_{\text{rec}}(T), \quad (195)$$

where - as before -  $\Gamma$  denotes the photoionization rate ( $\text{s}^{-1}$ ) and  $\alpha_{\text{rec}}(T)$  denotes the recombination coefficient ( $\text{cm}^3 \text{s}^{-1}$ , check that units work out!). We define the neutral fraction  $x \equiv n_{\text{HI}}/[n_{\text{HI}} + n_p] \equiv n_{\text{HI}}/n$  as the fraction of hydrogen that is neutral. The remaining fraction  $1 - x$  is ionised. We have  $n_{\text{HI}} = xn$ ,  $n_e = n_p = (1 - x)n$ . Substituting, we

have

$$xn\Gamma = (1 - x)^2 n^2 \alpha_{\text{rec}}(T). \quad (196)$$

Assuming that  $x \ll 1$  (which is highly accurate as we will see later, but for now it is just an assumption),  $1 - x \sim 1$ , and

$$x = \frac{n\alpha_{\text{rec}}(T)}{\Gamma} \equiv \frac{\bar{n}(1 + \delta_b)\alpha_{\text{rec}}(T)}{\Gamma}.$$

(197)

The first equality is a result that we obtained before (see Eq 170). In the second equality, we defined  $n \equiv \bar{n}(1 + \delta_b)$  (just as a previously defined overdensity as  $\rho = \bar{\rho}(1 + \delta)$ ). The opacity of the IGM at some arbitrary density  $\delta$  is then

$$\tau_{\text{IGM}}(\nu_e) = (1 + \delta_b[z_a])^2 \frac{\bar{n}(z_a)\alpha_{\text{rec}}(T)}{\Gamma(z_a)} 1.7 \times 10^5 \left(\frac{1 + z_a}{4}\right)^{1.5}, \quad (198)$$

where we used (i) Eq 194 to get the opacity of the neutral IGM at mean density, (ii) that this opacity is  $(1 + \delta_b)$  higher at overdensity  $\delta_b$  purely because of the increase in the number density of H nuclei, and (iii) that the local neutral fraction is  $\frac{\bar{n}(1 + \delta_b)\alpha_{\text{rec}}(T)}{\Gamma}$ .

If we now assume that the baryons trace the dark matter (and hence that  $\delta_b = \delta$ , in which  $\delta$  denotes the overdensity in both baryons + dark matter), substitute our expression for  $\bar{n}(z_{\text{res}})$  (from Eq 183), and plug-in numerical values, we obtain

$$\tau_{\text{IGM}}(\nu_e) \approx 2(1 + \delta[z_a])^2 \frac{\alpha_{\text{rec}}(T)}{\Gamma(z_a)} \left(\frac{1 + z_a}{4}\right)^{4.5}. \quad (199)$$

Figure 39 shows the observationally inferred 'mean optical depth' as a function of redshift in the Ly $\alpha$  forest. From

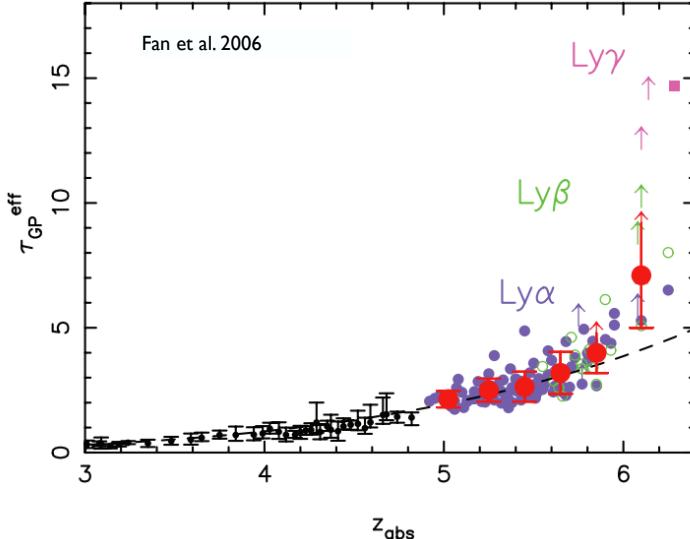


FIG. 39

this figure, and from Eq 199 and Eq 194 we can learn a few very interesting things:

- At  $z \sim 3 - 5$  the mean optical depth is order unity. For comparison, if the Universe were entirely neutral, we would have expected  $\tau$  to be of order  $10^5$ . In other words, the neutral fraction of hydrogen can only be  $x_{\text{HI}} \sim 10^{-5}$ . Remember how we talked about how the Universe recombined, and as a result became optically thin to radiation, which caused CMB photons to stream freely towards us. Observations of the CMB indeed indicate that the Universe became neutral. However, the intergalactic environment (the Universe outside of galaxies, which comprises most of the volume of the Universe) is highly ionized out to  $z \sim 6$ . The Universe must

have been *reionized* at some point following recombination. Understanding the *Epoch of Reionization* is one of the main science drivers for current and the next generation of telescopes. We will return to this in more detail in the next lecture.

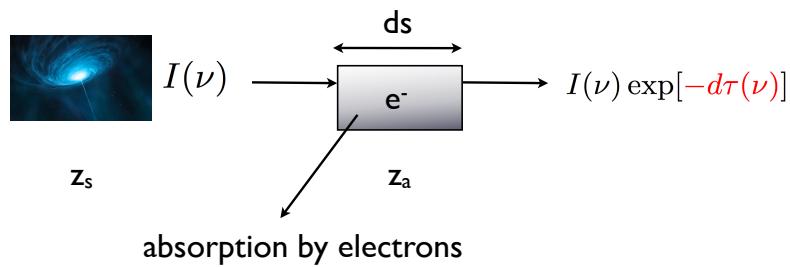
- The fact that opacity of the photoionized IGM is order unity implies that  $\alpha_{\text{rec}}(T)/\Gamma(z_a) \sim \mathcal{O}(0.1)$ . We previously discussed (see § XV.A) how photoionized gas has  $T \sim 1 - 2 \times 10^4$  K. Calculating the recombination coefficient depends only on atomic physics, and  $\alpha_{\text{rec}} \sim 10^{-13}$  cm<sup>3</sup> s<sup>-1</sup>. We therefore must have that  $\Gamma \sim 10^{-12}$ . This number is remarkably close to the values that have been inferred in the literature. In other words, a hydrogen atom that exists in the intergalactic medium will have to wait  $\sim 10^{12}$  s  $\sim 3 \times 10^4$  yrs before it gets ionized by a photon.
- The *dashed line* in Figure 39 shows  $\tau \propto (1+z)^{4.5}$ . This curve matches the observed redshift evolution of the Ly $\alpha$  forest opacity extremely well. The good match indicates that the ratio  $\alpha_{\text{rec}}(T)/\Gamma(z_a)$  cannot evolve much with redshift, which puts very interesting constraints on both  $T$ -evolution of the intergalactic medium and  $\Gamma$ . We will discuss this in more detail in the next lecture.

## XVIII. LY $\alpha$ FOREST II

In the previous lecture we found that the observed opacity of the intergalactic medium (IGM) is five (!) orders of magnitude less than expected if all the hydrogen in the Universe were neutral. We do not expect this hydrogen to have completely disappeared. Instead, hydrogen is highly ionized with only 1 part in  $10^5$  being neutral. We first show in § XVIII.A that the ionized IGM is also expected to not be entirely transparent to radiation, because free electrons can (Thomson)-scatter radiation at all wavelengths with equal probability (i.e. the cross-section for Thomson scattering does not depend on wavelength, unlike the Ly $\alpha$  absorption cross-section). Measuring the optical depth of the Universe to electron scattering,  $\tau_e$ , gives us direct measure for the amount of ionized gas between us and the CMB. We will talk about how the universe was reionized in § XVIII.B. Finally, we explain why the Ly $\alpha$  forest is such a powerful tool (see § XVIII.C).

### A. Derivation of Opacity of the IGM in Ionized Gas

We compute the opacity of the ionised component of the IGM. We use the same geometry as in the previous lecture and compute the change in intensity  $I(\nu)$  (evaluated at frequency  $\nu$ ) of radiation passing through a patch of ionised gas of length  $ds$ .



The change in intensity is  $I(\nu) \Rightarrow I(\nu) \exp[-d\tau(\nu)]$  (see Eq 189), in which  $d\tau(\nu)$  denotes the (differential) optical depth across the patch. For electron scattering (the main source of opacity), the optical depth  $d\tau(\nu)$  is given by

$$d\tau(\nu) = n_e(z_a) \sigma_T ds, \quad (200)$$

in which  $n_e(z_a)$  denotes the number density of free electrons in the patch,  $\sigma_T = \frac{8\pi}{3} r_e^2 = 6.65 \times 10^{-25}$  cm<sup>2</sup> denotes the cross-section for electron scattering, in which  $r_e = \frac{e^2}{m_e c^2}$  denotes the classical electron radius. This cross-section

is often referred to Thomson scattering cross-section.

The total optical depth is obtained by summing over all patches  $ds$  (i.e. integrating over  $ds$ )

$$\tau(\nu) \equiv \tau_e = \int_{\text{source}}^{\text{observer}} n_e(z_a) \sigma_T ds. \quad (201)$$

Note that we have dropped the frequency dependence. The Thomson scattering cross-section has no frequency dependence. The right-hand-side of the equation therefore has no frequency dependence. Therefore, the left-hand-side should also not depend on  $\nu$ . We labelled this frequency-independent opacity with subscript ‘e’  $\tau_e$  to emphasise that this is the opacity in free electrons. We recast the integral in terms of redshift (a quantity we can directly observe), and convert the integral over  $ds$  into an integral over  $dz$ . A photon that propagates a distance  $ds$  does so over a finite time interval  $dt = ds/c$ . During this time interval  $dt$ , the scale factor of the Universe changed by an amount  $da = \frac{da}{dt} dt$ . In other words

$$ds = cdt = c \frac{da}{\dot{a}} = c \frac{da}{H(a)a} = -c \frac{dz}{H(z)(1+z)}, \quad (202)$$

where we used that  $H \equiv \frac{\dot{a}}{a}$ , and  $a \equiv \frac{1}{1+z}$ . Substituting this into Eq 201 we find

$$\tau_e(z) = -c \int_{z_{\text{source}}}^{z_{\text{observer}}} \frac{n_e(z) \sigma_T dz}{(1+z) H(z)} = c \int_0^z \frac{n_e(z) \sigma_T dz}{(1+z) H(z)}, \quad (203)$$

which is the expression listed in the lecture notes. Note that we derived in the previous lecture that  $\bar{n}_{\text{HI}}(z) \sim 2.0 \times 10^{-7}(1+z)^3 \text{ cm s}^{-3}$ . If for simplicity we ignore helium, then we can say electrons are supplied exclusively by hydrogen, and when hydrogen is highly ionised  $n_e(z) = \bar{n}_{\text{HI}}(z)$ . Observations of the CMB allow us to directly constrain<sup>27</sup>  $\tau_e$ . The most recent Planck results give  $\tau_e = 0.058 \pm 0.012$  ([Planck Collaboration et al., 2016](#)). We know from the Ly $\alpha$  forest that the Universe is mostly ionized out to  $z \sim 6$ . For a fully ionized Universe out to  $z \sim 6$ ,  $\tau_e \sim 0.04$  and does not quite reach the observed value. In other words, the Planck data implies that the Universe was ionized out to  $z > 6$ .

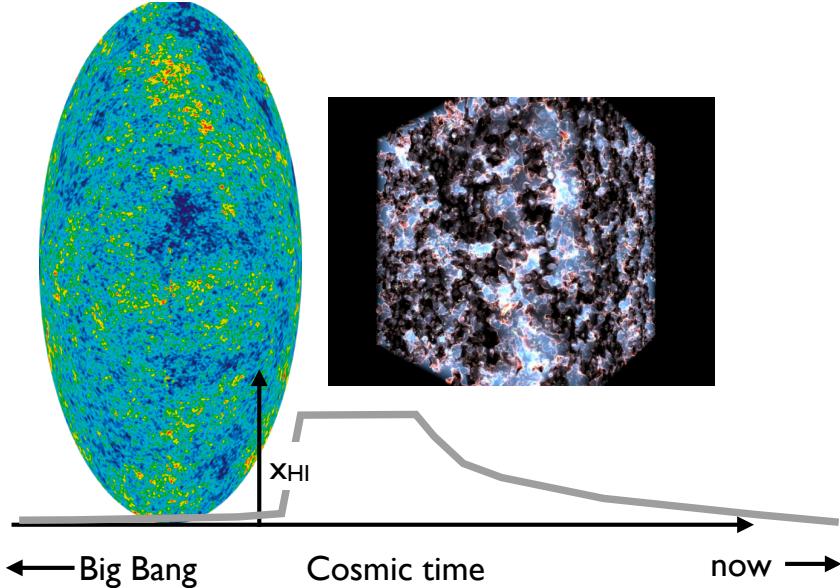


FIG. 40

The highly ionized state of the Universe as a whole contrasts sharply with the highly neutral state of the Universe following recombination. Figure 40 shows a schematic overview of the time evolution of the ionization state of the

---

<sup>27</sup> This is partially because electron scattering polarizes the CMB. The measured polarization therefore constrains  $\tau_e$ .

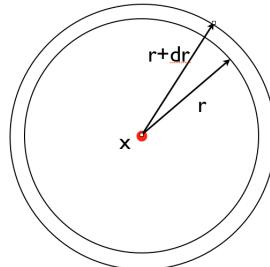
IGM. At early times - i.e. prior to recombination - the Universe was highly ionized. Recombination to a mostly neutral state occurred at  $z \sim 1090$ , which resulted in the release of CMB photons. At some point, the Universe started transforming back into an ionized state again. This reionization process was likely an extended process over time, driven by ionizing photons that were emitted by the increasing number of stars that formed in the Universe. Figure 40 shows a snapshot of a reionization simulation. Dark regions denote neutral gas, while ionized bubbles are colored. Figure 40 shows clearly that reionization likely was a highly inhomogeneous process, in which ‘bubbles’ of ionized gas formed around the first galaxies. These first galaxies were likely highly clustered, so the individual bubbles started overlapping, resulting in growth of large ionized bubbles driven by large ensembles of galaxies. To obtain some more insight into the reionization process, we first look in more detail into what is keeping the Universe ionized when reionization is complete. We do this in the next section.

## B. The Photoionization Rate, The Volume Emissivity of Ionizing Photons, Sources of Reionization

In previous lectures, I mentioned that the photoionisation rate  $\Gamma$  (units  $s^{-1}$ ) was directly proportional to the angle-averaged intensity  $J(\nu)$  (units  $\text{erg s}^{-1} \text{cm}^{-2} \text{Hz}^{-1}$ ):

$$\Gamma = C J(\nu_L), \quad (204)$$

where  $C$  is a constant<sup>28</sup> (with units  $\text{cm}^2 \text{Hz erg}^{-1}$ ). Our next goal is to connect  $J(\nu_L)$  to the total energy in ionising photons that are emitted per unit volume (the so-called ‘volume-emissivity’).



We evaluate the intensity  $J(\nu_L)$  (in  $\text{erg s}^{-1} \text{cm}^{-2} \text{Hz}^{-1}$ ) at a point  $x$ . We first compute the contribute  $dJ(\nu_L)$  from all sources of ionising photons separated from  $x$  by a distance  $[r, r + dr]$ . Note that  $dr$  is infinitesimally small. Each source ‘i’ within this shell has a luminosity  $L_i(\nu)$  (units  $\text{erg s}^{-1} \text{Hz}^{-1}$ ), and to compute the total contribution to the ionising intensity at  $x$  we have to sum over all sources. We have

$$dJ(\nu_L) = \sum_i \frac{L_i(\nu)}{4\pi r^2} \equiv \frac{dV_{\text{shell}}\epsilon(\nu_L)}{4\pi r^2}. \quad (206)$$

In the second step we used the definition of the volume emissivity, denoted with  $\epsilon(\nu_L)$ : which is just the total energy in ionising photons emitted per unit volume. We know the volume of the shell to be  $dV_{\text{shell}} = 4\pi r^2 dr$ , and therefore

$$dJ(\nu_L) = \epsilon(\nu_L) dr. \quad (207)$$

Summing over all shells gives as the following integral

$$J(\nu_L) = \int_0^\infty \epsilon(\nu_L) dr. \quad (208)$$

This integral diverges. This is known as ‘Olbers paradox’, which is named after German astronomer Olber, who

<sup>28</sup> We can calculate this constant if we specify the spectrum of the ionizing radiation field as:

$$\Gamma = \int_0^\infty d\nu \frac{J(\nu)}{h\nu} \sigma(\nu) \equiv J(\nu_L) \mathcal{G}(\alpha), \quad (205)$$

which assumes that the ionising radiation background has a power-law spectrum of the form  $J(\nu) = J(\nu_L) \left( \frac{\nu}{\nu_L} \right)^\alpha$ , where  $\nu_L$  denotes the frequency above which hydrogen can be photoionized, i.e.  $h\nu_L = 13.6 \text{ eV}$  ( $h$  denotes Planck’s constant). The function  $\mathcal{G}(\alpha)$  contains all relevant numerical constant, and shows explicitly that the overall value of the integral depends on the assumed slope of the power-law,  $\alpha$ . Note that  $\mathcal{G}(\alpha)$  gives the right dimensions.

concluded from similar reasoning that the night sky should be as bright as the sun, if the Universe were infinitely large, and infinitely old. In the lecture we discussed two solutions to this paradox applied to the ionizing radiation.

- In the first, we noticed that in reality, we have to take into account that we observe the shell at  $r$  at some earlier redshift  $z$ . Exactly as we argued above, a physical step  $dr$  corresponds to a step  $dz = \frac{H(z)(1+z)ds}{c}$  towards higher redshift. We should then take into account that the volume emissivity is changing with redshift (as we know that both the number density of galaxies, as well as their properties, change with redshift).
- In the second, we used that ionising photons are not free to travel uninterrupted through the IGM. Ly $\alpha$  forest observations indicate that the IGM is highly - but not completely - ionised. We see absorption features in the Ly $\alpha$  forest arising because of tiny left-over amounts of hydrogen. This hydrogen also places a limit on how far ionising photons can travel through the IGM. The ionized IGM is not fully transparent to ionising photons. A way to include the fact that ionising photons have a finite 'mean free path'  $\lambda_{\text{mfp}}^{\text{ion}}$  is simply by cutting off the integral at  $r = \lambda_{\text{mfp}}^{\text{ion}}$ , i.e.

$$J(\nu_L) = \int_0^{\lambda_{\text{mfp}}^{\text{ion}}} \epsilon(\nu_L) dr = \lambda_{\text{mfp}}^{\text{ion}} \epsilon(\nu_L). \quad (209)$$

Substituting Eq 209 into Eq 204 we are left with

$$\boxed{\Gamma = \lambda_{\text{mfp}}^{\text{ion}} \epsilon(\nu_L) C}. \quad (210)$$

This equation has interesting implications: the Ly $\alpha$  forest gives us  $\Gamma$  and as a function of redshift. While we have not discussed this, the Ly $\alpha$  forest can also constrain  $\lambda_{\text{mfp}}^{\text{ion}}$ . We can then directly infer  $\epsilon(\nu_L)C$ , the rate at which ionizing photons are emitted per unit volume. This is very interesting:  $\epsilon(\nu_L)C$  include the contribute from *all* galaxies in the Universe, irrespective whether the vast majority is too faint to have been detected by our telescopes. The Ly $\alpha$  forest enables us to make constrain certain properties of galaxies that we have not been able to see yet.

### C. Ly $\alpha$ Forest as a Cosmological Probe

## Interpreting the Ly $\alpha$ Forest

The Lyman alpha forest appears to probe only small overdensities.

$$\tau_{\text{IGM}}(z_a) \approx 2[1 + \delta(z_a)]^2 \frac{\alpha_{\text{rec}}(T)}{\Gamma} \left( \frac{1 + z_a}{4} \right)^{4.5}$$

$$\left. \begin{aligned} \alpha_{\text{rec}}(T) &\sim 10^{-13} \text{ cm}^3 \text{ s}^{-1} \text{ for } T \sim 10^4 K \\ \Gamma &\sim 10^{-12} \text{ s}^{-1} \end{aligned} \right\} \frac{\alpha_{\text{rec}}(T)}{\Gamma} \sim 0.1$$

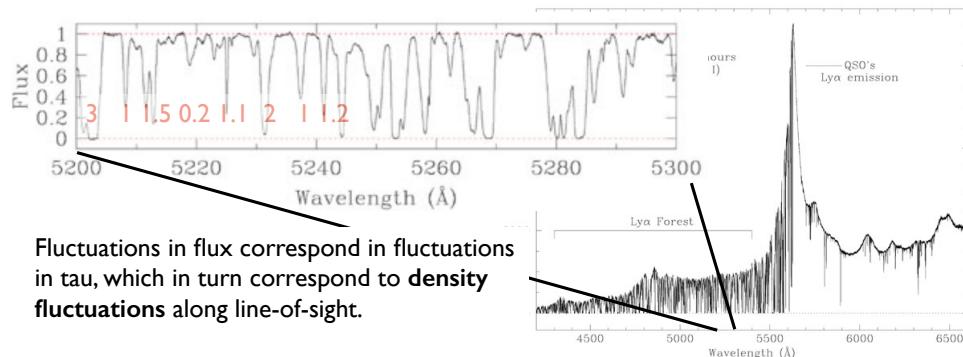


FIG. 41

Eq 198 showed that  $\tau \propto (1 + \delta)^2$  at a given redshift. As photons propagate from the quasar towards us, they will pass through the cosmological density field, which will be fluctuating along the line-of-sight. If we assume that the baryon overdensity closely tracks the dark matter density (this should be accurate on scales larger than the Jeans length, on smaller scales the baryons must look more ‘washed’ out), then the opacity in the Ly $\alpha$  forest should respond strongly to these density fluctuations: i.e., the opacity changes by  $\sim 40\%$  just going from  $\delta = 0 \rightarrow \delta = 0.2$ . Figure 41 illustrates this point: the measured flux in the forest at wavelength  $\lambda$  translates to a constraint on  $\tau_{\text{IGM}}(\lambda) = \tau_{\text{IGM}}(z_a)$ , which then can be converted into  $\delta(z_a)$ . This procedure allows us to map-out the mass (dominated by dark matter) density distribution along the line-of-sight. The accuracy of the inferred density distribution depends on how well the baryons track the dark matter. For many purposes, the baryons track the dark matter well. Indeed, it is possible to reproduce many statistical properties of the Ly $\alpha$  forest with dark matter only simulations.

In the lectures we discussed how the Baryon Oscillation Spectroscopic Survey (BOSS) uses this constrain the matter distribution throughout the Universe using millions of quasar spectra. The Ly $\alpha$  forest can recover some key features in the matter distribution, incl. for example the acoustic peak (see § XI.A).

## XIX. GALAXY FORMATION IN MORE DETAIL

We return to the question of how galaxies form and evolve. We apply concepts we learned in previous lectures, including the spherical top-hat model (see § V), gas cooling (§ VI.F), and the cooling criteria for gas collapse (§ VI.G), and see how these concepts apply in our current understanding of galaxy formation. We will also discuss how these concepts will allow us to understand some of the key challenges in galaxy formation.

### A. Cold vs Hot Model Accretion in Spherical Collapse

The spherical top-hat model assumed a uniform (enhanced, compared to the universal average) density inside a sphere. In § VI.D we briefly qualitatively discussed how the picture changed for more general, spherically symmetric density profiles in which  $\delta = \delta(r)$ . We can generally identify a maximum density for any density perturbation, and then measure the spherically averaged density contrast in concentric shells around it. Because we started out at the maximum, the *average* overdensity,  $\bar{\delta}(r)$ , within a (comoving) sphere of radius  $r$  centered on the maximum, *must* decrease with  $r$ . Compare the time evolution of two mass shells at  $r_1$  and  $r_2$ , where  $r_1 < r_2$ .

- We know that  $\bar{\delta}(r_1) < \bar{\delta}(r_2)$ .
- Linear theory tells us that both  $\bar{\delta}(r_1)$  and  $\bar{\delta}(r_2)$  grow with time as  $\propto a(t)$  (see Eq 55).
- We also learned that an object virializes when its linear overdensity reaches  $\delta_{\text{lin}} = 1.69$  (see Eq 67). It then follows that the mass shell at  $r_1$  will virialize earlier than the mass shell at  $r_2$ .

For more realistic density profiles, we therefore expect virialized objects to ‘grow’ from the inside out. This applies to both dark matter and baryons. What is different for the baryons is that they get shock heated to a temperature that increases in time. Figure 42 shows the time evolution of many radial gas shells in a full hydrodynamical

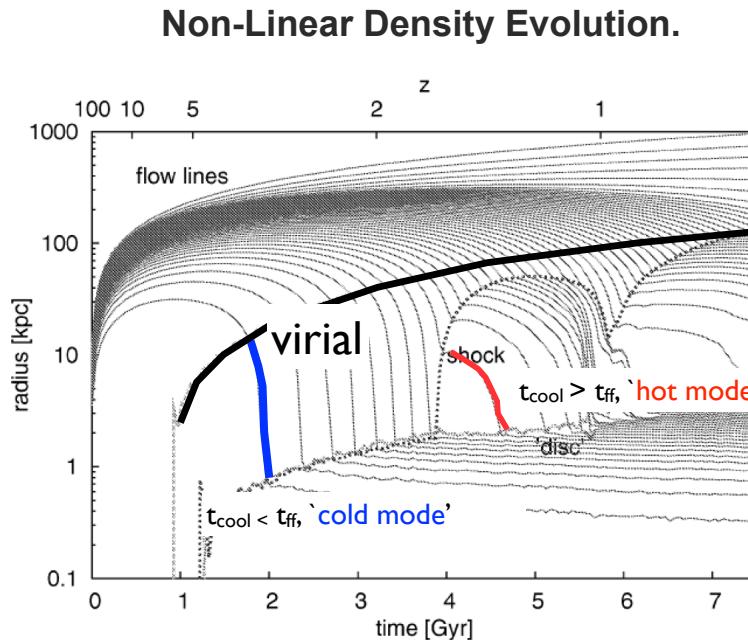


FIG. 42

simulation (each mass-shell is represented by a so-called ‘flow line’), with realistic cosmological initial conditions for  $\delta(r)$ . The Figure shows how the virial radius (*thick black solid line*, denoted with ‘virial’) grows in cosmic time. The virial radius traces the size of the virialized dark matter halos. The flow lines show clearly that the gas does not stop at all at the virial radius (here at  $t < 4$  Gyr, **but be aware that this number depends on the halo mass!**). This is because  $t_{\text{cool}} \ll t_{\text{dyn}}$ : the gas cools so fast that it cannot reach hydrostatic equilibrium. At these times, the

gas collapses to the center practically in free-fall. The gas still comes to a halt at  $r \sim 1$  kpc, which is where the actual galaxy is said to start<sup>29</sup>.

At later times (here at  $t > 4$  Gyr) the time-evolution of the gas becomes more complicated. The gas starts to shock, which is when you see clear kinks in the flow lines, which correspond to abrupt changes in both gas temperature, velocity, density. For  $4 \text{ Gyr} < t < 7 \text{ Gyr}$ , we see the gas shock at radii smaller than the virial radius. For  $t > 7 \text{ Gyr}$  the Figure shows that the gas shocks at the virial radius (just like we discussed in § VI.E).

As mentioned above, the gas that reaches the virial radius at  $t < 4$  Gyr collapses to the center of the dark matter halo practically in free-fall. There are no clear kinks in the flow lines associated with shocks. Indeed, the gas does not get shock heated at all, and collapses while staying at the temperature it had as intergalactic gas (which is  $\sim 10^4$  K, as we know from lectures on the Ly $\alpha$  forest, see § XVII). The gas is said to be accreted ‘cold’. In contrast, at later times  $4 \text{ Gyr} < t < 7 \text{ Gyr}$ , the gas still collapses but the collapse rate is said by the cooling rate of the gas. This gas has to be accreted onto the central ‘galaxy’ after having passed through a hotter phase. This mode of accretion is referred to as ‘hot-mode’ accretion.

### Abandoning Spherical Symmetry: Cold Flows

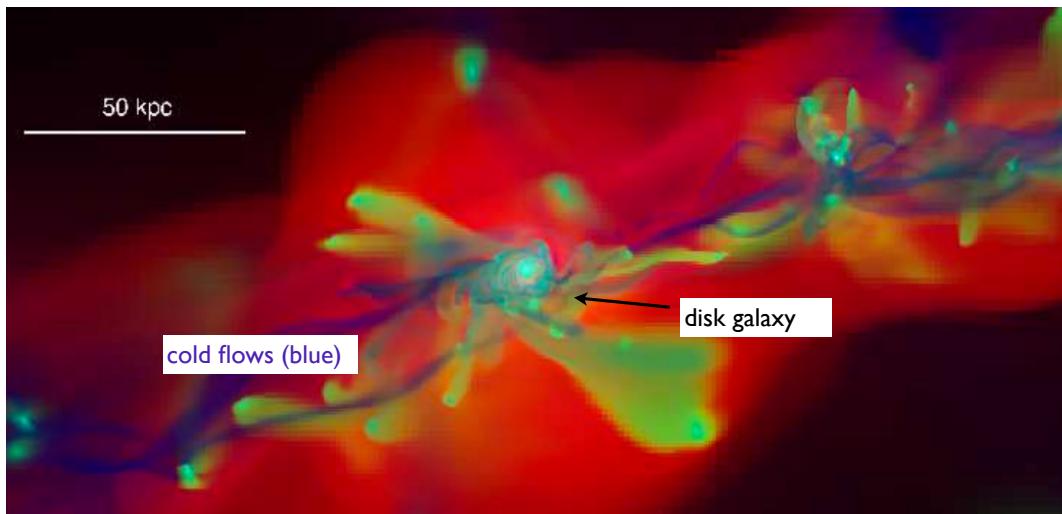


FIG. 43

The transition from cold-mode to hot-mode accretion is linked to the time when  $t_{\text{cool}} = t_{\text{dyn}}$ . Recall that  $t_{\text{dyn}} \propto \rho^{-1/2}$ , while  $t_{\text{cool}} \propto \rho^{-1}$ . In general, (over)density decreases with distance from the density maximum, and we expect that if the condition  $t_{\text{cool}} < t_{\text{dyn}}$  is more easily fulfilled<sup>30</sup> at small  $r$ . We can define the *cooling radius*,  $r_{\text{cool}}$ , where  $t_{\text{cool}} = t_{\text{dyn}}$ . For  $r < r_{\text{cool}}$  [ $r > r_{\text{cool}}$ ], we expect accretion to be cold [hot].

<sup>29</sup> Because the simulation is spherically symmetric, there is not really a galaxy in the center. In this simulation, the gas and dark matter were given a realistic amount of angular momentum which prevents the gas from collapsing to  $r = 0$ .

<sup>30</sup> In fact, it was realized that the efficient cooling at small  $r$  causes too efficient star formation. This is known as the ‘overcooling’ problem. The overcooling problem can be alleviated by invoking (again) feedback. Once stars start forming, they affect their environments in ways that can suppress subsequent star formation.

## B. Non-Spherical Gas Accretion and Cold Flows

The spherically symmetric simulations are very powerful for learning about the basic physics of galaxy formation, they do not give you galaxies. In terms of gas accretion, abandoning spherical symmetry introduces one major change: in spherically symmetric simulations, all physical properties of the gas depend entirely on  $r$ . In the real Universe, we expect density to fluctuate throughout a spherical shell of radius  $r$  and thickness  $dr$ . In the real Universe, we know that on large scales, matter is distributed in the cosmic web, which consists of filaments, walls, and knots. The more massive dark matter halos typically reside in knots connecting two or more filaments. We expect such filaments to introduce azimuthal density variations at some fixed  $r$ : along the high density filaments, we expect cooling to be more efficient, and accretion more likely to be ‘cold’ than in directions that do not intersect filaments.

Figure 43 shows a snap-shot from a cosmological hydrodynamical simulation by Agertz et al. (2009) which illustrates how this affects gas accretion onto simulated galaxies. A disk galaxy sits in the center of the snap-shot, taken at  $z = 3$ . The *blue filaments* show dense gas that is being accreted along filaments. The *red gas* has been shock heated to the virial temperature of the dark matter halo hosting this galaxy ( $T_{\text{vir}} \sim 10^6$  K). The green clouds show metal rich gas that was stripped from smaller galaxies. In this case, the hot, virialized gas has a temperature that is  $\sim 100$  times higher than that of the cold accretion streams, which causes the hot gas to compress the cold accretion streams into thin ‘cold flows’. The presence and properties of cold flows, and how they depend on redshift and dark matter halo mass, are still hotly debated subjects. One reason is that there currently is no direct observational evidence for the existence of cold flows in the Universe, though it has been argued that without cold flows it is not possible to reproduce the observed number of Lyman limit systems (absorbers in the Ly $\alpha$  forest with  $\log N_{\text{HI}}/(\text{cm}^{-2}) \sim 17 - 20$ ) in the Ly $\alpha$  forest.

## C. Galaxy Formation: Why Disks?

The previous discussion focusses on how galaxies get their gas. This picture does not explain the diverse structural properties of observed galaxies (as represented by Hubble’s classification scheme in Fig 44). It is beyond the scope of this lecture to give a quantitative explanation for the origin of different galaxy types. Instead, we will qualitatively explain some of the basic ideas.

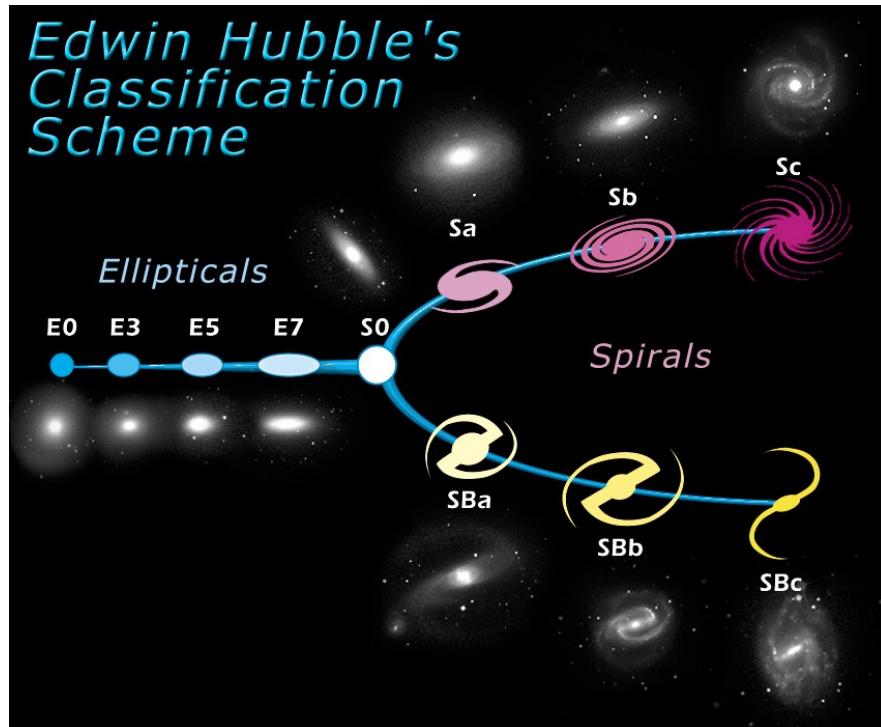


FIG. 44

We mentioned briefly above (see § XIX.A) that the non-zero angular momentum of the gas causes it to settle into an object of finite size, the galaxy. The natural configuration for such a gaseous system is a disk. The self-gravity of the disk can then cause the disk to collapse further along its minor axis. The conversion of gas into stars in the disk is discussed in more detail below in § XX. To understand why a collisionless system of stars settles into a beautiful spiral, or barred spiral pattern requires an understanding of galactic dynamics. We will not discuss this in our lecture. For those interested, I would refer you to the book by Binney & Tremaine on Galactic Dynamics. The main points are that spiral structure and bars arise naturally in rotating self-gravitating, disky systems.

#### D. Galaxy Formation: Why Ellipticals?

The previous discussion illustrated that it is not difficult to explain spiral, and barred spiral galaxies. The formation of elliptical galaxies is not fully understood. It is well known that smashing together two spiral galaxies - the more formal way to describe this is as the two spiral galaxies 'merging' - gives you an elliptical galaxy. As we discussed previously, in the current cosmological model, structure forms hierarchically (see § IX.A), with low mass galaxies forming first, and merging to form more massive objects. We generally do expect mergers to happen frequently. However, it is still unclear what fraction of ellipticals formed this way.

#### “Merger Sequence”

Simulation of the [Mice galaxies \(NGC 4676\)](#) from [Toomre & Toomre \(1972\)](#).

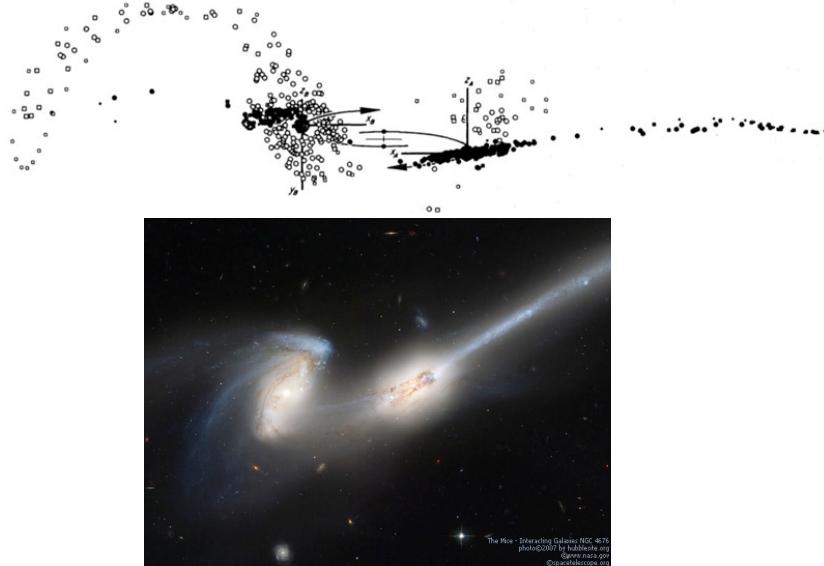


FIG. 45

Figure 45 illustrates that the Universe provides a rich laboratory for testing the collision of two merging disk galaxies. Simulations of two merging spiral galaxies have been carried out in the early 70s. It is stunning to see that most ‘snapshots’ of simulated mergers of galaxies can actually be seen somewhere in the Universe. Another example (not shown here) we showed in the lecture were the beautiful ‘shell’ elliptical galaxies, which are ordinary elliptical galaxies, with multiple very thin shells surrounding them. These objects arise naturally during the final stages of a merger between two elliptical galaxies.

## XX. STAR FORMATION

We spent quite some time discussing quantitatively how density fluctuations grow under the influence of gravity and resisted by pressure. A virialized dark matter halo is  $\sim 200$  times denser than the Universe as a whole at the time of

virialization. We already discussed that the structure of virialized dark matter halos is something that introduces complications as simulated dark matter halos do not seem to agree with constraints on the gravitational potential wells that we obtain from studies of gas kinematics. This underlines that collapse by  $\sim 2$  orders of magnitude in density is already complicated. In order to explain stars, we need to collapse further and increase density by an additional  $\sim 25$  orders of magnitude. Not surprisingly, the theory of star formation is extremely complicated. We will not take the approach in which we discuss star formation from first principles. Instead, we start with observations, which indicate that stars form inside so-called ‘Giant Molecular Clouds’ (GMCs).

GMCs are dense ( $n \sim 10^2 - 10^5 \text{ cm}^{-3}$ , i.e.  $\sim 10^9 - 10^{12}$  times denser than the average density in the present-day Universe), massive ( $10^4 - 10^6 M_\odot$ ) clouds that exist throughout galaxies. You can see the nearest GMC with your eyes in the Orion constellation (shown in Fig 46). GMCs have been observed to be sites of ongoing star formation, containing young (few Myr), hot, massive stars. Fig 46 also indicates the GMCs have significant structure to them, with embedded denser ‘clumps’, which in turn contain even denser ‘cores’. The cores are thought to be the birth places of stars. We decompose star formation into two different questions: (i) how do GMCs form?, and (ii) how do stars inside GMCs form?

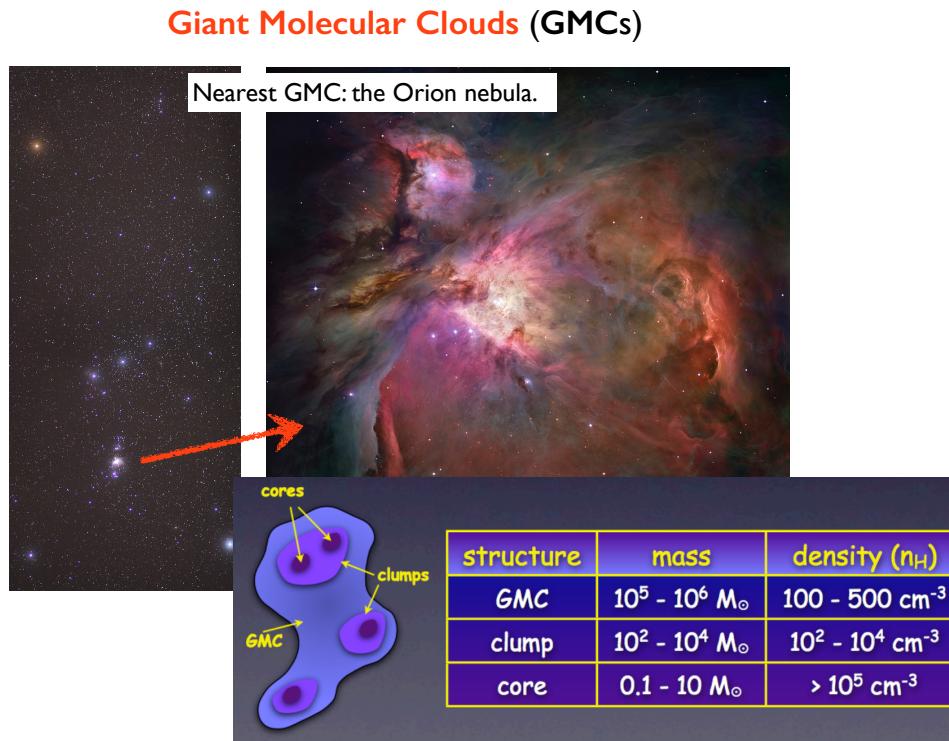


FIG. 46

#### A. How do Giant Molecular Clouds (GMCs) Form?

GMCs are generally thought to have formed from gaseous disks via some sort of instability. Two important examples of instabilities include the (*thermal instability*) and (*gravitational instability*).

**Thermal Instability.** The thermal instability can be understood from the cooling curve ( $\Lambda(T)$ , see Fig 10). Generally speaking we expect gas inside the gaseous disk to be subject to perturbations (for example, because extra gas is being accreted onto the disk). This perturbation can locally heat the gas. Suppose the gas was initially at  $T = 10^4 \text{ K}$ , then a slight boost in  $T$  to (say)  $T \sim 1.3^4 \text{ K}$  increases the cooling efficiency of the gas, as  $\frac{d\Lambda}{dT}|_{T=10^4 \text{ K}} > 0$ . The enhanced cooling rate then brings the gas back to its initial equilibrium  $T = 10^4 \text{ K}$ . This situation is stable. However, if we had started at a  $T$  where  $\frac{d\Lambda}{dT} < 0$ , then the increase in  $T$  would lead to reduced cooling, and further

heating of the gas. The increases pressure of this locally heated gas causes it to expand into its surroundings, which could compress the surrounding gas, which in turn would enable this denser surrounding to cool more efficiently, etc.

As before, we denote the cooling rate per unit volume  $C(T) = n^2\Lambda(T)$ , and heating rate per unit volume with  $n\Gamma$ . Gas is stable if

$$\frac{d}{dT} [n\Gamma - n^2\Lambda] > 0. \quad (211)$$

Stability criteria such as these generally cause the gas in disks to break-up into discrete phases, in which colder, denser clumps are embedded within a hot ionized medium. The theory that describes the structure of the ISM is beyond the scope of these lectures. The main take away point is that thermal instabilities can lead to the formation of ‘multi-phase’ gas with hot, tenuous gas co-existing with colder, denser gas. GMCs would form from this colder, denser phase.

**Gravitational/Toomre Instability.** The gaseous disks have contracted significantly compared to the dark matter. Self-gravity of the gas inside the disk can cause gas to collapse under the influence of its own gravity inside the disk. Whether this happens depends on something very analogous to the Jeans criterion which we discussed previously (in § IV.F)

The Jeans instability criteria states that only perturbations with a physical scale  $R$  larger than the Jeans-Length (here  $R_J$ ) can collapse under the influence of gravity, while smaller perturbations ( $R < R_J$ ) cannot collapse under the influence of gravity due to pressure forces. A quick way to get the Jeans length is by considering a spherical, self-gravitating ball of gas with radius  $R$ , mass  $M$ , and temperature  $T$ . The total energy,  $U_{\text{tot}}$ , of this cloud is

$$U_{\text{tot}} = \frac{1}{2}Mc_s^2 - \frac{GM^2}{R}, \quad (212)$$

where we used that the 3D velocity dispersion of atoms/nuclei inside the cloud equals the sound speed (to within a factor of order unity), i.e.  $\sigma^2 = c_s^2$ . The gas cloud is unstable and will collapse if  $U_{\text{tot}} < 0$ . The onset of instability thus occurs when, which translates to

$$R = c_s \sqrt{\frac{3}{8\pi G\rho}} \quad (213)$$

where we used that  $M = \frac{4}{3}\pi R^3\rho$ . This simple analysis almost gives us the Jeans length to within a factor of order unity compared to our more detailed analysis of § IV.F.

Gaseous disks are clearly not spherical, which quantitatively changes the argument presented above. We can repeat the analysis above, but applied to a 2D gaseous disk, for which  $M = \pi R^2\Sigma$ , where  $\Sigma$  is the *gas surface density* (density per unit area of the disk). The condition  $U_{\text{tot}} = 0$  now translates to a different expression for the analogous Jeans length:

$$R = \frac{c_s^2}{2G\pi\Sigma}. \quad (214)$$

Note the different dependence on sound speed, and surface density. This exercise serves mostly as a stepping stone towards the next, relevant calculation.

If we let the disk from the previous example rotate, then we add a term to the kinetic energy of the disk, i.e.

$$U_{\text{tot}} = \frac{1}{2}M[c_s^2 + R^2\Omega^2] - \frac{GM^2}{R} = \frac{\pi}{2}R^2\Sigma[c_s^2 + R^2\Omega^2] - G\pi^2R^3\Sigma^2. \quad (215)$$

The onset of instability happens when  $U_{\text{tot}} = 0$  which leads to

$$\frac{1}{2}[c_s^2 + R^2\Omega^2] - G\pi R\Sigma = 0. \quad (216)$$

This quadratic solution for  $R$  we can solve easily. However, solutions for  $R$  which are negative (or imaginary) are not physical. We would first like to know whether this equality has any positive real solutions. We can find this out

quickly by realizing that the term  $\frac{1}{2}[c_s^2 + R^2\Omega^2] - G\pi R\Sigma$  describes a parabola with a minimum at  $R_{\min} = \frac{G\pi\Sigma}{\Omega^2}$ . The minimum energy  $U_{\min}$  at this point is  $U_{\min} = \frac{1}{2}c_s^2 - \frac{1}{2}\frac{(G\pi\Sigma)^2}{\Omega^2}$ . There exist only at least one real, positive solution for  $R$  if  $U_{\min} < 0$ , i.e. when

$$Q \equiv \frac{c_s\Omega}{G\pi\Sigma} < 1. \quad (217)$$

That is,  $Q = 1$  (which translates to  $U_{\min} = 0$ ) marks the onset of gravitational instability. This criterion is known as the Toomre instability criterion, and the parameter  $Q$  is known as the Toomre Q parameter. Eq 217 is very powerful. Suppose we have detailed observations of a gaseous disk (e.g. through its 21-cm line, as in Fig 28), then we can measure  $R$ ,  $v_{\text{rot}} = \Omega R$ . From the HI observations we have a measure of the gas surface density  $\Sigma$ . If we also know  $T$ , we get  $c_s$ , and can compute  $Q$ . Eq 217 then can tell us whether gas can gravitationally collapse to higher densities.

The previous highlights that there once gas settles into a rotating gaseous disk in the center of a dark matter halo, there are physical mechanisms that can cause to break up and collapse into denser structures. The densest structures should lead to the formation of GMCs.

## B. How do Stars form from GMCs?

To go from GMCs with  $n_{\text{H}} \sim 10^2 - 10^5 \text{ cm}^{-3}$  to stars that have  $n_{\text{H}} \sim 10^{24} \text{ cm}^{-3}$ , we need to collapse by a an additional linear scale of  $10^7$  (i.e  $R_{\text{GMC}} \rightarrow 10^{-7}R_{\text{GMC}}$ ). Contracting by such a large linear factor immediately introduces computational/theoretical challenges. For example, (i) GMCs rotate. Collapse by a linear factor of  $10^7$ , requires the GMC to lose an enormous amount of angular momentum, otherwise stars would be spinning very fast, (ii) The gravitational binding energy changes by a factor of  $U_{\text{bind}} \propto R^{-1}$ , which causes the binding energy to be  $\sim 10^7$  times more negative. All this extra binding energy must have been radiated away, even though observations indicate that GMCs are optically thick at most wavelengths, which implies that radiation is not efficient at carrying away binding energy at all. We again approach this problem by considering observational data, which provides some hints as to how star formation proceeds. We will see that we can then automatically address some of the challenges mentioned above.

Observations of GMCs indicate that these are associated with stars with life times of up to  $\sim 10$  Myr. For

**GMCs are ‘supersonically turbulent’.**

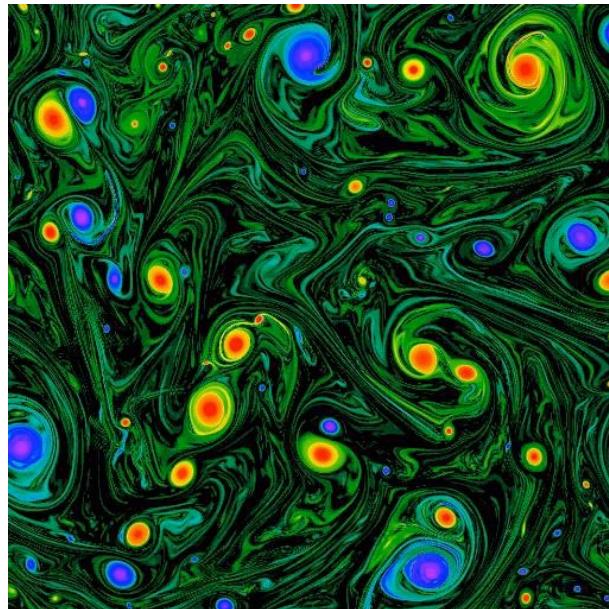
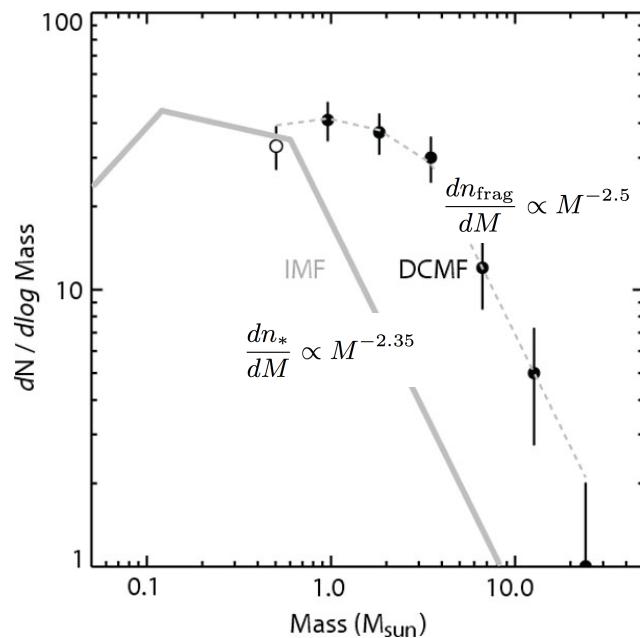


FIG. 47

comparison, the collapse time of GMCs is significantly shorter at  $t_{\text{coll}} = 1/\sqrt{4\pi G\rho} \sim 3(n/10^2 \text{ cm}^{-3})^{-1/2}$  Myr. GMCs must therefore not been in free-fall, and so something must be preventing gravitational collapse from happening. This is paradoxical, since we want stars to form in GMCs. Another way of phrasing this is to say that GMCs are somehow inefficient at converting gas into stars. Observations indicate that GMCs convert a few % of their gas into stars for each free-fall time. So something is preventing the GMC from collapsing under its own weight. It cannot be ordinary gas pressure. For the gas temperatures and densities inferred for GMCs from observations, we the Jeans mass is  $M_J \sim 50 \left(\frac{n}{10^3 \text{ cm}^{-3}}\right)^{-1/2} \left(\frac{T}{10 \text{ K}}\right)^{3/2}$ . The Jeans mass is much smaller than the mass of the GMC, and we do not expect pressure to prevent the collapse of the GMC.

The previous shows that it is not ‘conventional’ thermal pressure that is preventing the GMC from collapsing. Observations indicate that GMCs are highly turbulent (see Fig 47 for Van Gogh’s - what turns out to be quantitatively accurate - depiction of turbulence), with the velocity dispersion of the gas that is inferred from observed line widths is  $\sigma_{\text{turb}} \sim 5 \text{ km s}^{-1}$ , which is much ( $\sim 15 - 30$  times) larger than the thermal velocity (or sound speed,  $c_s$ ) that is associated with the inferred temperatures for the GMCs. In other words, gas inside GMCs is moving around at velocities  $\sim 15 - 30$  times the local sound speed. GMCs are thus highly supersonically turbulent. In supersonically turbulent media, the Jeans mass is enhanced to

### Initial Mass Function vs Fragment Mass Function



$$M_J \propto c_{s,\text{eff}}^3, \quad c_{s,\text{eff}} = \sqrt{c_s^2 + \sigma_{\text{turb}}^2}. \quad (218)$$

Since  $\sigma_{\text{turb}} \sim 15 - 30 c_s$ , supersonic turbulence boosts the Jeans mass by a factor of  $15^3 - 30^3 \sim 10^3 - 10^4$ . What is further interesting is that observations indicate that  $\sigma_{\text{turb}}$  is scale-dependence. That is,  $\sigma_{\text{turb}}$  depends on the scale  $R$  over which we measure it (just like  $\sigma(M)$  depends on smoothing scale  $M$ , see § VII) as  $\sigma_{\text{turb}}(R) \propto R^{1/2}$ . In other words, turbulent motion of the gas is most important on the largest scales. This is important: *supersonic turbulence can prevent the cloud as a whole from collapsing, while on the smallest scales gas can collapse to higher densities*. In addition, simulations indicate that supersonic turbulence gives rise to a log-normal density distribution<sup>31</sup> of the gas. Supersonic turbulence therefore also promotes the formation of high-density regions, in which the local Jeans mass

<sup>31</sup> This means that  $\log \rho$  has a Gaussian distribution for which the standard deviation is set by  $\sigma_{\text{turb}}(R_{\text{GMCs}})$ .

is reduced and which are therefore more likely to collapse.

Supersonic turbulence naturally gives rise to a distribution of collapsed ‘clumps’ in simulations. The *black data points* in Figure XX.B shows the number (density) of clumps as a function of their mass for a simulation of a supersonically turbulent cloud. For comparison, the *grey thick line* shows the stellar initial mass function (IMF), which gives the number (density) of newly born stars in star forming regions as a function of their mass. The similarity of the slope  $\frac{dn}{dM} \propto M^{-2.35}$  is stunning, and provides further support that supersonic turbulence may play a key role in converting GMC gas into stars.

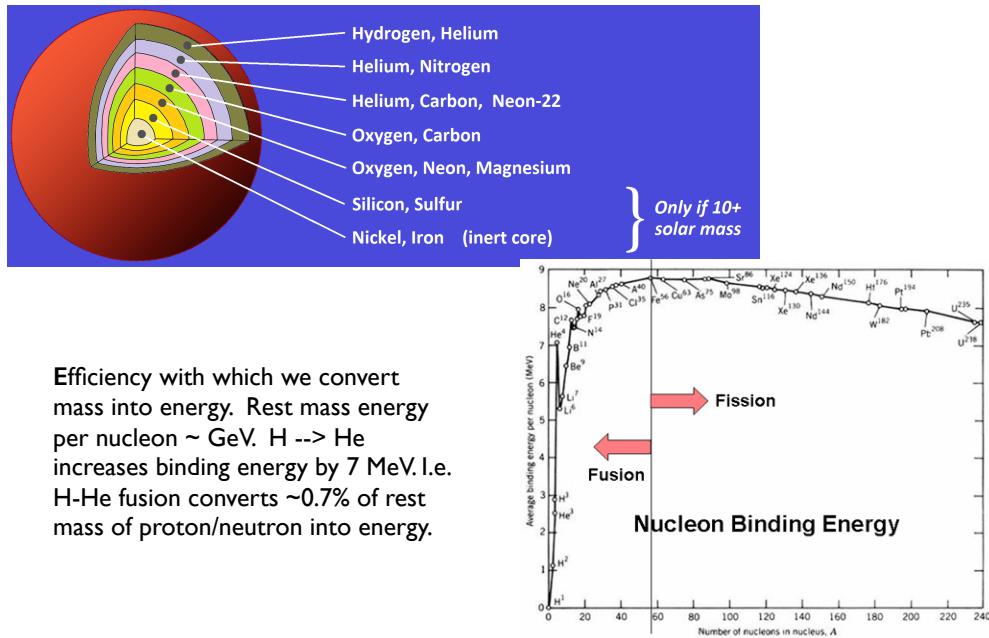
If supersonic turbulence is what drives star formation in GMCs, we have not addressed the origin of this turbulence. This is again something that has not been determined definitively yet. Possible explanations include: accretion of new gas onto the galaxy, spiral density waves (i.e. if the GMC passes through a spiral arm of the galaxy), but also processes that result from star formation such as supernova explosions (external to the GMC of interest), protostellar outflows, stellar winds, ionizing radiation from young, hot stars. The latter category arguably falls in the classic ‘chicken-or-egg’ category<sup>32</sup>.

## XXI. BLACK HOLES & THEIR FORMATION

### A. Origin of Stellar Mass Black Holes

To properly understand the formation of stellar mass black holes, we need to understand stellar evolution. Stellar evolution is a topic on its own, and we did not discuss it in this lecture. Instead, I will briefly describe how stellar mass black holes are thought to have formed. A more detailed derivation of the Chandrasekhar mass - the maximum mass that can be prevented from collapsing into a black hole by the electrons degeneracy pressure - is presented in § XXI.B. It is not necessary to know this derivation for the exam. I have kept it in the notes anyway.

### Formation of Stellar Mass Black Holes.



Efficiency with which we convert mass into energy. Rest mass energy per nucleon  $\sim$  GeV.  $H \rightarrow He$  increases binding energy by 7 MeV. i.e. H-He fusion converts  $\sim 0.7\%$  of rest mass of proton/neutron into energy.

<sup>32</sup> To clarify: it is unclear whether the chicken went supernova first, or whether it decided to cross the road later on.

Stars are powered by thermonuclear fusion of hydrogen into helium for at least a significant fraction of their life time. The energy that is released traverses the stars interior, prevents it from collapsing, and is radiated away. Fusion of hydrogen into helium converts 0.7% of the hydrogen's rest mass into energy. The more massive the star, the higher its central pressure. The enhanced central pressure of more massive stars enables them to fuse helium into nitrogen, helium into carbon etc. Figure XXI.A shows the ‘union’ like structure of massive star, where the heaviest nuclei are synthesized towards the center of the star. Figure XXI.A also shows how the nuclear binding energy of becomes more negative until  $A \sim 57$  (Fe), after which energy is required to make more massive nuclei.

Once the iron core forms, nuclear fusion in the core no longer helps to prevent gravitational collapse. As the central iron core gets compressed, eventually electron degeneracy pressure kicks in. The physical origin of this pressure source, of quantum mechanical origin, is that electrons are fermions, and that no two fermions can be in the same quantum state (see § XXI.B). The core, held up by degeneracy pressure, is no longer able to prevent collapse once the mass of the core exceeds  $M \sim 1.5M_{\odot}$ . This threshold mass is known as the Chandrasekhar mass. Once the core becomes more massive than this mass, the core collapses straight into a black hole. The black hole is characterized by its mass  $M$  and its Schwarzschild radius,  $R_s$ :

$$R_s = \frac{2GM}{c^2} \sim 3(M/M_{\odot}) \text{ km.} \quad (219)$$

This defines the radius from which nothing can escape. A simple way to view this is to see it as the ‘radius’ of the black hole.

After its formation, a black hole can grow in mass by accreting additional matter. (The majority of) Accretion onto a black hole has to go through an accretion disk, as direct flows into a black hole is only possible for gas that has practically zero angular momentum (i.e. when the velocity vector of far away gas is pointed almost perfectly at the black hole). Gas inside an accretion disk slowly moves inward on practically circular orbits. We show in § XXI.C that accretion disks are generally thought to be geometrically thin. In § XXI.D we compute how the temperature of the accretion disks varies as a function of distance from the black hole (and as a function of black hole mass). This information allows us to calculate the spectrum emerging from such a ‘multi-colored’ disk (see § XXI.E). This spectrum depends on - among other things - the innermost radius of the accretion disk. This innermost radius of the accretion disk corresponds to the innermost stable circular orbit (ISCO, see § XXI.F). We then show how the ISCO relates to ‘radiative efficiency’ of gas being accreted onto a black hole (§ XXI.G). We then discuss the maximum luminosity can be associated with an accretion disk (§ XXI.H), which we can then relate to the maximum black hole growth rate in § XXI.I. This final discussion shows why it is presently a puzzle why the Universe appear to contain super massive black holes ( $M \sim 10^9 M_{\odot}$ ) when the Universe was less than 1 Gyr old. Yeah!

## B. Quantitative Derivation of the Chandrasekhar Mass: Not Needed for Exam!

We did not discuss this in the lecture, and it is not part of the material you need to know for the exam, but for completeness I have given a simple derivation of the Chandrasekhar mass here. In a previous lecture in which we derived the density profile for the ‘isothermal’ sphere, we introduced the ‘distribution function’  $f(\mathbf{x}, \mathbf{v})$  which denotes the density of particles in 6-dimensional phase-space. We had to invoke this distribution function to describe a system of stars. The ordinary number density (number per unit volume),  $n(\mathbf{x})$  [we will drop the  $\mathbf{x}$  from now on], relates to the distribution function as

$$n = \int d^3 \mathbf{p} f(\mathbf{x}, \mathbf{v}). \quad (220)$$

We are going to use this phase-function to describe the electron degeneracy pressure in a gas.

To do this we start with Heisenberg’s uncertainty principle that states that particles cannot be confined in a phase-space volume element smaller than

$$(\Delta x \Delta p)^3 \geq \hbar^3. \quad (221)$$

You can think of  $\hbar^3$  as the minimum volume in phase-space into which we can squeeze a particle. Electrons (and also neutrons & protons) are fermions, and for (spin 1/2) fermions we can only fit a maximum of 2 particles in such a volume element. The maximum phase-space density of fermions is therefore  $f_{\max}(\mathbf{x}, \mathbf{v}) = 2\hbar^{-3} \sim \hbar^{-3}$  (where we drop factors of order unity).

Let us now compute the number density,  $n$ , under the assumption that the phase-space density is maximal. This number density is given by

$$n = \int d^3\mathbf{p} f_{\max}(\mathbf{x}, \mathbf{v}) \sim \int d^3\mathbf{p} \hbar^{-3} = \int_0^{p_F} dp p^2 \hbar^{-3} \sim p_F^3 \hbar^{-3}, \quad (222)$$

where we truncated the integral over momentum at the so-called ‘Fermi-momentum’  $p_F$ . Inverting this equation gives

$$\boxed{p_F \sim \hbar n^{1/3}} \quad (223)$$

This equation is important: it shows that for a degenerate gas, as we increase the number density, we add electrons with increasingly large momentum. This is a strange statement since momentum is associated with motion of particles, which is associated with temperature. Strangely, if you have taken a course on statistical physics, you know that the assumption that the phase-space density is maximal corresponds to having zero temperature! That is, we assumed zero-temperature gas, yet as we increase the density of this gas, an increasing fraction of electrons obtains larger momenta.

The Fermi-energy of the particles relates to momentum via  $E_F = \frac{p_F^2}{2m}$  in the non-relativistic regime. However, if we increase the density enough (via e.g. gravitational compression) there can be particles with large enough momenta that they are relativistic<sup>33</sup>. In this case  $E_F = p_F c$ .

Let us focus on the non-relativistic regime first. Consider a star of radius  $R$  that contains  $N$  particles (say  $N$  electrons and  $N$  protons). The number density of particles is then  $n \sim N/R^3$ . We can then compute the Fermi-energy per *electron* (which is higher than that of the protons) from

$$E_F = \frac{1}{2m_e} (\hbar n^{1/3})^2 = \frac{\hbar^2 N^{2/3}}{2m_e R^2} \sim \frac{\hbar^2 N^{2/3}}{m_e R^2}. \quad (224)$$

We compare this to the gravitational binding energy per electron which is

$$E_g \sim \frac{GMm_p}{R}, \quad (225)$$

where we used the mass of the *proton* instead of the mass per electron. If we were to ‘lift’ an electron out of the gravitational potential, then a small displacement of the electron would create an electric field, which would pull along the proton. The effective mass of the electron is then set by that of the proton.

The total energy of the electron is

$$E_{\text{tot}} = E_F + E_g = \frac{\hbar^2 N^{2/3}}{m_e R^2} - \frac{GMm_p}{R} = \frac{\hbar^2 N^{2/3}}{m_e R^2} - \frac{GNm_p^2}{R}. \quad (226)$$

Nature wants to minimise the energy of this electron, which sets a radius  $R_{\min}$  (we will look into this in the **assignment**).

One thing we can see is that as we increase  $N$  the total energy decreases (the **assignment** will make this clear as well): the objects becomes more compact and the density of electrons increases. This in turn increases the Fermi-momentum until the gas becomes relativistic. For relativistic gas, we have to write a different set of equations.

For a relativistic gas we have  $E_F = p_F c \sim \hbar n^{1/3} c \sim \hbar N^{1/3} c R^{-1}$ . The total energy of the electron is then

$$E_{\text{tot}} = \frac{\hbar N^{1/3} c}{R} - \frac{GMm_p}{R} = \frac{\hbar N^{1/3} c - GNm_p^2}{R}. \quad (227)$$

Minimizing the energy now has two possible implications: (*i*) if  $\hbar N^{1/3} c - GNm_p^2 > 0$  then minimising  $E_{\text{tot}}$  corresponds to increasing  $R$ . Degeneracy pressure of the relativistic gas pushes the star apart (possibly back into the non-relativistic regime [?]); (*ii*) if  $\hbar N^{1/3} c - GNm_p^2 < 0$  then minimising  $E_{\text{tot}}$  corresponds to decreasing  $R$ . As  $R$  decreases, the energy of the system decreases, and run-away collapse occurs. The condition  $\hbar N^{1/3} c - GNm_p^2 = 0$  therefore marks the onset

<sup>33</sup> In other words, formally we can have a relativistic degenerate gas with zero temperature.

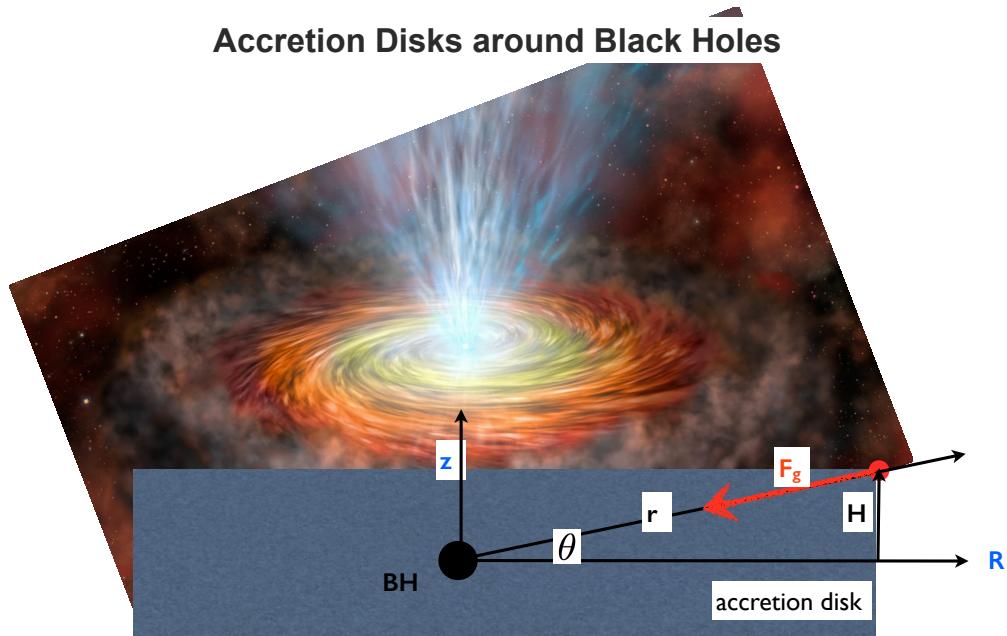
of the stability. Solving for  $N$  gives us the maximum number of particles which we can have in a degenerate gas

$$N_{\max} \sim \left( \frac{\hbar c}{G m_p^2} \right)^{3/2} \sim 2 \times 10^{57} \Rightarrow M_{\max} \sim 1.5 M_\odot. \quad (228)$$

This Maximum mass is the so-called Chandrasekhar mass. Note that  $p_F$  is identical for protons. However, protons have a higher mass, and at fixed momentum they have a lower velocity. The pressure exerted by particles generally scales as  $P \propto p v$ , where  $v$  denotes the velocity. The proton degeneracy pressure is therefore lower.

Further note that runaway collapse of the electron-degenerate gas actually does not lead to a black hole directly. At sufficiently high densities the free electrons are forced into the protons (this is known as inverse beta-decay) to make neutrons. We can then do the same calculation as we did here for a neutron degenerate gas. At the level of the discussion in class, this calculation is identical. A more detailed calculation -taking into account the ‘equation of state’ of nuclear matter - gives rise to the so-called ‘Tolman-Oppenheimer-Volkoff limit’, which is slightly different than what we derived. Objects above this limit would collapse into a black hole.

### C. Geometrically Thin Accretion Disks: Not Needed for Exam!



Consider a disk with half-thickness  $H$ .

Hydrostatic equilibrium [in cylindrical coordinates  $(z, R)$ ] implies that  $c_s = \left( \frac{H}{r} \right) v_{\text{circ}}$

Accretion disks are often depicted as geometrically thin. The goal of this calculation is to check what physical conditions this requires. The thickness of the disk is determined by pressure inside the disk balancing the gravitational pull by the black hole. We therefore consider the equation of hydrostatic equilibrium ( $\nabla p = -\rho_{\text{gas}} \nabla \phi$ ), and use cylindrical coordinates. We are interested in the thickness of the disk, and hence focus on the  $z$ -component (see slide 15.). Under the assumption that gravity is dominated by the black hole we have

$$\frac{dP}{dz} = -\rho_{\text{gas}} \frac{GM \sin \theta}{r^2}. \quad (229)$$

We approximate the pressure gradient in the  $z$ -direction with  $\frac{dP}{dz} = \frac{P(z=H) - P(z=0)}{H} \sim -\frac{P}{H}$ . We also know that  $\sin \theta = H/R$ , where  $H$  denotes the half-thickness of the disk (see slide 15.). The equation of hydrostatic equilibrium

becomes

$$\frac{P}{H} = \rho_{\text{gas}} \frac{GM}{r^2} \frac{H}{r} = \frac{\rho_{\text{gas}} H}{r^2} \frac{GM}{r}. \quad (230)$$

We know that  $\frac{GM}{r} = v_{\text{rot}}^2$ . Rearranging some terms gives us

$$\frac{P}{\rho_{\text{gas}}} = \left(\frac{H}{r}\right)^2 v_{\text{rot}}^2. \quad (231)$$

The term on the left-hand side  $\frac{P}{\rho_{\text{gas}}} \sim c_s^2$  (see our lectures on the Jeans mass<sup>34</sup>), and we therefore have

$c_s^2 = \left(\frac{H}{r}\right)^2 v_{\text{rot}}^2.$

(232)

This equation is important. In accretion disks the rotational velocity close to the black holes usually greatly exceeds the sound-speed, i.e  $c_s/v_{\text{rot}} = H/r \ll 1$ . This suggests that at any  $r$ , the half-thickness of the disk is much smaller than  $r$ . However, this analysis also implies that the geometrically thin disk approximation breaks down when the disk becomes very hot, which is the case very close to the BH (see next analysis), and/or for very high accretion rates (also see next analysis).

#### D. Temperature Structure in Geometrically Thin Accretion Disks.

The total energy per particle of mass  $m$  on a circular orbit around a BH is

$$E_{\text{tot}} = \frac{1}{2}mv_{\text{rot}}^2 - \frac{GMm}{r} = \frac{-GMm}{2r} \quad (233)$$

The change in the energy  $dE_{\text{tot}}$  of the particle as it moves inward by  $dr$

$$dE_{\text{tot}} = \frac{dE_{\text{tot}}}{dr} dr = \frac{GMmdr}{2r^2} \sim \frac{GMmdr}{r^2}, \quad (234)$$

where again I drop constants of order unity. The total luminosity  $dL$  from gas moving from  $r$  to  $r - dr$  is

$$dL = \frac{GM\dot{m}dr}{r^2}, \quad (235)$$

where  $\dot{m}$  denotes the total mass that is flowing through the shell at  $r$  per unit time. Now we assume that this luminosity comes out in black body radiation (which is the case if the gas is ‘optically thick’ to the radiation). Under this assumption  $dL = dA\sigma_{\text{SB}}T^4 = 2\pi r dr \sigma_{\text{SB}}T^4$ . Setting this equal to the equation above, and dropping constants of order unit we find

$$T = \left(\frac{GM\dot{m}}{r^3\sigma_{\text{SB}}}\right)^{1/4} \propto \left(\frac{M\dot{m}}{r^3}\right)^{1/4}. \quad (236)$$

This equation has several implications: (i) the temperature increases towards smaller  $r$ , and (ii) the temperature increases with mass accretion rate, and (iii) accretion disks become *hotter* around lower mass black holes  $M$ . To see this latter point, we evaluate the temperature at some fixed radius  $R$  in terms of the Schwarzschild Radius  $R_s$  of the black hole, i.e.  $R = xR_s$  (where we keep  $x$  fixed). We then know that at fixed  $x$   $r \propto R_s \propto M$ . We will see later that it is also reasonable to assume that the mass accretion rate is proportional to the BH mass  $M$ , i.e.  $\dot{m} \propto M$ . Under

<sup>34</sup> Formally  $c_s^2 = \frac{dP}{dp}$ . For an ideal gas this introduces an extra factor of order unit which we can ignore in the present analysis.

these assumptions we have

$$T(xR_s) \propto \left(\frac{M^2}{M^3}\right)^{1/4} \propto M^{-1/4}. \quad (237)$$

### E. Spectra Emerging from Accretion Disks: Not Needed for Exam!

With the temperature profile  $T(r, M)$  determined for an optically thick (but geometrically thin) accretion disk, we can compute the spectrum emerging from the accretion disk as a whole. We know that each shell emits as a black body, for which the spectrum is determined by its temperature,  $B(\nu, T[r])$ . Summing over all shells gives us the full spectrum. That is

$$L(\nu) = 2\pi \int_{R_{in}}^{R_{out}} dr r B(\nu, T[r]), \quad (238)$$

where  $R_{in}$  denotes the inner radius of the accretion disk, and  $R_{out}$  denotes the outer radius. We know the blackbody spectrum to be

$$B(\nu, T) \propto \nu^3 \left( \exp \left[ \frac{h\nu}{k_b T} \right] - 1 \right)^{-1} \quad (239)$$

Throughout the accretion disks there exists a one-to-one relation between radius  $r$  and temperature  $T$  as  $T \propto r^{-3/4}$  (see [slide 19](#)). When  $T \propto r^{-3/4}$  we have  $r \propto T^{4/3}$  and  $dr \propto T^{-1/3} d[1/T]$ . This yields (dropping all constants of proportionality):

$$L(\nu) \propto \nu^3 \int_{1/T_{in}}^{1/T_{out}} d[1/T] \frac{T^{-4/3} T^{-1/3}}{\exp(h\nu/[k_b T]) - 1}. \quad (240)$$

Substituting  $x = \frac{h\nu}{k_b T}$ , which means  $T \propto \nu x^{-1}$  and  $d[1/T] = \frac{k_b T}{h\nu} dx$ . We are then left with

$$L(\nu) \propto \nu^{1/3} \int_{h\nu/k_b T_{in}}^{h\nu/k_b T_{out}} dx x^{5/3} [\exp(x) - 1]^{-1}. \quad (241)$$

The gas has the highest temperature  $T_{in}$  at  $R_{in}$ . We expect the spectrum to be cut-off as  $L(\nu) \propto \nu^3 \exp \left[ -\frac{h\nu}{k_b T_{in}} \right]$  at  $h\nu \gg k_b T_{in}$ . On the other hand, the gas has the lowest temperature  $T_{out}$  at  $R_{out}$ . At  $h\nu \ll k_b T_{out}$ , we expect that  $L(\nu) \propto \nu^3 \left( \exp \left[ \frac{h\nu}{k_b T_{out}} \right] - 1 \right)^{-1} \propto \nu^2$  (because  $\exp(x) \sim 1 + x$  for small  $x$ ).

### F. The Innermost Stable Circular Orbit (ISCO)

In Newtonian gravity, a particle can be in any stable orbit around a black hole for<sup>35</sup>  $R > R_s$ . In general relativity however, particles cannot be in circular orbits with  $R_{ISCO} < 3R_s$ , where the precise numerical value of  $R_{ISCO}$  decreases with black hole spin (see lecture slides for the precise dependence). Interestingly, at  $R < R_{ISCO}$ , particles flow into the black hole without dissipating any energy. The ISCO therefore also determines how much energy is radiated by gas accreting onto a black hole (see § [XXI.G](#)).

<sup>35</sup> You could say that the maximum circular velocity of a particle is  $c$ . Then gravity balance the centrifugal acceleration when  $\frac{GM}{r^2} = \frac{c^2}{r}$ , and  $r = \frac{GM}{c^2}$  which is half the Schwarzschild radius, and is therefore not relevant.

## G. Radiative Efficiency $\eta$

The total energy of a proton on a circular orbit of radius  $R$  around a black hole of mass  $M$  is (see Eq 233)

$$E_{\text{tot}} = \frac{-GMm_p}{2R}. \quad (242)$$

In general relativity there is a minimum radius at which particles can be on circular orbits. This radius is referred to as the ‘ISCO’ radius (Innermost Stable Circular Orbit),  $R_{\text{ISCO}}$ . This is a general realistic effect. In Newtonian gravity there is no analogous radius. As we showed above, for a stationary (non-spinning) black hole  $a = 0$  and  $R_{\text{ISCO}} \sim \frac{6GM}{c^2}$  (see lecture slides how  $R_{\text{ISCO}}$  depends on black hole spin). The total energy of a proton orbiting at  $R_{\text{ISCO}}$  is therefore

$$E_{\text{tot}} = \frac{-GMm_p}{12GM/c^2} = -\frac{1}{12}mc^2 \sim -0.08m_pc^2. \quad (243)$$

That is, the total kinetic + binding energy of the proton is (minus) 8% of its rest-mass energy. It is common to express this total energy as

$$E_{\text{tot}} = -\eta mc^2. \quad (244)$$

The standard assumption is that  $\eta \sim 0.1$ , which corresponds to a stationary black hole. The parameters  $\eta$  is the ‘**radiative efficiency**’, as it indicates how much binding energy is radiated away. Finally, mass-energy equivalence states that this negative kinetic + binding energy corresponds to a negative mass of  $\eta m_p$ . If a proton finally is ‘absorbed’ by the black hole, it contributes a total of  $\Delta M_{\text{BH}} = (1 - \eta)m_p$  to the mass of the black hole.

## H. Derivation of Eddington Luminosity

As we increase the gas accretion rate  $\dot{m}$ , we increase the total luminosity if the accretion rate  $L$ . This radiation can exert a pressure on the in falling material. Consider gas at a distance  $r$  from the central black hole. The gravitational force on an electron equals  $a_{\text{grav}} = \frac{GMm_p}{r^2}$  (where I deliberately used the proton mass, see the discussion above). On the other, the electron experience a force outwards because of the incoming radiation. This force is given by

$$F_{\text{rad}} = \frac{dp_\gamma}{dt}, \quad (245)$$

in which  $\frac{dp_\gamma}{dt}$  denotes the rate at which transfer momentum to the electron. Each time an electron ‘absorbs’ a photon it absorbs its momentum, which is given by  $h\nu/c$ . Electrons do of course not absorb photons, but instead scatter them. The precise amount of momentum that each photon gives to the electron depends on the angle at which the photon is scattered. However, it can be shown that *on average* each scattering event transfers an amount  $h\nu/c$ .

The total force by the radiation on the electron is then

$$\frac{dp_\gamma}{dt} = \Gamma \frac{\langle h\nu \rangle}{c}, \quad (246)$$

where  $\Gamma$  denotes the total rate at which the electron scatters photons, and  $\langle h\nu \rangle$  denotes the average energy of the photon. We have compute scattering rates before. The scattering rate  $\Gamma$  (with units  $\text{s}^{-1}$ ) is related to the luminosity of the accretion disk as

$$\Gamma = \frac{\dot{N}\sigma_T}{4\pi r^2} = \frac{L\sigma_T}{4\pi r^2 \langle h\nu \rangle} \quad (247)$$

where  $\dot{N}$  is the total number of emitted photons. Substituting this into the Eq 246 gives

$$F_{\text{rad}} = \frac{dp_\gamma}{dt} = \frac{L\sigma_T}{4\pi r^2 c}, \quad (248)$$

This radiation pressure force balances gravity when

$$L = \frac{4\pi GMm_p c}{\sigma_T} \equiv L_{\text{edd}}. \quad (249)$$

which is the ‘Eddington luminosity’.

### I. Black Hole Mass Growth in the Eddington Limit

§ XXI.H contains a derivation of the Eddington luminosity  $L_{\text{edd}}$ , which is the maximum luminosity that accretion is allowed to generate, without stopping further accretion onto the black hole. This luminosity increases proportionally to the mass accretion rate  $\dot{m}$ , and the Eddington luminosity therefore has a corresponding maximum mass accretion rate. The goal of this section is to compute how the mass of a black hole grows, under the assumption that it accretes at the maximum rate possible.

Our starting point is that each particle of mass  $m$  contributes  $(1 - \eta)m_p$  to the mass of the black hole. In other words, we can write

$$\frac{dM_{\text{BH}}}{dt} = (1 - \eta)\dot{m}, \quad (250)$$

where  $\dot{m}$  denotes the *total mass accretion rate* onto the black hole. We also know that each particle of mass  $m$  needs to radiate away  $\eta mc^2$  to reach the ISCO,  $R_{\text{ISCO}}$ . After the particle reaches the ISCO it is dragged into the black hole without emitting any more energy. The total luminosity that is generated by mass accreting at some rate  $\dot{m}$  is therefore  $L = \eta\dot{m}c^2$ , i.e.  $\dot{m} = L/[\eta c^2]$ . Substituting this into Eq 250 we have

$$\frac{dM_{\text{BH}}}{dt} = \frac{1 - \eta}{\eta} \frac{L}{c^2}. \quad (251)$$

We wanted to know the accretion rate for the maximum mass accretion rate, which had a corresponding Eddington luminosity. If we replace  $L$  with  $L_{\text{edd}}$ , we get a differential expression for the mass growth rate for the maximum accretion rate:

$$\frac{dM_{\text{BH}}}{dt} = \frac{1 - \eta}{\eta} \frac{L_{\text{edd}}}{c^2} = \frac{1 - \eta}{\eta} \frac{4\pi GMm_p}{c\sigma_T} \equiv \frac{1 - \eta}{\eta} \frac{M_{\text{BH}}}{t_E}, \quad (252)$$

where we substituted the expression for  $L_{\text{edd}}$ , and where we defined the ‘Eddington time’  $t_E \equiv c\sigma_T/[4\pi Gm_p] \sim 0.45$  Gyr. Eq 252 is a first order differential equation with the solution

$$M_{\text{BH}}(t) = M_{\text{BH}}(t = 0) \exp\left(\frac{1 - \eta}{\eta} \frac{t}{t_E}\right) \sim M_{\text{BH}}(t = 0) \exp\left(\frac{t}{t_s}\right) \quad (253)$$

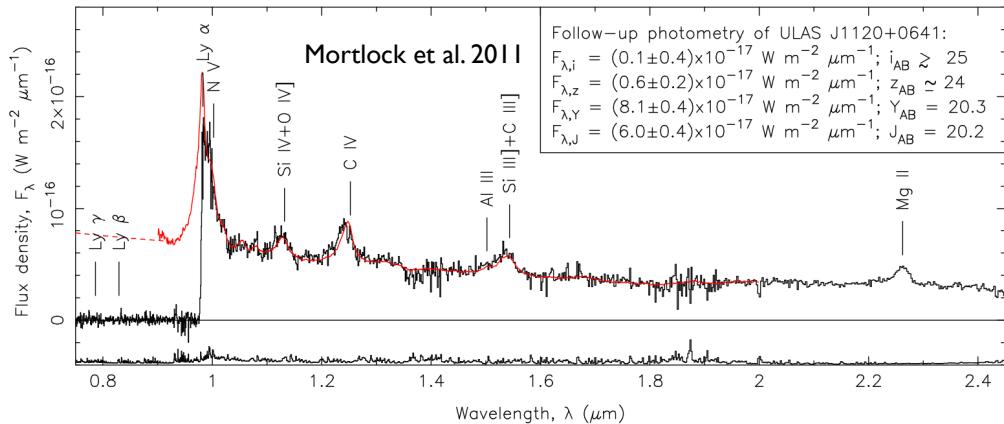
where  $t_s \sim 50$  Myr. This is an important result. A black hole that is accreting at its maximum possible rate, increases its mass by a factor of  $e$  over  $t_s \sim 50$  Myr. We used this to argue that the existence of super massive black holes of mass  $M_{\text{BH}} \sim 10^9 M_\odot$  at  $z \sim 7.1$  [when the Universe was only 0.77 Gyr old!] is difficult to explain by having stellar mass black holes grow by  $\sim 7 - 8$  orders of magnitude via the maximum allowed accretion rate over its entire life time. Fig XXI.I shows something equally puzzling, namely that the spectra of accreting super massive black holes at  $z \sim 7.1$  are virtually identical to that of their lower redshift counterpart.

## XXII. USEFUL READING

Useful reading material includes

- The derivation of the isothermal density profile for the dark matter follows Binney & Tremaine ‘*Galactic Dynamics*’ (page 226-228).

## High-z vs Low-z Quasars



- Reading material: suggested review article by Weinberg et al (<https://arxiv.org/abs/1306.0913>), as mentioned in the slides provides a good, brief summary of both missing satellite and too-big-to-fail problems.

**Acknowledgments** A special thanks to

### References

- Agertz, O., Teyssier, R., & Moore, B. 2009, MNRAS, 397, L64  
 Peter, A. H. G. 2012, arXiv:1201.3942  
 Planck Collaboration, Adam, R., Aghanim, N., et al. 2016, arXiv:1605.03507  
 Vogelsberger, M., Zavala, J., & Loeb, A. 2012, MNRAS, 423, 3740  
 Weinberg, D. H., Bullock, J. S., Governato, F., Kuzio de Naray, R., & Peter, A. H. G. 2015, Proceedings of the National Academy of Science, 112, 12249