8.902: Astrophysics II

Massachusetts Institute of Technology
Department of Physics

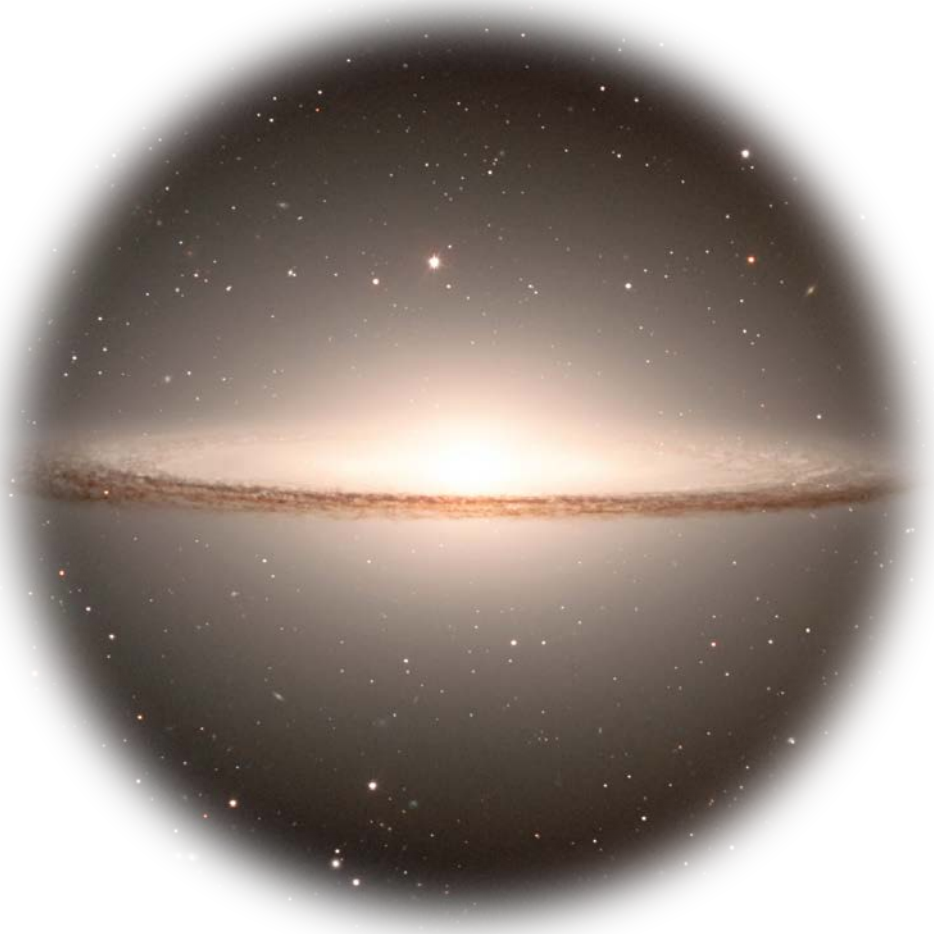
Lecture notes by
Mark Vogelsberger, Stephanie O'Neil, & David DePalma

Fall 2023

Contents

Part I

Galaxies



Messier 104, Sombrero Galaxy. Credit: ESO/P. Barthel (Kapteyn Institute, Groningen). License CC-BY.

1 Key observations of galaxies

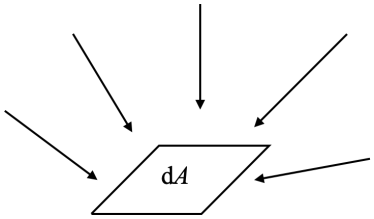
1.A Basic units of radiative transfer

We first define some fundamental quantities.

Flux:

$$\boxed{F_\nu = \frac{\mathrm{d}E_\nu}{\mathrm{d}A\ \mathrm{d}t\ \mathrm{d}\nu}} \tag{1}$$

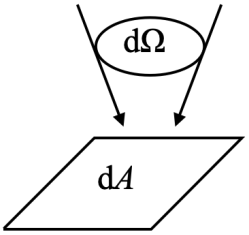
with units $[F_\nu] = \mathrm{erg\ s^{-1} cm^{-2} Hz^{-1}}$.
F_ν is the flux at a specific frequency ν.



Specific intensity:

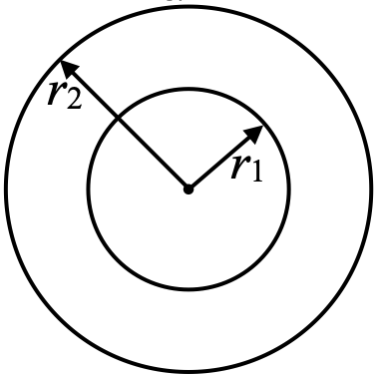
$$\boxed{I_\nu = \frac{\mathrm{d}E_\nu}{\mathrm{d}A\ \mathrm{d}t\ \mathrm{d}\nu\ \mathrm{d}\Omega}} \tag{2}$$

with units $[I_\nu] = \mathrm{erg\ s^{-1} cm^{-2} Hz^{-1} sr^{-1}}$.
This is the flux per solid angle.



Note:

- $F_\nu(r) \propto \frac{1}{r^2}$.
  Due to energy conservation:



$$\begin{aligned} &F_\nu(r_1) \times 4\pi r_1^2 = \\ &(\mathrm{d}E_\nu)_1 = (\mathrm{d}E_\nu)_2 = \\ &F_\nu(r_2)^2 \times 4\pi r_2^2 \end{aligned} \tag{3}$$

$$\Rightarrow \frac{F_\nu(r_1)}{F_\nu(r_2)} = \left(\frac{r_2}{r_1}\right)^2 \tag{4}$$

$$\Rightarrow F_\nu(r) \propto \frac{1}{r^2}$$

- $I_\nu(r) \propto$ constant because:

$$\begin{aligned} &I_\nu = \frac{F_\nu}{\mathrm{d}\Omega} \text{ and } \mathrm{d}\Omega \propto \frac{1}{r^2} \\ &\Rightarrow F_\nu \propto \frac{1}{r^2} \text{ and } \mathrm{d}\Omega \propto \frac{1}{r^2} \\ &\Rightarrow I_\nu \propto \text{constant} \end{aligned} \tag{5}$$

Magnitude scale:
Define the *apparent magnitude* m, i.e. how bright an object appears:

$$m_1 - m_2 = -2.5 \log \left(\frac{(F_\nu)_1}{(F_\nu)_2} \right) \tag{6}$$

With this definition, a brighter object has a lower magnitude. There are two main magnitude systems: Vega and AB.

The *Vega system* is calibrated using the flux of the AO V star Vega $(F_\nu)_{\text{Vega}}$, which has a non-flat distribution (flux changes for different frequencies). The *AB system* is calibrated to a hypothetical source with flux

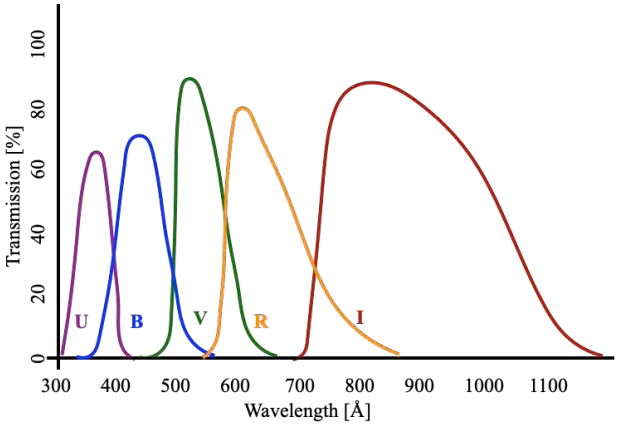
$$(F_\nu)_{\text{AB}} = 3.63 \times 10^{-20} \text{erg s}^{-1}\text{cm}^{-2}\text{Hz}^{-1} \tag{7}$$

which has a flat distribution.

We also have the *monochromatic magnitude*, i.e. the magnitude at a single wavelength, defined for each system:

$$\begin{aligned} \text{Vega} : m_\nu &= -2.5 \log \left(\frac{F_\nu}{(F_\nu)_{\text{Vega}}} \right) \\ \text{AB} : m_\nu &= -2.5 \log \left(\frac{F_\nu}{(F_\nu)_{\text{AB}}} \right) \end{aligned} \tag{8}$$

A more practical quantity is the *band magnitude*. In most observations, the fluxes are integrated over a filter bandpass with a transmission function $T_X(\nu)$ for band X. An example of a set of filters (U,B,V,R,I) and the transmission function (what percent of the flux is let through at a given frequency or wavelength) is shown to the right.

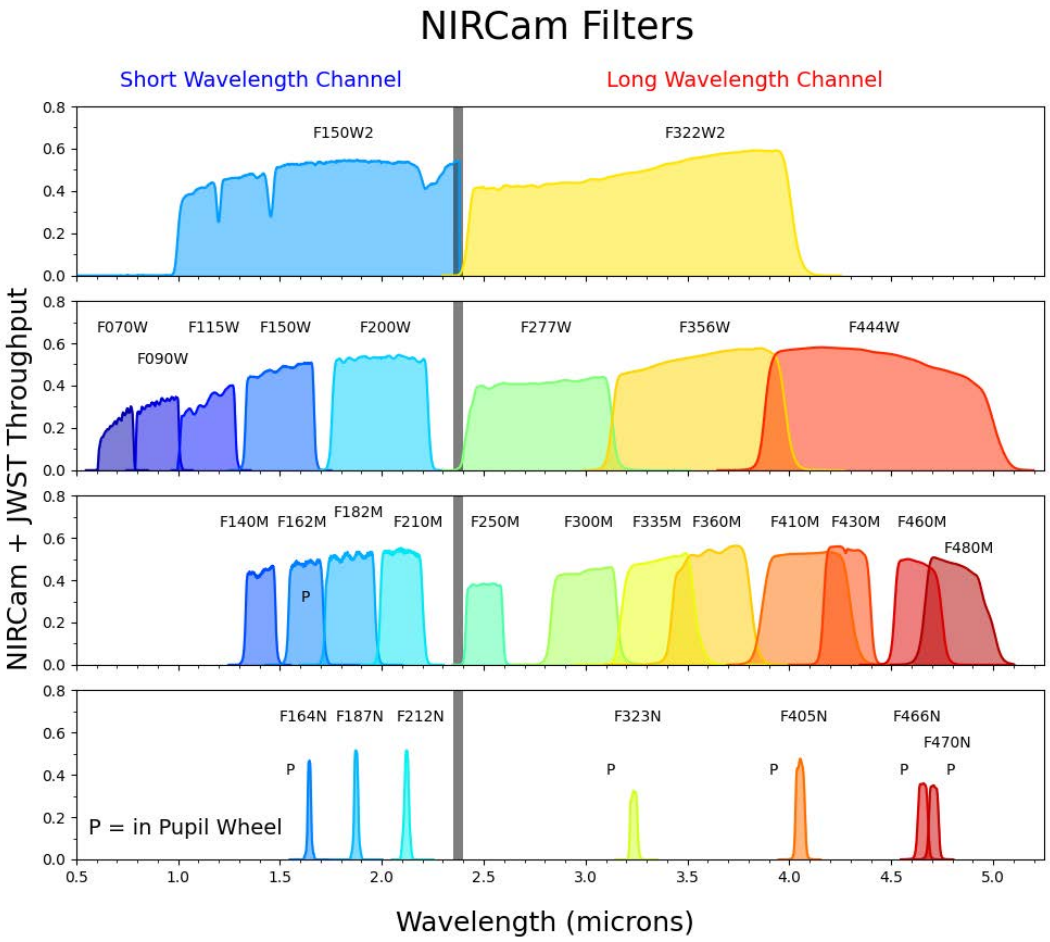


$$\begin{aligned} \text{Vega} : m_X &= -2.5 \log \left(\frac{\int F_\nu T_X(\nu) \mathrm{d}\nu}{\int (F_\nu)_{\text{Vega}} T_X(\nu) \mathrm{d}\nu} \right) \\ \text{AB} : m_X &= -2.5 \log \left(\frac{\int F_\nu T_X(\nu) \mathrm{d}\nu}{\int (F_\nu)_{\text{AB}} T_X(\nu) \mathrm{d}\nu} \right) \end{aligned} \tag{9}$$

$(F_\nu)_{\text{AB}} =$ constant and $\int T_X(\nu) \mathrm{d}\nu = 1$, so

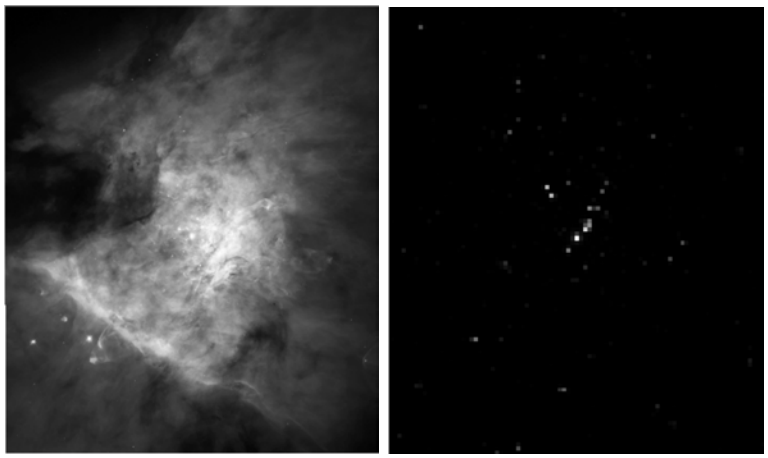
$$\int (F_\nu)_{\text{AB}} T_X(\nu) \mathrm{d}\nu = (F_\nu)_{\text{AB}} \, . \tag{10}$$

Telescopes like Hubble and SDSS observe primarily in the visible light spectrum. JWST measures slightly longer wavelengths and is sensitive to the infrared range. The NIRCam instrument filters are shown in below. We show the total throughput (photon-to-election conversion efficiency) for extra-wide, wide, medium, and narrow filters for NIRCam (image from https://jwst-docs.stsci.edu).



Each filter measures a different energy range of electromagnetic waves and therefore probes different physics. As an example of this, we show the Orion Nebula as viewed in visible light from Hubble below on the left and in X-ray from Chandra on the right. In the visible range, we can see the diffuse gas while in the X-ray, we can see point-like sources from stars.

Figure is in the public domain. JWST User Documentation (JDox). Baltimore, MD: Space Telescope Science Institute; 2016-2024-07-25. https://jwst-docs.stsci.edu



For all filters, $-2.5\log(3.63\times10^{-20}) = 48.6$, so (for the AB system)

$$\begin{aligned} m_X &= -2.5\log\left(\int F_\nu T_X(\nu)\mathrm{d}\nu\right) - 48.6 \\ m_\nu &= -2.5\log(F_\nu) - 48.6\ . \end{aligned} \tag{11}$$

The value for $(F_\nu)_{\rm AB}$ was chosen such that $m_V(\mathrm{AB}) = m_V(\mathrm{Vega})$ and they have the same magnitude in the V-band. For other bands, one must apply the conversion

$$m_X(\mathrm{AB}) - m_X(\mathrm{Vega}) = -2.5\log\left(\frac{\int (F_\nu)_{\rm AB} T_X(\nu)\mathrm{d}\nu}{(F_\nu)_{\rm Vega} T_X(\nu)\mathrm{d}\nu}\right)\ . \tag{12}$$

This gives us, for example:

$$\begin{aligned} U_{\rm AB} &= U_{\rm Vega} - 0.8 \\ B_{\rm AB} &= B_{\rm Vega} - 0.11 \\ V_{\rm AB} &= V_{\rm Vega} \end{aligned} \tag{13}$$

Be careful which magnitude is quoted! SDSS uses u, g, r, i, z filters.

Define the *absolute magnitude* as the apparent magnitude if the object were at a distance of 10 pc. Apparent magnitude depends on both the brightness of the object and its distance. Absolute magnitude is related to the intrinsic brightness of the object.

$$m_X - M_X = 5\log\left(\frac{D}{10\,\mathrm{pc}}\right) \equiv \mu \tag{14}$$

Image of Orion Nebula created by Mark Vogelsberger using SAOImage DS9 image display and visualization tool for astronomical data. https://sites.google.com/cfa.harvard.edu/saoimageds9/home

Since

$$
\begin{aligned}
F_{\rm app} = \frac{L}{4\pi D^2} \qquad \text{and} \qquad F_{\rm abs} &= \frac{L}{4\pi (10\,{\rm pc})^2} \\
\Rightarrow \quad m_X &= -2.5 \log\left(\frac{L}{4\pi D^2}\right) + \text{constant} \\
M_X &= -2.5 \log\left(\frac{L}{4\pi (10\,{\rm pc})^2}\right) + \text{constant} \\
\Rightarrow \quad m_X - M_X &= +2.5\log(D^2) - 2.5\log((10\,{\rm pc})^2) \\
&= 5\log(D) - 5\log(10\,{\rm pc}) \\
&= 5\log(D/{\rm pc}) - 5
\end{aligned}
\tag{15}
$$

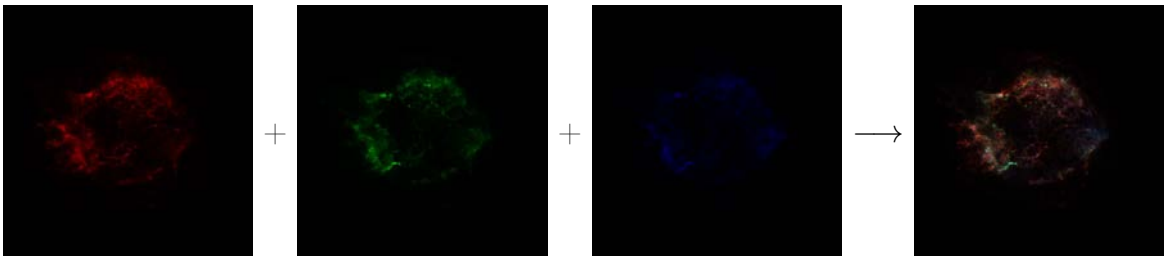
$\mu = m_X - M_X$ is the *distance modulus* (a measure of distance).

Colors:
If observations are made in more than one filter (X, Y), then one can define a color as the difference in magnitudes between the two bands:

$$(X - Y) = m_X - m_Y = M_X - M_Y \tag{16}$$

Stars and galaxies can be "red" or "blue", for example. It is common to use the difference between g and r filters to get $g - r$ color. A higher $g - r$ value is red and a lower value is blue. Note that higher $g - r$ has a higher g relative to r, but a higher magnitude is less bright.

We can take images of the same object through different filters and combine them for a more complete view of the object. Here we show images of the supernova remnant Cassiopeia A taken in three wavelength ranges (0.6-1.65 keV, 1.67-2.25 keV, and 2.25-7.5 keV) shown in red, green, and blue and then combined into a single image.



Surface brightness:
We measure the luminosity ([erg s^{-1}]) per area. This is often called Σ or I. It effectively measures the magnitude per square arcsecond:

$$M \propto -2.5\log(I)\ . \tag{17}$$

Spiral and elliptical galaxies show different surface density profiles:

$$
\begin{aligned}
\text{exponential}: I(r) &= I_0 e^{-r/r_s} && \text{(spirals)} \\
\text{de Vaucouleurs}: I(r) &= I_0 e^{-7.67(r/r_e)^{\frac{1}{4}}} && \text{(ellipticals)}
\end{aligned}
\tag{18}
$$

Images of supernova remnant created by Mark Vogelsberger using SAOImage DS9 image display and visualization tool for astronomical data. https://sites.google.com/cfa.harvard.edu/saoimageds9/home

More generally, we have the Sérsic profile with Sérsic index n:

$$I(r) = I_0 e^{-(r/r_0)^{\frac{1}{n}}} \tag{19}$$

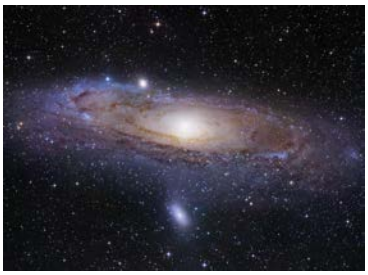
Observationally, $n \approx 4$ for ellipticals and $n \approx 1$ for spirals. Theory needs to explain this!

1.B Basic properties of the galaxy population

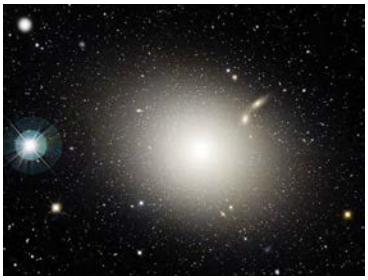
Types of galaxies:
Images of galaxies show mainly three types:

- Spirals (Sp):
  - dominate in the field (outside clusters)
  - disks with gas and stars
  - young stellar population
  - rotationally supported
  - blue color
  - exponential surface brightness profile
- Ellipticals (E):
  - cluster environment
  - spheroidal
  - old stellar population
  - pressure supported
  - red color
  - de Vaucouleurs surface brightness profile
- Lenticular (SO):
  - stellar disk
  - no gas disk
  - link between spiral and elliptical galaxies

Andromeda M31: Courtesy of Robert Gensler. Used with permission.



Andromeda



M87

NGC 2787

Galaxy luminosity distribution:
The luminosity L of an object is

$$L = \frac{\mathrm{d}E}{dt} = \int I_\nu \mathrm{d}A d\Omega \mathrm{d}\nu \ . \tag{20}$$

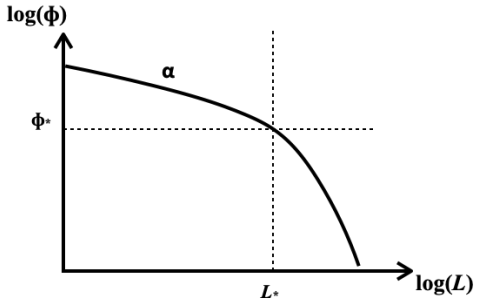
M87 Image: Courtesy of Canada-France-Hawaii Telescope / Coelum. Used with permission.

NGC 2787 Image: NASA and The Hubble Heritage Team (STScI/AURA); Acknowledgment: M. Carollo (Swiss Federal Institute of Technology, Zurich)

What is the distribution function of L for galaxies? We commonly use the *Schechter function* to describe the number density of galaxies at a given luminosity:

$$\phi(L)\mathrm{d}L = \phi_* \left(\frac{L}{L_*}\right)^{\alpha} e^{-L/L_*} \frac{\mathrm{d}L}{L_*} \tag{21}$$

ϕ_*: normalization	$\phi_* \approx 0.02\, h^3 \mathrm{Mpc}^{-3}$
α: faint-end slope	$\alpha \approx -1.09$
L_*: characteristic L at the normalization point	$L_* \approx 10^{10}\, L_\odot h^{-2}$



Velocity structure of galaxies:
Spectral data of galaxies allows us to measure velocities. Spiral galaxies have ordered, circular motion with $V_c \sim 200 \pm 50\,\mathrm{km/s}$. We can measure the circular velocity through the motions of stars and, further out, from the spectral lines of gas. Outside the galaxy, one finds that v_c remains constant, but one would expect:

$$\frac{mv_c^2}{r} = \frac{GMm}{r^2} \tag{22}$$

for a circular orbit. This implies $v_c \propto r^{-1/2}$, for centralized mass, which is not constant. To have v_c constant, we need

$$\begin{aligned} v_c &\propto \frac{1}{r}\int_0^r 4\pi r^2 \rho(r) dr \\ \Rightarrow \rho(r) &\propto r^{-2} \end{aligned} \tag{23}$$

to large radii. This was one of the first hints for dark matter.
What is dark matter? A few things we know:

- It can't be non-luminous gas since we would have seen it through absorption lines
- Dim stars or other dense objects at larger distances (MACHOS: Massive Compact Halo Object) have been ruled out since microlensing (the temporary brightening of a distant object due to a closer massive object bending the light rays closer together) does not occur frequently enough
- Neutrinos have been ruled out since they lead to the wrong structure formation because they move so fast (hot dark matter). Since neutrinos move close to the speed of light, they have too much kinetic energy to be bound in low-mass potential wells.

- It could possibly be WIMPs (Weakly Interacting Massive Particles). However, there are no detections of WIMPs so far.
- General theories:
  - Cold Dark Matter (CDM): dark matter is a particle that moves slowly ($v \ll c$) and is collisionless, interacting solely through gravity.
  - Self-interacting dark matter (SIDM): dark matter interacts through gravity as well as through self-interactions that allow particles to scatter and transfer energy and momentum.
  - Warm dark matter (WDM): dark matter is still collisionless but moves with a faster velocity than CDM, which makes it harder to form less massive halos.
  - Bose-Einstein condensate (very low mass) dark matter: dark matter particles are very low mass such that their de Broglie wavelength is on the length scale of galaxies and leads to interference patterns in halos.
  - Modified Newtonian dynamics (MOND): Dark matter is not a type of matter but is accounted for through modifications to our theory of gravity.

Elliptical galaxies have a random motion velocity structure with velocity dispersion $\sigma_v \sim 200 - 300\,\mathrm{km/s}$. There is negligible circular motion, typically $v_c \sim 50 - 100\,\mathrm{km/s}$.

Spectra can also be used to measure redshift/recession velocity z of galaxies:

$$z = \frac{\lambda_{\mathrm{obs}} - \lambda_0}{\lambda_0} \qquad \text{or} \qquad 1 + z = \frac{\lambda_{\mathrm{obs}}}{\lambda_0} \tag{24}$$

For low z ($z \ll 1$), one finds that the distance d is related to the redshift through the present-day Hubble constant H_0:

$$d \approx \frac{cz}{H_0}\,, \tag{25}$$

which yields

$$\boxed{v = H_0 d \approx cz\,, \quad z \ll 1} \tag{26}$$

The redshift directly yields the recession velocity. A more formal proof of $d \approx \frac{cz}{H_0}$ will be discussed later (low z limit for all distances).

Note: h is defined so the Hubble constant today is $H_0 = 100\,h$ km/s/Mpc

1.C Stellar population synthesis

So far, we have used spectral information only to derive velocities. However, we can also use this information to derive the *spectral energy distribution* (SED) of a galaxy. Stellar SEDs are blackbodies with different temperatures. The types of stars are referred to as O, B, A, F, G, K, M, L, and T, each with different temperatures that contribute differently to the

spectrum of the galaxy. Galaxies are a combination of these, so the total flux at a given frequency is a combination of the flux from each star:

$$F_\nu = N_O F_{\nu,O} + N_B F_{\nu,B} + ... \tag{27}$$

Determining $N_O, N_B, N_A, N_F...$ is the basic idea of stellar population synthesis.

2 Structure and a qualitative picture of galaxies

Goal: look at the most basic dynamical properties of a galaxy.

A *galaxy* is a collisionless fluid of stars and dark matter orbiting together with collisional gas in a common self-gravitational potential.

With this definition, we can try to understand the main dynamical properties.

2.A Virial Theorem

Derivation: assume stars orbit in a galaxy with mass, position, and velocity $(m_i, \vec{r}_i, \vec{v}_i)$. We then define the virial G:

$$G = \sum_i \vec{p}_i \cdot \vec{r}_i \tag{28}$$

which we can rewrite:

$$G = \sum_i \left(m_i \frac{d\vec{r}_i}{dt} \right) \cdot \vec{r}_i \, . \tag{29}$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}t}(\vec{r} \cdot \vec{r}) = \dot{\vec{r}} \cdot \vec{r} + \vec{r} \cdot \dot{\vec{r}} = 2\dot{\vec{r}}\,\vec{r}\,, \tag{30}$$

we get

$$\begin{aligned} G &= \frac{1}{2} \sum_i m_i \frac{\mathrm{d}}{\mathrm{d}t}(\vec{r}_i \cdot \vec{r}_i) \\ &= \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \sum_i m_i r_i^2 \, . \end{aligned} \tag{31}$$

Defining $I = \sum_i m_i r_i^2$ as the moment of inertia about the origin, we get

$$\boxed{G = \frac{1}{2} \frac{\mathrm{d}I}{\mathrm{d}t}} \, . \tag{32}$$

Now consider the time derivative of G:

$$\begin{aligned} \frac{\mathrm{d}G}{\mathrm{d}t} &= \sum_i \dot{\vec{p}}_i \cdot \vec{r}_i + \sum_i \vec{p} \cdot \dot{\vec{r}}_i \\ &= \sum_i \vec{F}_i \cdot \vec{r}_i + \sum_i m_i v_i^2 \\ &= \sum_i \vec{F}_i \cdot \vec{r}_i + 2T \end{aligned} \tag{33}$$

where T is the kinetic energy. Because gravity is a pairwise force, we can write

$$\vec{F}_k = \sum_{i=1}^{N} \vec{F}_{jk} \ . \tag{34}$$

$F_{ii} = 0$ and $1 \leq j \leq N$, so we can split F_{jk} into two parts, the upper and lower portions of the matrix

$$F_{jk} = \begin{array}{c} k\downarrow \ j\rightarrow \\ \\ \\ \end{array} \begin{pmatrix} 0 & & \textcircled{2} \\ & \ddots & \\ \textcircled{1} & & 0 \end{pmatrix} \tag{35}$$

with

$$\textcircled{1} = \sum_{k=2}^{N} \sum_{j=1}^{k-1} \vec{F}_{jk} \cdot \vec{r}_k \qquad \text{and} \qquad \textcircled{2} = \sum_{k=1}^{N-1} \sum_{j=k+1}^{N} \vec{F}_{jk} \cdot \vec{r}_k \tag{36}$$

so

$$\sum_{k=1}^{N} \vec{F}_k \cdot \vec{r}_k = \sum_{k=2}^{N} \sum_{j=1}^{k-1} \vec{F}_{jk} \cdot \vec{r}_k + \sum_{k=1}^{N-1} \sum_{j=k+1}^{N} \vec{F}_{jk} \cdot \vec{r}_k \ . \tag{37}$$

F is pairwise, so $-\vec{F}_{kj} = \vec{F}_{jk}$, which gives

$$\sum_{k=1}^{N} \vec{F}_k \cdot \vec{r}_k = \sum_{k=2}^{N} \sum_{j=1}^{k-1} \vec{F}_{jk} \cdot \vec{r}_k \ - \sum_{k=1}^{N-1} \sum_{j=k+1}^{N} \vec{F}_{kj} \cdot \vec{r}_k \ . \tag{38}$$

The second term in the above equation can be rewritten:

$$\sum_{k=1}^{N-1} \sum_{j=k+1}^{N} \vec{F}_{kj} \cdot \vec{r}_k = \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} \vec{F}_{jk} \cdot \vec{r}_j = \sum_{k=2}^{N} \sum_{j=1}^{k-1} \vec{F}_{jk} \cdot \vec{r}_j \tag{39}$$

which has the same matrix elements as the first term, so we get

$$\sum_{k=1}^{N} \vec{F}_k \cdot \vec{r}_k = \sum_{k=2}^{N} \sum_{j=1}^{k-1} \vec{F}_{jk} \cdot (\vec{r}_k - \vec{r}_j) \ . \tag{40}$$

We now assume that there is a potential V such that:

$$\begin{aligned} \vec{F}_{jk} &= -\nabla_k V \left(|\vec{r}_{jk}| \right) = -\nabla_k V(r_{jk}) \\ &= -\frac{\mathrm{d}V}{\mathrm{d}r} \left(\frac{\vec{r}_k - \vec{r}_j}{r_{jk}} \right) \end{aligned} \tag{41}$$

so we get

$$\begin{aligned}\sum_{k=1}^{N}\vec{F}_k\cdot\vec{r}_k &= \sum_{k=2}^{N}\sum_{j=1}^{k-1}\vec{F}_{jk}\cdot(\vec{r}_k-\vec{r}_j)\\ &= -\sum_{k=2}^{N}\sum_{j=1}^{k-1}\frac{\mathrm{d}V}{\mathrm{d}r}\frac{|\vec{r}_k-\vec{r}_j|^2}{r_{jk}}\\ &= -\sum_{k=2}^{N}\sum_{j=1}^{k-1}\frac{\mathrm{d}V}{\mathrm{d}r}r_{jk}\ .\end{aligned} \tag{42}$$

We now assume the special case $V(r_{jk}) = \alpha r_{jk}^n$. This gives us

$$\begin{aligned}\frac{\mathrm{d}V}{\mathrm{d}r} &= n\alpha r_{jk}^{n-1}\\ \Rightarrow\frac{\mathrm{d}V}{\mathrm{d}r}r_{jk} &= nV\end{aligned} \tag{43}$$

so

$$\begin{aligned}\sum_{k=1}^{N}\vec{F}_k\cdot\vec{r}_k &= -\sum_{k=2}^{N}\sum_{j=1}^{k-1}nV(r_{jk})\\ &= -n\sum_{k-2}^{N}\sum_{j=1}^{k-1}V(r_{jk})\\ &= -nV_{\mathrm{tot}}\ .\end{aligned} \tag{44}$$

Finally:

$$\frac{\mathrm{d}G}{\mathrm{d}t} = \sum_i \vec{I}_i\cdot\vec{r}_i + 2T = 2T - nV_{\mathrm{tot}} \tag{45}$$

With $\frac{dG}{dt} = \frac{1}{2}\frac{\mathrm{d}^2I}{\mathrm{d}t^2}, U = V_{\mathrm{tot}}$, and $n = -1$ (for gravity):

$$\boxed{\frac{1}{2}\frac{\mathrm{d}^2I}{\mathrm{d}t^2} = 2T + U}\ . \tag{46}$$

We now take the time average:

$$\begin{aligned}\left\langle\frac{\mathrm{d}G}{\mathrm{d}t}\right\rangle_{\mathcal{T}} &= \frac{1}{\mathcal{T}}\int_0^{\mathcal{T}}\frac{\mathrm{d}G}{\mathrm{d}t}\mathrm{d}t = \frac{G(\mathcal{T})-G(0)}{\mathcal{T}}\\ \Rightarrow\left\langle\frac{\mathrm{d}G}{\mathrm{d}t}\right\rangle_{\mathcal{T}} &= 2\langle T\rangle_{\mathcal{T}} - n\langle V_{\mathrm{tot}}\rangle_{\mathcal{T}}\end{aligned} \tag{47}$$

For a steady state system and long time average, $\frac{G(T)-G(0)}{\mathcal{T}} \approx 0$, so we get

$$\boxed{\begin{aligned}0 &= 2\langle T\rangle_{\mathcal{T}} - n\langle V_{\mathrm{tot}}\rangle_{\mathcal{T}}\\ 0 &= 2\langle T\rangle_{\mathcal{T}} + \langle U\rangle_{\mathcal{T}} \qquad \text{for } n=-1\end{aligned}}\ . \tag{48}$$

This is the *virial theorem*, often written simply as $0 = 2T + U$.

Note the three important assumptions for the virial theorem to hold:

- F is a pairwise force
- The potential V has the form $V \propto r^n$
- We have a steady state time averaged quantity $\frac{d^2I}{dt^2} = 0$.

Applications:
We now apply the virial theorem to a galaxy.

$$2T + U = 0 \,. \tag{49}$$

Assuming that a galaxy is made of N stars all with the same mass m (so total mass $M = Nm$) and average velocity $\vec{v}$, we get a total kinetic energy for the system

$$T = \frac{1}{2}\sum_i m_i v_i^2 \approx \frac{1}{2}M\bar{v}^2 = \frac{1}{2}Mv^2 \,. \tag{50}$$

From dimensional analysis for a galaxy of size R, we get a total potential energy

$$U = -\frac{GM^2}{R} \,. \tag{51}$$

The virial theorem then implies

$$\begin{aligned} Mv^2 + \left(-\frac{GM^2}{R}\right) &= 0 \\ \Rightarrow v &= \sqrt{\frac{GM}{R}} \,. \end{aligned} \tag{52}$$

Using some typical numbers:
$R \approx 10\text{kpc}$ $\quad M_\odot = 2 \times 10^{33}\text{g}$
$M \approx 10^{11} M_\odot$ $\quad G = 0.0043 M_\odot^{-1}\text{pc}\left(\frac{\text{km}}{\text{s}}\right)^2$

$$\begin{aligned} v &= \sqrt{\frac{0.0043 M_\odot^{-1}\text{pc}(\text{km/s})^2 \; 10^{11} M_\odot}{10\,000\text{pc}}} \\ &\approx \sqrt{4 \times 10^4}\,\text{km/s} \approx 200\,\text{km/s} \end{aligned} \tag{53}$$

which is in good agreement with observations.

We can also use the virial to get the ideal gas law.
For an ideal gas with N particles at temperature T is

$$K = \frac{3}{2}NkT \tag{54}$$

where we use K for kinetic energy to differentiate from temperature and k is the Boltzmann constant. The force comes from the pressure from the particles, so the force per unit area is

$$\mathrm{d}\vec{F} = -P\mathrm{d}\vec{A} \,. \tag{55}$$

Then the potential energy is

$$-\frac{1}{2}\left\langle \sum_i \vec{F}_i \cdot \vec{r} \right\rangle = \frac{P}{2}\int \vec{r}_i \cdot \mathrm{d}\vec{A} \tag{56}$$

and $n = 2$. From the divergence theorem

$$\begin{aligned} \int \vec{r}_i \cdot \mathrm{d}\vec{A} &= \int \vec{\nabla} \cdot \vec{r} \mathrm{d}V \\ &= 3\int \mathrm{d}V = 3V \ . \end{aligned} \tag{57}$$

The virial theorem gives

$$\begin{aligned} 0 &= 2\left\langle K \right\rangle_{\mathcal{T}} - 2\left\langle V_{\mathrm{tot}} \right\rangle \\ &= 2\frac{3}{2}NkT - 2\frac{P}{2}3V \end{aligned} \tag{58}$$

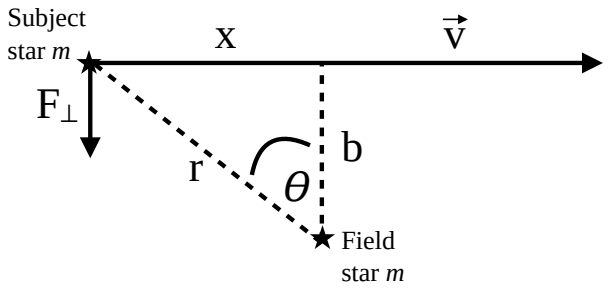
so

$$NkT = PV \ . \tag{59}$$

2.B Relaxation times

The virial theorem gave us some first insight into the dynamics of galaxies. Now we will show that stars are collisionless, i.e. that two-body collisions are rare in galaxies. Since this is true, we can describe the distribution of stars as a smooth density field and gravitational potential.

Frequency of strong encounters between stars:
Goal: estimate the change in velocity $\delta\vec{v}$ by which the encounter deflects the velocity $\vec{v}$ of the subject star.



We assume that $|\delta\vec{v}|/|\vec{v}| \ll 1$ and that the field star is stationary. This means that $\delta\vec{v}$ is perpendicular to $\vec{v}$ since the accelerations parallel to $\vec{v}$ cancel out as the subject star passes by the field star. We calculate $\delta v = |\delta\vec{v}|$ by integrating $F_\perp$:

$$F_\perp = \frac{Gm^2}{b^2+x^2}\cos\theta = \frac{Gm^2 b}{(b^2+x^2)^{3/2}} = \frac{Gm^2}{b^2}\left[1+\left(\frac{vt}{b}\right)^2\right]^{-3/2} \ . \tag{60}$$

Newton's law $m\dot{\vec{v}} = \vec{F}$ gives us the change in the perpendicular velocity

$$\begin{aligned}\delta v &= \frac{1}{m}\int_{-\infty}^{+\infty} \mathrm{d}t F_\perp \\ &= \frac{Gm}{b^2}\int_{-\infty}^{+\infty} \frac{\mathrm{d}t}{\left[1+\left(\frac{vt}{b}\right)^2\right]^{3/2}} \\ &= \frac{Gm}{bv}\int_{-\infty}^{+\infty} \frac{\mathrm{d}s}{(1+s^2)^{3/2}} \\ &= \frac{2Gm}{bv}\end{aligned} \tag{61}$$

using $s = \frac{vt}{b}$. Thus, δv is roughly equal to the acceleration at closest approach, $\frac{GM}{b^2}$, times the duration of the acceleration, $\frac{2b}{v}$.

Strong encounters:
An encounter is strong if $\delta v \sim v$ (which also causes the calculation to break down). This is also when a star will have its path deflected by $\sim 90^\circ \equiv b_{\text{strong}}$.

$$\delta v \sim v \Leftrightarrow b \lesssim b_{90} = \frac{GM}{v^2} \equiv b_{\text{strong}} \,. \tag{62}$$

The cross section for strong encounters is

$$\sigma_{\text{strong}} = \pi b_{\text{strong}}^2 \tag{63}$$

From the virial theorem, we have

$$\begin{aligned} v^2 &\sim \frac{GM}{R} = \frac{GNm}{R} \\ \Rightarrow b_{\text{strong}} &\approx \frac{2R}{N_*}\end{aligned} \tag{64}$$

so we get:

$$\sigma_{\text{strong}} \approx \frac{4\pi}{N_*^2} R^2 \tag{65}$$

which is small since $N \sim 10^{11}$. This means that the probability p of a strong encounter over a single crossing of a star through a galaxy with an average number density of stars n is

$$\begin{aligned} p &= n\sigma_{\text{strong}} R \\ &= \frac{N}{\frac{4}{3}\pi R^3}\frac{4\pi R^2}{N^2} R \\ &= \frac{3}{N} \sim 10^{-11} \,.\end{aligned} \tag{66}$$

This is a tiny probability! So there are likely no strong encounters in a galaxy. For globular clusters, $N \sim 10^4$, so strong encounters are more common.

What about weak encounters?
We have seen that strong encounters are rare, i.e. they practically never happen. Nevertheless, if a star crosses a galaxy many times, it will encounter many weak encounters. Each of those will slightly perturb its velocity until $v_\perp \approx v$. The time it takes for this to happen is the relaxation time of the system.

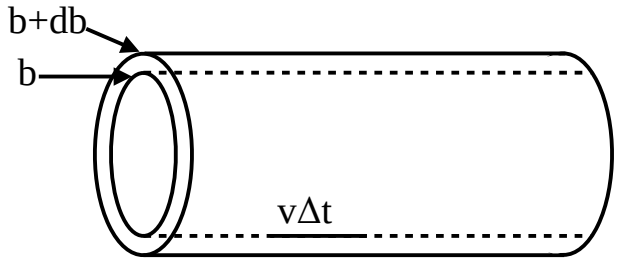
Multiple weak encounters

A star makes a random walk through a galaxy. Its total deviation from its path is the sum of each of its encounters with other stars. For N encounters,

$$\delta v_{\rm tot}^2 = \sum_{i=1}^{N} (\delta v_i)^2 \, . \tag{67}$$

The strength of each encounter depends on the impact parameter b. The number of encounters N within $(b, b+db)$ is

$$N = \underbrace{(2\pi b db)}_{\text{area}} \; \underbrace{(v\Delta t)}_{\text{length}} \; \underbrace{n}_{\text{density}} \, . \tag{68}$$



We then have

$$\begin{aligned} \sum_i (\delta v_i)^2 &= \int_{b_{\rm min}}^{b_{\rm max}} (\text{number of encounters in } (b, b+\mathrm{d}b)) \times (\delta v \text{ for each encounter with } b) \\ &= \int_{b_{\rm min}}^{b_{\rm max}} (2\pi \, v \, \Delta t \, n \, b \, \mathrm{d}b) \left(\frac{2Gm}{bv}\right)^2 \\ &= \frac{8\pi G^2 m^2 n}{v} \Delta t \int_{b_{\rm min}}^{b_{\rm max}} \frac{\mathrm{d}b}{b} \, . \end{aligned} \tag{69}$$

Now we determine the limits on the integral:

$$\begin{aligned} b_{\max} &\approx R \approx 10\,\mathrm{kpc} \\ b_{\min} &\approx b_{\mathrm{strong}} = \frac{2R}{v_*} = 10^{-10}\mathrm{kpc}\ (0.01\mathrm{AU}) \end{aligned} \tag{70}$$

so

$$\begin{aligned} \int_{b_{\min}}^{b_{\max}} \frac{\mathrm{d}b}{b} &= \ln\left(\frac{b_{\max}}{b_{\min}}\right) \equiv \ln\Lambda \ \ \text{(Coulomb logarithm)} \\ &= \ln\left(\frac{10\,\mathrm{kpc}}{10^{-10}\,\mathrm{kpc}}\right) \\ &= \ln\left(10^{11}\right) \approx 25\ . \end{aligned} \tag{71}$$

Relaxation time:
We define the relaxation time t_{relax} through

$$\begin{aligned} \sum_i (\delta v_i)^2 \approx v^2 \Rightarrow \frac{8\pi G^2 m^2 n}{v} \ln\Lambda t_{\mathrm{relax}} \\ \Rightarrow t_{\mathrm{relax}} = \frac{v^3}{8\pi G^2 m^2 n \ln\Lambda}\ . \end{aligned} \tag{72}$$

We now compare this to the dynamical time $t_{\mathrm{orbit}} \approx R/v$ of the system:

$$\frac{t_{\mathrm{relax}}}{t_{\mathrm{orbit}}} = t_{\mathrm{relax}}\frac{v}{R} = \frac{v^4}{8\pi G^2 m^2 n \ln\Lambda R}\ ,. \tag{73}$$

Using the virial theorem $v^2 = \frac{GM}{R}$ and number density $n = \frac{M/m}{\frac{4\pi}{3}R^3}$ gives us:

$$\begin{aligned} &= \frac{(GM/R)^2}{8\pi G^2 m^2 \frac{M/m}{\frac{4\pi}{3}R^3} R \ln\Lambda} \\ &= \frac{M}{8\pi m \frac{3}{4\pi}\ln\Lambda} \\ &= \frac{N}{6\ln\Lambda} = \frac{N_*}{6\ln\left(\frac{b_{\max}}{b_{\min}}\right)} = \frac{N_*}{6\ln\left(\frac{R}{2R/N_*}\right)} \\ &\sim \frac{N_*}{6\ln N_*} \end{aligned} \tag{74}$$

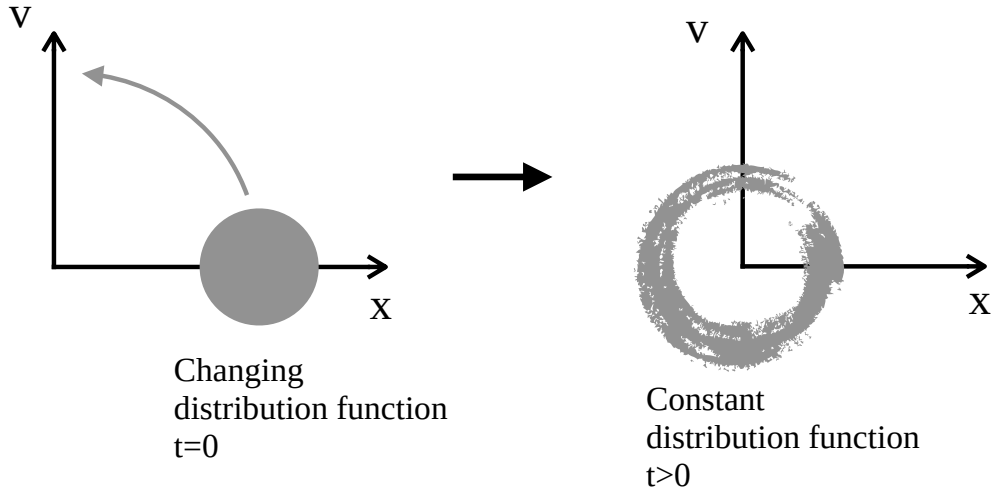
which is very large! Thus, stars are orbiting in an unperturbed collective potential (collisionless)!

2.C Collisionless relaxation

Relaxation occurs in two ways within a galaxy: the collisional gas with interactions reaches a Maxwellian distribution through two-body interactions, but the collisionless systems (stars

and dark matter) must relax through a different process otherwise galaxies and galaxy clusters would not reach a relaxed state within the age of the Universe. We say a system is *relaxed* when its coarse grained phase-space distribution function does not change any more.

Collisionless relaxation processes:
Phase mixing:
The coarse grained phase-space distribution function is distributed over time so doesn't change with time.



Violent relaxation:
Since energy in the stellar and dark matter systems in galaxies can't be efficiently exchanged through collisions, we must find another way for energy exchange. The energy of an individual star (specific energy) is:

$$E = \frac{1}{2}v^2 + \phi \,. \tag{75}$$

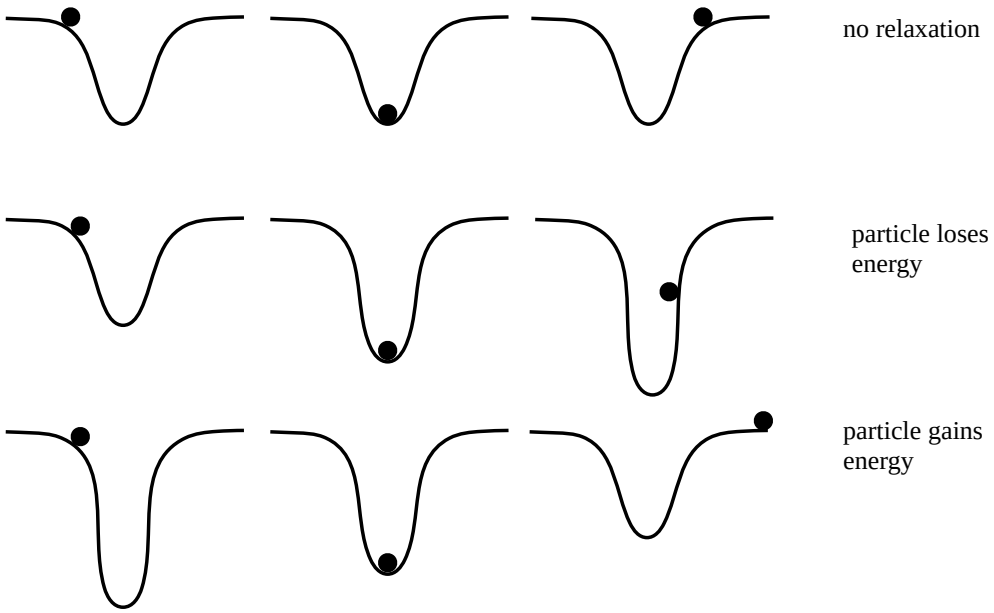
Then the change in energy over time is

$$\begin{aligned} \frac{\mathrm{d}E}{\mathrm{d}t} &= \frac{\partial E}{\partial \vec{v}}\frac{\mathrm{d}\vec{v}}{dt} + \frac{\partial E}{\partial \phi}\frac{\mathrm{d}\phi}{\mathrm{d}t} \\ &= -\vec{v}\cdot\vec{\nabla}\phi + \frac{\mathrm{d}\phi}{\mathrm{d}t} \\ &= -\vec{v}\cdot\vec{\nabla}\phi + \frac{\partial\phi}{\partial t} + \frac{\partial\phi}{\partial\vec{x}}\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} \\ &= -\vec{v}\cdot\vec{\nabla}\phi + \frac{\partial\phi}{\partial t} + \vec{v}\vec{\nabla}\phi \\ &= \frac{\partial\phi}{\partial t} \end{aligned} \tag{76}$$

Thus, the only way for a star to change its energy is by having a time-dependent potential.

To think of this intuitively, we can consider an object moving through a potential well. If the potential is constant with time, the particle will recover the same energy as it comes out

the other side and there is no relaxation. If the potential grows with time, the particle will need to expend more energy to cross it and will not have enough energy to get back out of the potential well, thus losing energy. If the potential shrinks with time, the particle will gain energy as it crosses the well.



As a galaxy or cluster forms, the gravitational potential changes significantly as mass accretes and collapses into a halo. Averaging over all particles, the timescale for violent relaxation $t_{\rm vr}$ is

$$\begin{aligned} t_{\rm vr} &= \left\langle \frac{\left(\frac{{\rm d}E}{{\rm d}t}\right)^2}{E^2} \right\rangle^{-1/2} \\ &= \left\langle \frac{\left(\frac{\partial\phi}{\partial t}\right)^2}{E^2} \right\rangle^{-1/2} \\ &\sim \left\langle \frac{\dot{\phi}^2}{\phi} \right\rangle^{-1/2} \end{aligned} \tag{77}$$

where in the last step we used the time-dependent virial theorem (see Lynden-Bell 1967). This occurs on roughly the same timescale as free-fall since this is the timescale at which the potential changes during collapse. It's very fast, hence 'violent' relaxation!

3 Modelling galaxies

So far, we have looked at the basic dynamical properties of galaxies. Now we discuss the main ingredients of modelling galaxies:

- potential-density pairs (the common potential)

- orbits (trajectories of stars orbiting in a potential)
- phase-space distribution function (distribution of orbits, Vlasov equation)
- stability (Jeans criterion)
- composition of stars (stellar populations), star formation rate, initial mass function
- chemical evolution of galaxies
- active galaxies

3.A Potential-density pairs

Stars move in a collective potential. What are interesting potentials and the related density functions?

Scalar potential:

$$-\vec{\nabla}\phi = \frac{1}{m}\vec{F} \tag{78}$$

Note that $m\phi = U$ is the potential energy of the system and using Poisson's equation $\nabla^2\phi = 4\pi G\rho$, we get

$$\phi(\vec{r}) = G\int \frac{\rho(\vec{r'})}{|\vec{r'}-\vec{r}|}\mathrm{d}^3\vec{r} \tag{79}$$

$$\Rightarrow \text{potential } \phi - \text{density } \rho - \text{pairs!} \tag{80}$$

Examples:

- Kepler/point mass potential:

$$\phi = -\frac{GM}{r} \tag{81}$$

To find $\vec{F}$, we take the gradient of ϕ

$$\frac{1}{m}\vec{F} = \frac{GM}{r^2}\hat{e}_r\ . \tag{82}$$

- Homogeneous sphere:

$$\begin{aligned}\rho(\vec{r}) &= \frac{M}{\frac{4}{3}\pi R^3}\\ \vec{F} &= ?\end{aligned} \tag{83}$$

We do not have ϕ, so we need a different way to get $\vec{F}$. We can use Gauss's theorem for gravity for a surface S_r with radius R enclosing a volume V_r:

$$\begin{aligned}\int_{S_r}\vec{F}\cdot\mathrm{d}\vec{S} &= \int_{V_r}(\vec{\nabla}\cdot\vec{F})\mathrm{d}V\\ &= -m\int_V(\vec{\nabla}^2\phi)\mathrm{d}V\\ &= -4\pi Gm\int\rho(\vec{r})\mathrm{d}V\\ &= -4\pi GM(<r)m\end{aligned} \tag{84}$$

Since we're working with gravity, we have $\vec{F}(r) = -F(r)\hat{e}_r$ and

$$\int_{S_r} \vec{F} \cdot \mathrm{d}\vec{S} = \int_{S_r} F(r)(-\hat{e}_r)\mathrm{d}\vec{S} = -4\pi r^2 F(r) \tag{85}$$

$$\Rightarrow 4\pi r^2 F(r) = r\pi GM(<r)m \tag{86}$$

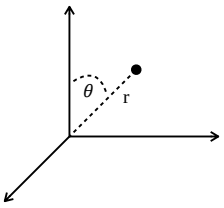
So

$$\begin{aligned} \text{Outside the sphere} : r > R \Rightarrow F(r) &= \frac{GMm}{r^2} \\ \text{Inside the sphere} : r < R \Rightarrow F(r) &= 4\pi G\frac{\rho r}{3}m \end{aligned} \tag{87}$$

From these, we can now also get $\phi : \frac{1}{m}\vec{F} = -\vec{\nabla}^2\phi$

$$\begin{aligned} \text{Outside the sphere} : r > R \Rightarrow \phi(r) &= \frac{GMm}{r} + \text{constant} \\ \text{Inside the sphere} : r < R \Rightarrow \phi(r) &= 2\pi G\frac{\rho r^2}{3}m + \text{constant} \end{aligned} \tag{88}$$

- Mestel disk (example of a disk potential):



$$\phi(r,\theta) = v_c^2\left[\ln\frac{r}{r_0} + \ln\frac{1+|\cos\theta|}{2}\right] \tag{89}$$

Is this a disk? It's hard to see based on the potential, so we need to find ρ. Let's look at Poisson's equation:

$$\nabla^2\phi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\phi}{\partial\theta}\right) + \underbrace{\frac{1}{r^2\sin^2\theta}\frac{\partial^2\phi}{\partial\varphi}}_{=0,\ \text{since no } \varphi \text{ dependence}} \tag{90}$$

Using $\phi = v_c^2\phi_0$:

$$\begin{aligned} \nabla^2\phi &= \frac{v_c^2}{r^2}\frac{\partial}{\partial}\left(r^2\frac{1}{r}\right) + \frac{v_c^2}{r^2\sin\theta}\left(\cos\theta\frac{\partial\phi_0}{\partial\theta} + \sin\theta\frac{\partial^2\phi_0}{\partial\theta^2}\right) \\ &= \frac{v_c^2}{r^2}\left[1 + \left(\frac{\cos\theta}{\sin\theta}\frac{\partial\phi_0}{\partial\theta} + \frac{\partial^2\phi_0}{\partial\theta^2}\right)\right] \end{aligned} \tag{91}$$

We now calculate $\frac{\partial\phi_0}{\partial\theta}$ and $\frac{\partial^2\phi_0}{\partial\theta^2}$. We assume $\cos\theta > 0$. The calculations are the same or $\cos\theta < 0$ except for an overall sign change $\cos\theta \to -\cos\theta$.

$$\phi_0 = \ln\left(\frac{r}{r_0}\right) + \ln\left(\frac{1+\cos\theta}{2}\right) \tag{92}$$

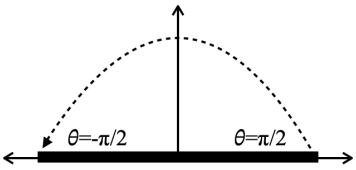
Then

$$\begin{aligned}\frac{\partial\phi}{\partial\theta} &= \frac{2}{1+\cos\theta}\left(-\frac{\sin\theta}{2}\right) = \left(-\frac{\sin\theta}{1+\cos\theta}\right)\\ \frac{\partial^2\phi_0}{\partial\theta^2} &= -\frac{\cos\theta}{1+\cos\theta} - \frac{\sin^2\theta}{(1+\cos\theta)^2}\end{aligned} \tag{93}$$

So

$$\begin{aligned}\frac{\cos\theta}{\sin\theta}\frac{\partial\phi_0}{\partial\theta} + \frac{\partial^2\phi_0}{\partial\theta^2} &= -\frac{2\cos\theta}{1+\cos\theta} - \frac{\sin^2\theta}{(1+\cos\theta)^2}\\ &= \frac{-2\cos\theta - 2\cos^2\theta - \sin^2\theta}{(1+\cos\theta)^2}\\ &= -\frac{1+\cos^2\theta+2\cos\theta}{(1+\cos\theta)^2}\\ &= -\frac{(1+\cos\theta)^2}{(1+\cos\theta)^2}\\ &= -1\end{aligned} \tag{94}$$

For $\cos\theta \neq 0$, this gives $\nabla^2\phi = \frac{v_c^2}{r^2}(1-1) = 0$, so there is no density for $\theta \neq \pi/2$ and all mass is in a thin plane with infinite density ρ (3D density).



We can calculate the surface density

$$\Sigma(r) = \int_{-\infty}^{+\infty} \frac{1}{4\pi G}\vec{\nabla}^2\phi \,\mathrm{d}z \tag{95}$$

With $z = r\cos\theta$ so $\mathrm{d}z = -r\sin\theta\mathrm{d}\theta + \cos\theta dr \approx -r\, d\theta$ since $\theta \approx \pi/2$, we get

$$\begin{aligned}\Sigma(r) &= \int \rho \,\mathrm{d}z\\ &= \int_{\frac{\pi}{2}+\epsilon}^{\frac{\pi}{2}-\epsilon} \frac{1}{4\pi G}\vec{\nabla}^2\phi(-r\,\mathrm{d}\theta)\ .\end{aligned} \tag{96}$$

We go from $\frac{\pi}{2}+\epsilon$ where $z<0$ to $\frac{\pi}{2}-\epsilon$ where $z>0$. We can then switch the bounds

and change the overall sign

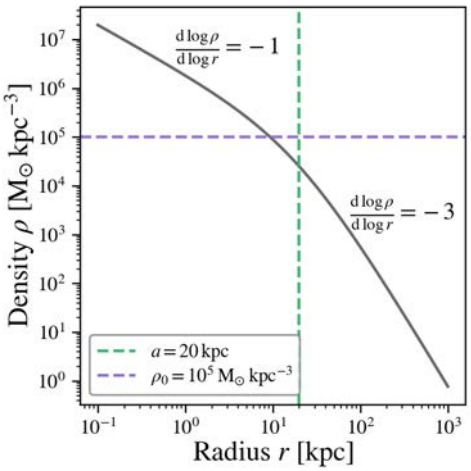
$$\begin{aligned}
\Sigma(r) &= \frac{1}{4\pi G}\int_{\frac{\pi}{2}-\epsilon}^{\frac{\pi}{2}+\epsilon} \vec{\nabla}^2\phi\, r\, \mathrm{d}\theta \\
&= \frac{1}{4\pi G}\int_{\frac{\pi}{2}-\epsilon}^{\frac{\pi}{2}+\epsilon} \frac{v_c^2}{r^2}\left(1+\left[\frac{\cos\theta}{\sin\theta}\frac{\partial\phi_0}{\partial\theta}+\frac{\partial^2\phi_0}{\partial\theta^2}\right]\right) r\, \mathrm{d}\theta \\
&\approx \frac{1}{4\pi G}\frac{v_c^2}{r}\int_{\frac{\pi}{2}-\epsilon}^{\frac{\pi}{2}+\epsilon}\left(\frac{\cos\theta}{\sin\theta}\frac{\partial\phi_0}{\partial\theta}+\frac{\partial^2\phi_0}{\partial\theta^2}\right)\mathrm{d}\theta \\
&\quad \text{(continuous functions} \to 0 \text{ for } \epsilon \to 0) \\
&\approx \frac{1}{4\pi G}\frac{v_c^2}{r}\int_{\frac{\pi}{2}-\epsilon}^{\frac{\pi}{2}+\epsilon}\frac{\partial^2\phi_0}{\partial\theta^2}d\theta \\
&= \frac{1}{4\pi G}\left[\frac{\partial\phi_0}{\partial\theta}\right]_{\frac{\pi}{2}-\epsilon}^{\frac{\pi}{2}+\epsilon}
\end{aligned} \tag{97}$$

When $\theta > \frac{\pi}{2}$, $\cos\theta < 0$ and $|\cos\theta| = -\cos\theta$, and when $\theta < \frac{\pi}{2}$, $\cos\theta > 0$ and $|\cos\theta| = \cos\theta$. So we take the derivative using $-\cos\theta$ in the first term and $\cos\theta$ in the second term

$$\begin{aligned}
\Sigma(r) &= \frac{1}{4\pi G}\left(\left[\frac{\sin\theta}{1-\cos\theta}\right]_{\frac{\pi}{2}+\epsilon} - \left[\frac{-\sin\theta}{1+\cos\theta}\right]_{\frac{\pi}{2}-\epsilon}\right) \\
&\quad (\epsilon \to 0) \\
&= \frac{1}{4\pi G}\frac{v_c^2}{r}(1+1) \\
\Rightarrow \Sigma(r) &= \frac{1}{2\pi G}\frac{v_c^2}{r}\ .
\end{aligned} \tag{98}$$

- Navarro-Frenk-White profile (NFW):
  empirical profile found in simulations of CDM halos.

$$\rho(r) = \frac{\rho_0}{\left(\frac{r}{a}\right)\left(1+\frac{r}{a}\right)^2} \propto \begin{cases} r^{-1} & r \ll a \\ r^{-3} & r \gg a \end{cases} \tag{99}$$



Simulations showed the ρ_0 and a are strongly correlated for CDM halos, so halos are approximately members of a 1-parameter family. The conventional choice for this

parameter is r_{200}, the distance which has an enclosed density 200 times the cosmic critical density ρ_c (which we will cover later) or $M_{200} = 200\rho_c \frac{4}{3}\pi r_{200}^3$.

The *concentration* of a halo is

$$\boxed{c = \frac{r_{200}}{a}} \tag{100}$$

Central result:
The second parameter c is only a very weak function of mass and for fixed mass, and it is the same for all halos in that mass range.

$$\phi = -4\pi\rho_0 a^2 \frac{\ln\left(1+\frac{r}{a}\right)}{\frac{r}{a}} + \text{constant} \tag{101}$$

Related topics:

- Core-cusp problem: From observations of stellar dynamics, the inner profile of halos flattens to a slope ~ 0 (core) instead of -1 (cusp). This is possibly due to supernova feedback, but it could also be resolved through modifications of cold dark matter.
- Diversity of shapes problem: Observationally, halos display diversity in the shapes of their profiles with some cuspier and some more cored profiles whereas, in simulations, halos are universally described by the NFW profile and self-similar across mass ranges (the profiles look the same when scaled).
- Missing satellite problem: Simulations produce more satellite halos than there are observed satellite galaxies. It's possible that not all subhalos form stars, so we need to be able to find "dark subhalos." This could be done by looking for disruptions in stellar streams or through gravitational lensing. Recently, however, there have been many more satellites found as our observational techniques improve.
- Too-big-to-fail problem: This is related to the missing satellites problem, where the number of predicted large halos doesn't match the number of large galaxies observed (but the total number of satellite halos is consistent). The gravitational potential of these galaxies, however, is large enough that they should have collected enough gas and stars to form galaxies and maintain their evolution (e.g. not lose the stars through stripping).

3.B Orbits

Now that we have looked at potential-density pairs, we can study orbits in these potentials. *Orbits* refer to the motion of stars through 6D phase space $(\vec{x}(t), \vec{v}(t))$. Often, the integrals of motion restrict the dimensionality of the orbit (1 per integral of motion).

Integrals of motion:
The orbital energy E is:

$$E = \frac{1}{2}v^2 + \phi(r) = \frac{1}{2}\dot{r}^2 + \phi(r) \tag{102}$$

Taking the time derivative (for a time-independent potential) gives us

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \dot{r}\ddot{r} + \frac{\mathrm{d}\phi}{\mathrm{d}r}\dot{r} = \dot{r}\ddot{r} - \ddot{r}\dot{r} = 0 \tag{103}$$

which implies that the energy is constant along the orbit.

The angular momentum $\vec{L}$ (for a central force potential) is:

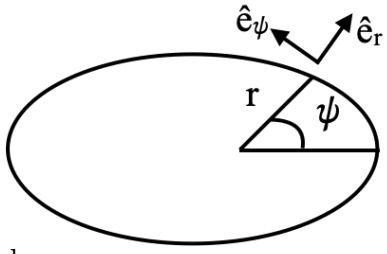
$$\vec{L} = \vec{r} \times \dot{\vec{r}} \tag{104}$$

and the time derivative is

$$\frac{\mathrm{d}\vec{L}}{\mathrm{d}t} = \dot{\vec{r}} \times \dot{\vec{r}} + \vec{r} \times \ddot{\vec{r}} = \vec{r} \times (F(r)\hat{e}_r) = 0 \tag{105}$$

so angular momentum is also constant along the orbit. This means we have a 4D phase space instead of 6D for time-independent, central force potentials, which is often a good approximation.

Central potentials: $\phi = \phi(r)$
Goal: derive equations for radial and tangential components, which is sufficient to describe motion since it is a 4D phase space.



$$\begin{aligned} \hat{e}_r &= \begin{pmatrix} \cos\psi \\ \sin\psi \end{pmatrix} \\ \hat{e}_\psi &= \begin{pmatrix} \sin\psi \\ -\cos\psi \end{pmatrix} \end{aligned} \tag{106}$$

We have:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\vec{r} &= \frac{\mathrm{d}}{\mathrm{d}t}(r\hat{e}_r) \\ &= \dot{r}\hat{e}_r + r\frac{\mathrm{d}}{\mathrm{d}t}(\hat{e}_r) \\ &= \dot{r}\hat{e}_r + r\left(\underbrace{\frac{\mathrm{d}\hat{e}_r}{\mathrm{d}r}}_{=0}\frac{\mathrm{d}r}{\mathrm{d}t} + \frac{\mathrm{d}\hat{e}_r}{\mathrm{d}\psi}\frac{\mathrm{d}\psi}{\mathrm{d}t}\right) \\ &= \dot{r}\hat{e}_r + r\dot{\psi}\underbrace{\frac{\mathrm{d}}{\mathrm{d}\psi}\hat{e}_r}_{= \frac{\mathrm{d}}{\mathrm{d}\psi}\begin{pmatrix}\cos\psi \\ sin\psi\end{pmatrix} = \begin{pmatrix}-\sin\psi \\ cos\psi\end{pmatrix} = \hat{e}_\psi} \\ &= \dot{r}\hat{e}_r + r\dot{\psi}\hat{e}_\psi \end{aligned} \tag{107}$$

so

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}t^2}\vec{r} &= \frac{\mathrm{d}}{\mathrm{d}t}(\dot{r}\hat{e}_r + r\dot{\psi}\hat{e}_\psi) \\
&= (\ddot{r}\hat{e}_r + \dot{r}\dot{\psi}\hat{e}_\psi) + \frac{\mathrm{d}}{\mathrm{d}t}(r\dot{\psi}\hat{e}_\psi) \\
&= \ddot{r}\hat{e}_r + \dot{r}\dot{\psi}\hat{e}_\psi + \dot{r}(\dot{\psi}\hat{e}_\psi) + r\frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\psi}\hat{e}_\psi\right) \\
&= \ddot{r}\hat{e}_r + \dot{r}\dot{\psi}\hat{e}_\psi + \dot{r}\dot{\psi}\hat{e}_\psi + r\ddot{\psi}\hat{e}_\psi + \underbrace{r\dot{\psi}\frac{\mathrm{d}}{\mathrm{d}t}\hat{e}_\psi}_{= \frac{\mathrm{d}\hat{e}_\psi}{\mathrm{d}\psi}\frac{\mathrm{d}\psi}{\mathrm{d}t}} \\
&= \ddot{r}\hat{e}_r + \dot{r}\dot{\psi}\hat{e}_\psi + \dot{r}\dot{\psi}\hat{e}_\psi + r\ddot{\psi}\hat{e}_\psi - r\dot{\psi}^2\hat{e}_r \\
&= \left(\ddot{r} - r\dot{\psi}^2\right)\hat{e}_r + \left(2\dot{r}\dot{\psi} + r\ddot{\psi}\right)\hat{e}_\psi
\end{aligned} \tag{108}$$

and, since we are using a central force,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\vec{r} = F(r)\hat{e}_r \tag{109}$$

where F is the force per unit mass. Combining the two above equations, we get the scalar equations for 4D orbits for the radial and tangential components of motion:

$$\begin{aligned}
\text{radial} &: \ddot{r} - r\dot{\psi}^2 = F(r) \\
\text{tangential} &: 2\dot{r}\dot{\psi} + r\ddot{\psi} = 0 \ .
\end{aligned} \tag{110}$$

For now, we focus on the radial equation and substitute $u = \frac{1}{r}$ to avoid the singularity at $r = 0$. Then

$$F(r) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\left(\frac{1}{u}\right) - \frac{1}{u}\left(\frac{\mathrm{d}\psi}{\mathrm{d}t}\right)^2 \ . \tag{111}$$

With

$$\vec{L} = \vec{r} \times \vec{v} \Rightarrow L = r^2\frac{\mathrm{d}\psi}{\mathrm{d}t} \tag{112}$$

we can parameterize t with ψ to get $u = u(\psi)$:

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{L}{r^2}\frac{\mathrm{d}}{\mathrm{d}\psi} = Lu^2\frac{\mathrm{d}}{\mathrm{d}\psi} \tag{113}$$

Then

$$\begin{aligned}
\vec{L} &= \vec{r} \times \frac{\mathrm{d}}{\mathrm{d}t}\vec{r} \\
&= \vec{r} \times \left(\dot{r}\hat{e}_r + r\dot{\psi}\hat{e}_\psi\right) \\
&= r\hat{e}_r \times \left(\dot{r}\hat{e}_r + r\dot{\psi}\hat{e}_\psi\right) \\
&= r^2\dot{\psi} \\
&= r^2\frac{\mathrm{d}\psi}{\mathrm{d}t} \ .
\end{aligned} \tag{114}$$

Then we get

$$\begin{aligned} F(u) &= Lu^2 \frac{\mathrm{d}}{\mathrm{d}\psi}\left(Lu^2 \frac{\mathrm{d}}{\mathrm{d}\psi}\frac{1}{u}\right) - \frac{1}{u}\left(Lu^2 \frac{\mathrm{d}\psi}{\mathrm{d}\psi}\right) \\ &= Lu^2 \frac{\mathrm{d}}{\mathrm{d}\psi}\left(Lu^2 \frac{-1}{u^2}\frac{\mathrm{d}u}{\mathrm{d}\psi}\right) - \frac{1}{u}\left(Lu^2\right)^2 \\ &= -L^2u^2 \frac{\mathrm{d}^2u}{\mathrm{d}\psi^2} - L^2u^3 \end{aligned} \tag{115}$$

which gives us the orbit equation:

$$\boxed{\frac{\mathrm{d}^2u}{\mathrm{d}\psi^2} + u = -\frac{F(u)}{L^2u^2}} \tag{116}$$

with $u = u(\psi)$. Note that there is no time dependence.

We now examine some examples using this equation.

Examples:

- Kepler:

$$\phi(r) = -\frac{GM}{r} \rightarrow F(r) = -\frac{\mathrm{d}}{\mathrm{d}r}\phi = -\frac{GM}{r^2} = -GMu^2 \tag{117}$$

  so we get the orbital equation:

$$\frac{\mathrm{d}^2u}{\mathrm{d}\psi^2} + u = -\frac{F(u)}{L^2u^2} = \frac{GM}{L^2}\ . \tag{118}$$

  Note that this is a harmonic oscillator, so we know the solution:

$$u(\psi) = C\,cos(\psi - \psi_0) + \frac{GM}{L^2} \tag{119}$$

  Then for $\psi = 0$ to $\psi = 2\pi$, $u(\psi = 0) = u(\psi = 2\pi)$ and we get closed orbits (frequency $\omega = 1$).

- Post-Newtonian relativistic correction:

$$\phi(r) = -\frac{GM}{r}\left(1 + \frac{2GM}{rc^2}\right) \tag{120}$$

  so

$$\begin{aligned} F(r) &= -\frac{\mathrm{d}}{\mathrm{d}r}\phi \\ &= -\frac{GM}{r^2} - \frac{4G^2M^2}{r^3c^2} \\ &= -GMu^2 - \frac{4G^2M^2}{c^2}u^3\ . \end{aligned} \tag{121}$$

So we get the orbital equation:

$$\begin{aligned} &\frac{\mathrm{d}^2u}{\mathrm{d}\psi^2} + u = -\frac{F(u)}{L^2u^2} = \frac{GM}{L^2} + \frac{4G^2M^2}{L^2c^2}u \\ \Rightarrow &\frac{\mathrm{d}^2u}{\mathrm{d}\psi^2} + \left(1 - \underbrace{\frac{4G^2M^2}{L^2c^2}}_{\text{constant}}\right)u = \frac{GM}{L^2}\ . \end{aligned} \tag{122}$$

The term in parentheses implies that $k^2 \neq 1$, so $\omega < 1$, which means that $u(\psi = 0) \neq u(\psi = 2\pi)$. This accounts for the precession of Mercury.

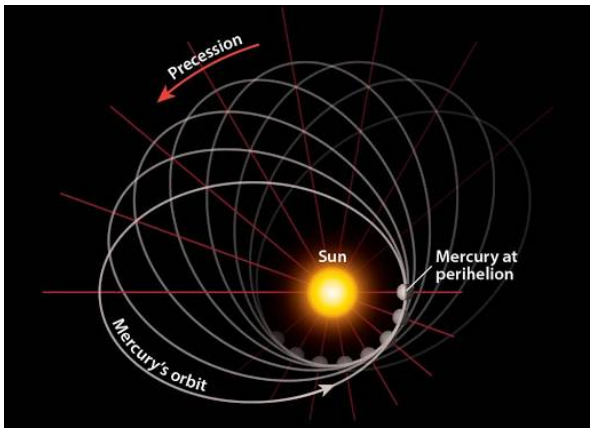
The solution for a harmonic oscillator

$$m\ddot{u} + k\dot{u} = \text{const} \tag{123}$$

is

$$u = \cos(\omega t - \phi) \tag{124}$$

where $\omega^2 = k/m$. Then if $w \neq 1$ and $k \neq 1$, $u(0) \neq u(2\pi)$. Here, t is analogous to ψ, so if the position u after one orbit when $\psi = 2\pi$ is not the same as when $\psi = 0$, the mass has not returned to its previous position and the orbit is not closed.



Axisymmetric Potentials: $\phi = \phi(R, |z|)$
We will derive equations for R, z, and ψ.

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\vec{r} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(r\hat{e}_r + z\hat{e}_z\right) \\ &= \dot{r}\hat{e}_r + r\dot{\psi}\hat{e}_\psi + \dot{z}\hat{e}_z + z\underbrace{\frac{\mathrm{d}}{\mathrm{d}t}(\hat{e}_z)}_{= 0 \text{ since } \hat{e}_z = \begin{pmatrix}0\\0\\1\end{pmatrix}} \\ &= \dot{r}\hat{e}_r + r\dot{\psi}\hat{e}_\psi + \dot{z}\hat{e}_z\ . \end{aligned} \tag{125}$$

Note that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\hat{e}_r &= \dot{\psi}\hat{e}_\psi \\ \frac{\mathrm{d}}{\mathrm{d}t}\hat{e}_\psi &= -\dot{\psi}\hat{e}_r \\ \frac{\mathrm{d}}{\mathrm{d}t}\hat{e}_z &= 0\end{aligned} \tag{126}$$

then

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}t^2} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{r}\hat{e}_r + r\dot{\psi}\hat{e}_\psi + \dot{z}\hat{e}_z\right) \\ &= \ddot{r}\hat{e}_r + \dot{r}\dot{\psi}\hat{e}_\psi + \dot{r}\dot{\psi}\hat{e}_\psi + r\frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\psi}\hat{e}_\psi\right) + \ddot{z}\hat{e}_z \\ &= \ddot{r}\hat{e}_r + \dot{r}\dot{\psi}\hat{e}_\psi + \dot{r}\dot{\psi}\hat{e}_\psi + r\ddot{\psi}\hat{e}_\psi - r\dot{\psi}^2\hat{e}_r + \ddot{z}\hat{e}_z \\ &= \left(\ddot{r} - r\dot{\psi}^2\right)\hat{e}_r + \underbrace{\left(2\dot{r}\dot{\psi} + r\ddot{\psi}\right)}\hat{e}_\psi + \ddot{z}\hat{e}_z \\ &\qquad\qquad = \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}t}\left(r^2\dot{\psi}\right)\end{aligned} \tag{127}$$

and for an axisymmetric potential

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\vec{r} = \vec{F} = \left(-\frac{\partial\phi}{\partial r}, 0, -\frac{\partial\phi}{\partial z}\right) \tag{128}$$

so we get each component of the force:

$$\begin{aligned}\text{radial}: \ddot{r} - r\dot{\psi}^2 &= -\frac{\partial\phi}{\partial r} \\ \text{tangential}: \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}t}\left(r^2\dot{\psi}\right) &= 0 \\ \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\left(r^2\dot{\psi}\right) &= 0 \\ &\text{(using conservation of } L_z = r^2\psi = \text{constant)} \\ \text{vertical}: \ddot{z} &= -\frac{\partial\phi}{\partial z}\ .\end{aligned} \tag{129}$$

We then rewrite this in terms of the effective potential:

$$\phi_{\text{eff}} = \phi + \frac{L_z^2}{2r^2} \tag{130}$$

where the last term is the centrifugal barrier. Since

$$\vec{v} = \frac{\mathrm{d}}{\mathrm{d}t}\vec{r} \tag{131}$$

then, using the above from $\dot{\vec{r}}$,

$$\begin{aligned}E &= \frac{1}{2}\left(\dot{r}^2 + r^2\dot{\psi}^2 + \dot{z}^2\right) + \phi \\ &= \frac{1}{2}\left(\dot{r}^2 + \dot{z}^2\right) + \phi_{\text{eff}}\end{aligned} \tag{132}$$

so

$$\begin{aligned}\ddot{r} = r\dot{\psi}^2 &= \frac{\partial \phi}{\partial r} \\ &= r\dot{\psi}^2 - \frac{\partial}{\partial r}\left(\phi_{\text{eff}} - \frac{L_z^2}{2r^2}\right) \\ &= r\dot{\psi}^2 - \frac{\partial \psi_{\text{eff}}}{\partial r} - \frac{L_z^2}{r^3} \\ &= r\dot{\psi}^2 - \frac{\partial \phi_{\text{eff}}}{\partial r} - \frac{r^4\dot{\psi}^2}{r^3} \text{ (using } L_z = r^2\dot{\psi}) \\ &= -\frac{\partial \phi_{\text{eff}}}{\partial r}\end{aligned} \tag{133}$$

and

$$\ddot{z} = -\frac{\partial \phi_{\text{eff}}}{\partial z} \ . \tag{134}$$

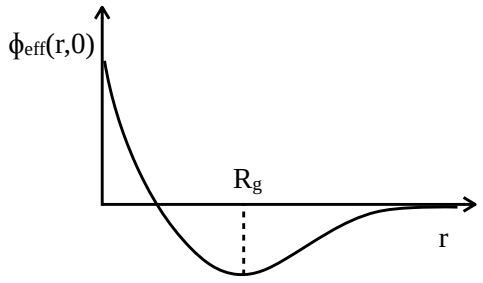
So finally we get the scalar equations for 4D orbits (E, L_z):

$$\begin{aligned}\ddot{r} &= -\frac{\partial \phi_{\text{eff}}}{\partial r} \\ \ddot{z} &= -\frac{\partial \phi_{\text{eff}}}{\partial z} \\ E &= \frac{1}{2}\left(\dot{r}^2 + \dot{z}^2\right) + \phi_{\text{eff}} \ .\end{aligned} \tag{135}$$

Note that the orbits have a uniform rotation around the symmetry axis (z) with $\dot{\psi} = \frac{L_z}{r^2}$, but we have oscillations in r and z. If r is not oscillating, then $z = 0$, and any perturbation leads to oscillations in z and r.

Guiding center and circular orbits:
ϕ_{eff} has a minimum at some R_g such that for a given L_z, $\phi(R_g, 0)$ is minimal:
At the minimum:

$$\begin{aligned}\left.\frac{\partial \phi_{\text{eff}}}{\partial r}\right|_{r=R_g, z=0} &= 0 \\ \left.\frac{\partial \phi_{\text{eff}}}{\partial z}\right|_{r=R_g, z=0} &= 0\end{aligned} \tag{136}$$



where symmetry implies that there is no force at $z = 0$.
Then we get:

$$\begin{aligned}0 &= \left.\frac{\partial \phi_{\text{eff}}}{\partial r}\right|_{R_g, 0} \\ \Rightarrow \left.\frac{\partial \phi_{\text{eff}}}{\partial r}\right|_{R_g, 0} &= \frac{L_z^2}{R_g^3} = R_g\dot{\psi}^2\end{aligned} \tag{137}$$

since

$$\frac{\partial \phi_{\text{eff}}}{\partial r} = \frac{\partial \phi}{\partial r} - \frac{L_z^2}{r^3} \ . \tag{138}$$

We also have

$$\Omega^2 = \dot{\psi}^2 = \frac{L_z^2}{R_g^4} \,. \tag{139}$$

And since

$$\left.\frac{\partial \phi_{\text{eff}}}{\partial r}\right|_{R_g,0} = \left.\frac{\partial \phi_{\text{eff}}}{\partial z}\right|_{R_g,0} = 0 \tag{140}$$

then

$$\ddot{r} = \ddot{z} = 0 \tag{141}$$

so we have a circular orbit with speed $\Omega = \dot{\psi}$. The minimum of ϕ_{eff} occurs at a radius R_g at which a circular orbit has angular momentum L_z and $E = \phi_{\text{eff}}$. This orbit is called the *guiding center*. If an object is pushed off the guiding center, there is a restoring force that leads to oscillations, or *epicycles*.

Epicycle approximation:
In disk galaxies, many stars are on mostly circular orbits, but they are not exactly circular. We look for small perturbations around the circular orbit.
We will define our coordinate system (x, y) as

$$\begin{aligned} x &= r - R_g \\ y &= z \end{aligned} \tag{142}$$

and expand ϕ_{eff} around $(x, y) = (0, 0)$.
Keeping only second-order terms for the epicycle approximation, we get

$$\tilde{\phi}_{\text{eff}}(x,y) = \tilde{\phi}_{\text{eff}}(0,0) + (\tilde{\phi}_{\text{eff},x})x + (\tilde{\phi}_{\text{eff},y})y + (\tilde{\phi}_{\text{eff},xy})xy + \frac{1}{2}(\tilde{\phi}_{\text{eff},xx})x^2 + \frac{1}{2}(\tilde{\phi}_{\text{eff},yy})y^2 + \dots \tag{143}$$

where

$$\begin{aligned} \tilde{\phi}_{\text{eff},x} &= \left.\frac{\partial \tilde{\phi}_{\text{eff}}}{\partial x}\right|_{0,0} = 0 \\ \tilde{\phi}_{\text{eff},y} &= \left.\frac{\partial \tilde{\phi}_{\text{eff}}}{\partial y}\right|_{0,0} = 0 \\ \tilde{\phi}_{\text{eff},xy} &= \left.\frac{\partial^2 \tilde{\phi}_{\text{eff}}}{\partial x \partial y}\right|_{0,0} = 0 \qquad \text{(by symmetry)}. \end{aligned} \tag{144}$$

We define κ and ν:

$$\begin{aligned} \kappa^2 &\equiv \tilde{\phi}_{\text{eff},xx} = \phi_{\text{eff},rr} = \left.\frac{\partial^2 \phi_{\text{eff}}}{\partial r^2}\right|_{R_g,0} \\ \nu^2 &\equiv \tilde{\phi}_{\text{eff},yy} = \phi_{\text{eff},zz} = \left.\frac{\partial^2 \phi_{\text{eff}}}{\partial z^2}\right|_{R_g,0} . \end{aligned} \tag{145}$$

so

$$\begin{aligned} \tilde{\phi}_{\text{eff}} &\approx \frac{1}{2}\kappa^2 x^2 + \frac{1}{2}\nu^2 y^2 + \tilde{\phi}_{\text{eff}}(0,0) \\ &= \frac{1}{2}(\kappa x)^2 + \frac{1}{2}(\nu y)^2 + \tilde{\phi}_{\text{eff}}(0,0) \,. \end{aligned} \tag{146}$$

We can write down the equations of motion for

$$\tilde{\phi}_{\text{eff}}(x,y) = \tilde{\phi}_{\text{eff}}(0,0) + \frac{1}{2}\kappa^2 x^2 + \frac{1}{2}\nu^2 y^2 \tag{147}$$

so

$$\begin{aligned} \ddot{x} = \ddot{r} &= -\frac{\partial \phi_{\text{eff}}}{\partial r} = -\frac{\partial x}{\partial r}\frac{\partial \tilde{\phi}_{\text{eff}}}{\partial x} = -\frac{\partial \tilde{\phi}_{\text{eff}}}{\partial x} = -\kappa^2 x \\ \ddot{y} = \ddot{z} &= -\frac{\partial \phi_{\text{eff}}}{\partial z} = -\frac{\partial y}{\partial z}\frac{\partial \tilde{\phi}_{\text{eff}}}{\partial y} = -\frac{\partial \tilde{\phi}_{\text{eff}}}{\partial y} = -\nu^2 y \end{aligned} \tag{148}$$

and we get the final equations of motion:

$$\boxed{\begin{aligned} \ddot{x} &= -\kappa^2 x \\ \ddot{y} &= -\nu^2 y \end{aligned}}\,. \tag{149}$$

This is harmonic oscillation with epicycle frequency κ and vertical frequency ν in addition to the circular frequency

$$\Omega(r) = \frac{L_z}{r^2} = \frac{v_c}{r} = \sqrt{\frac{1}{r}\frac{\partial \phi}{\partial r}} \tag{150}$$

where for the last equality we used the fact that the centripetal force is equal to the gravitational force $\frac{v_c^2}{r} = \frac{\partial \phi}{\partial r}$ on a circular orbit.

At $r = R_g$ and $\Omega(R_g)$, we can rewrite κ in terms of Ω:

$$\begin{aligned}
&\left(r\frac{\mathrm{d}\Omega^2}{\mathrm{d}r} + 4\Omega^2\right)\Big|_{R_g} \\
=&\left(r\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{1}{r}\frac{\partial\phi}{\partial r}\right) + 4\frac{L_z^2}{r^4}\right)\Big|_{R_g} \\
=&\left(r\left(-\frac{1}{r^2}\frac{\partial\phi}{\partial r} + \frac{1}{r}\frac{\partial^2\phi}{\partial r^2}\right) + 4\frac{L_z^2}{r^4}\right)\Big|_{R_g} \\
=&\left(\underbrace{-\frac{1}{r}\frac{\partial\phi}{\partial r}}_{=\Omega^2=\frac{L_z^2}{r^4}} + \frac{\partial^2\phi}{\partial r^2} + 4\frac{L_z^2}{r^4}\right)\Big|_{R_g} \\
=&\left(-\frac{L_z^2}{r^4} + \frac{\partial^2\phi}{\partial r^2} + 4\frac{L_z^2}{r^4}\right)\Big|_{R_g} \\
=&\left(\frac{\partial^2\phi}{\partial r^2} + 3\frac{L_z^2}{r^4}\right)\Big|_{R_g} \\
=&\left(\frac{\partial}{\partial r}\left(\frac{\partial\phi}{\partial r} - \frac{L_z^2}{r^3}\right)\right)\Big|_{R_g} \\
=&\left(\frac{\partial}{\partial r}\left(\frac{\partial}{\partial r}\underbrace{\left(\phi + \frac{L_z^2}{2r^2}\right)}_{=\phi_{\text{eff}}}\right)\right)\Big|_{R_g} \\
=&\frac{\partial^2\phi_{\text{eff}}}{\partial r^2}\Big|_{R_g} \\
=&\kappa^2 \\
\Rightarrow \kappa^2 =&\left(r\frac{\mathrm{d}(\Omega^2)}{\mathrm{d}r} + 4\Omega^2\right) .
\end{aligned} \tag{151}$$

Motion in the epicycle approximation (valid for $x, y, z \ll R_g$):
We look at each component of the motion:

$$\begin{aligned}
\text{radial} &: r(t) = r_0\cos(\kappa t + \alpha) + R_g \\
\text{vertical} &: z(t) = z_0\cos(\nu t + \beta) \\
\text{tangential} &: \dot{\psi} = \frac{L_z}{r^2} = \frac{L_z}{R_g^2}\left(1 + \frac{x}{R_g}\right)^{-2} = \Omega(R_g)\left(1 + \frac{x}{R_g}\right)^{-2}
\end{aligned} \tag{152}$$

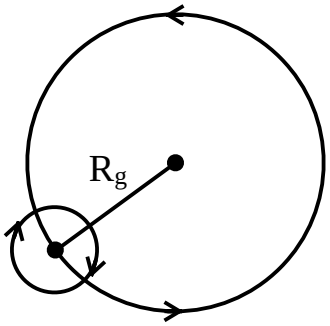
where we use $r = x + R_g$ in the tangential equation. Assuming that $x \ll R_g$ and defining $\Omega_g \equiv \Omega(R_g)$, we can approximate the tangential component to be:

$$\dot{\psi} \approx \Omega_g\left(1 - \frac{2x}{R_g}\right) . \tag{153}$$

We then integrate over time, so

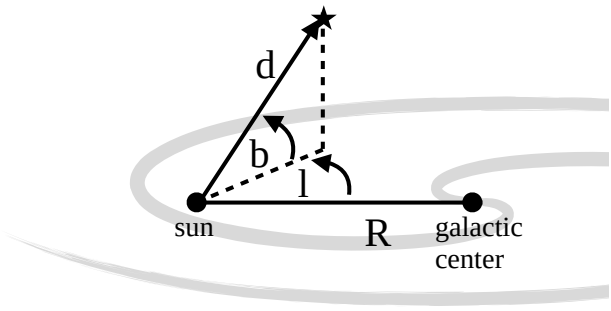
$$\psi(t) = \Omega_g t + \psi_0 - \frac{2\Omega_g r_0}{\kappa R_g} \sin(\kappa t + \alpha) \ . \tag{154}$$

This gives us circular motion of the guiding center with a closed retrograde elliptical orbit in the frame of the guiding center. We also have oscillations in the z direction with frequency ν.



Oort constants: (see Problem Set 3)
Goal: measure the epicycle frequency κ at the position of the sun in the Milky Way by measuring the motion of nearby stars (proper motion an the sky and line of sight velocity).

We use the *galactic coordinate system* to measure the location of stars in the sky (l, b):



l: galactic longitude
b: galactic latitude

R is the distance from the sun to the galactic center ($\sim$ 8kpc) and d is the distance from the sun to the star. l measures the angle in the plane of the Milky Way away from the line of sight to the galactic center, and b measures the angle above the plane of the galaxy. Within this system, we find:

$$\begin{aligned} \text{proper motion} &: \mu \approx d(A\cos(2l) + B) \\ \text{line of sight motion} &: v_{\parallel} \approx dA\sin(2l) \end{aligned} \tag{155}$$

where A and B are the *Oort constants* given by:

$$\begin{aligned} A &= -\frac{1}{2}\frac{d\Omega}{dR} \\ B &= -\left(\Omega + \frac{1}{2}R\frac{\mathrm{d}\Omega}{\mathrm{d}R}\right) \ . \end{aligned} \tag{156}$$

More importantly, they can be related to κ:

$$\boxed{\kappa^2 = -4B(A - B)} \ . \tag{157}$$

Luminosity-velocity relations:
We can relate properties of a galaxy to observables through several equations:

$$\begin{aligned} \theta &= \frac{R}{d} \text{ (apparent size)} \\ F &= \frac{L}{4\pi d^2} \\ v^2 &= \frac{GM}{R} \ . \end{aligned} \tag{158}$$

Introducing surface brightness Σ

$$\begin{aligned} \Sigma = \frac{F}{\theta^2} &= \frac{L}{4\pi d^2} \cdot \frac{d^2}{R^2} \\ &= \frac{L}{4\pi} \cdot \frac{v^4}{G^2 M^2} \end{aligned} \tag{159}$$

then

$$L = \frac{v^4}{\Sigma 4\pi G^2 (M/L)^2} \ . \tag{160}$$

If we assume, for a given class of galaxies, that the surface brightness and the mass-to-light ratio are the same, then

$$\boxed{L \propto v^4} \, . \tag{161}$$

This introduces two important relations.

The *Tully-Fischer relation* is used for spiral galaxies and relates the maximum velocity in the rotation curve $v_{\max}$, which can be measured from HII spectra, and the luminosity:

$$L \propto v_{\max}^4 \ . \tag{162}$$

The *Faber-Jackson relation* is used for ellipticals and relates the velocity dispersion $\sigma_{\rm v}$ to the luminosity:

$$L \propto \sigma_{\rm v}^4 \ . \tag{163}$$

Thus, we can get an estimate of the intrinsic luminosity of a galaxy be measuring stellar velocities. The constant of proportionality is roughly $L_*/(220\,\mathrm{km/s})^4$, where L_* is the characteristic galaxy luminosity.

3.C Phase-space distribution function

We have described the individual orbits in a potential, but this is not sufficient to describe galactic dynamics. We want information of the configuration of all particles. Each star is described by its position $\vec{x}$ and velocity $\vec{v}$, and we need to know this for all stars, i.e. how stars are distributed in the 6D phase space $(\vec{x}, \vec{v})$.

We define a *phase-space distribution function*

$$f(\vec{x}, \vec{v}, t) \mathrm{d}^3\, \vec{x}\, \mathrm{d}^3 \vec{v} \tag{164}$$

as the probability that at time t, a randomly chosen star has $(\vec{x}_*, \vec{v}_*) \in ([\vec{x}, \vec{x} + d\vec{x}], [\vec{v}, \vec{v} + d\vec{v}])$. This means that the function must be normalized for all t, i.e.

$$\int f(\vec{x}, \vec{v}, t) \mathrm{d}^3\, \vec{x}\, \mathrm{d}^3 \vec{v} = 1\,. \tag{165}$$

Collisionless Boltzmann equation:
We want to describe the time evolution of $f(\vec{x}, \vec{v}, t)$. Since probability cannot be destroyed, the 6D continuity equation must hold.

We define the 6D phase-space vector

$$\vec{w} = (\vec{x}, \vec{v}) \tag{166}$$

then

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial \vec{w}} \left(f \dot{\vec{w}} \right) = 0\,. \tag{167}$$

This is the same form as the standard 3D continuity equation. We can rewrite this by expanding out $\vec{w}$ and using velocity $\vec{v} = \dot{\vec{x}}$ and acceleration $\vec{a} = \dot{\vec{v}}$:

$$\begin{aligned} 0 &= \frac{\partial f}{\partial t} + \frac{\partial}{\partial \vec{w}} \left(f \dot{\vec{w}} \right) \\ &= \frac{\partial f}{\partial t} + \frac{\partial}{\partial \vec{x}} (f \dot{\vec{x}}) + \frac{\partial}{\partial \vec{v}} (f \dot{\vec{v}}) \\ &= \frac{\partial f}{\partial t} + \frac{\partial}{\partial \vec{x}} (f \vec{v}) + \frac{\partial}{\partial \vec{v}} \left(f(-\vec{\nabla} \phi) \right) \\ &= \frac{\partial f}{\partial t} + \vec{v} \frac{\partial f}{\partial \vec{x}} - \frac{\partial \phi}{\partial \vec{x}} \frac{\partial f}{\partial \vec{v}}\,. \end{aligned} \tag{168}$$

This gives us the *collisionless Boltzmann equation* (CBE):

$$\boxed{\frac{\partial f}{\partial t} + \vec{v} \frac{\partial f}{\partial \vec{x}} - \frac{\partial \phi}{\partial \vec{x}} \frac{\partial f}{\partial \vec{v}} = 0}\,. \tag{169}$$

Note that another way to see this is by writing out $\frac{df}{dt} = 0$ and taking the limits $\lim_{\vec{x} \to \infty} = 0$ and $\lim_{\vec{v} \to \infty} = 0$.

General Jeans equations:
A solution to the collisionless Boltzmann equation is difficult to obtain, so we instead study moments of the CBE and the phase-space distribution.

Moments of the phase-space density give us some average quantities of the system.

a) The first moment gives the density n of the system:

$$n = \int f\, \mathrm{d}^3 \vec{v}\,. \tag{170}$$

b) The second moment gives the average velocity $\bar{v}_i$:

$$\bar{v}_i = \frac{1}{n} \int v_i f \, \mathrm{d}^3\vec{v} \, . \tag{171}$$

c) The third moment gives the velocity dispersion σ_{ij}^2:

$$\begin{aligned} \overline{v_i v_j} &= \frac{1}{n} \int v_i v_j f \, \mathrm{d}^3\vec{v} \\ \sigma_{ij}^2 &= \overline{v_i v_j} - \bar{v}_i \bar{v}_j = \overline{(v_i - \bar{v}_i)(v_j - \bar{v}_j)} \, . \end{aligned} \tag{172}$$

We now examine moments of the collisionless Boltzmann equation more closely. We break each integral into three terms to simplify each individually.

a) First moment:

$$\begin{aligned} &\int \mathrm{d}^3\vec{v} \left(\frac{\partial f}{\partial t} + \vec{v} \frac{\partial f}{\partial \vec{x}} - \frac{\partial \phi}{\partial \vec{x}} \frac{\partial f}{\partial \vec{v}} \right) = 0 \\ &\underbrace{\int \mathrm{d}^3\vec{v} \, \frac{\partial f}{\partial t}}_{①} + \underbrace{\int \mathrm{d}^3\vec{v} \, \vec{v} \frac{\partial f}{\partial \vec{x}}}_{②} - \underbrace{\int \mathrm{d}^3\vec{v} \, \frac{\partial \phi}{\partial \vec{x}} \frac{\partial f}{\partial \vec{v}}}_{③} = 0 \end{aligned} \tag{173}$$

$$\begin{aligned} ① &: \int \mathrm{d}^3\vec{v} \, \frac{\partial f}{\partial t} = \frac{\partial}{\partial t} \int \mathrm{d}^3\vec{v} f = \frac{\partial n}{\partial t} \\ ② &: \int \mathrm{d}^3\vec{v} \, \vec{v} \frac{\partial f}{\partial \vec{x}} = \frac{\partial}{\partial \vec{x}} \left(\int \mathrm{d}^3\vec{v} \, \vec{v} f \right) = \frac{\partial}{\partial \vec{x}} \left(n\bar{\vec{v}} \right) = \sum_i \frac{\partial}{\partial x_i} \left(n \bar{v}_i \right) \\ ③ &: \int \mathrm{d}^3\vec{v} \, \frac{\partial \phi}{\partial \vec{x}} \frac{\partial f}{\partial \vec{v}} = \frac{\partial \phi}{\partial \vec{x}} \int \mathrm{d}^3\vec{v} \, \frac{\partial f}{\partial \vec{v}} = \frac{\partial \phi}{\partial \vec{x}} \left[f \right]_{\vec{v}=-\infty}^{\vec{v}=+\infty} = 0 \end{aligned} \tag{174}$$

For the third term, we used the fact that phase-space distribution goes to 0 at $\pm\infty$ for physical systems.

This gives us the *3D continuity equation*:

$$\boxed{\frac{\partial n}{\partial t} + \frac{\partial}{\partial \vec{x}} \left(n\bar{\vec{v}} \right) = 0} \, . \tag{175}$$

b) Second moment:

$$\begin{aligned} &\int \mathrm{d}^3\vec{v} \, v_j \left(\frac{\partial f}{\partial t} + \vec{v} \frac{\partial f}{\partial \vec{x}} - \frac{\partial \phi}{\partial \vec{x}} \frac{\partial f}{\partial \vec{v}} \right) = 0 \\ &\underbrace{\int \mathrm{d}^3\vec{v} \, v_j \frac{\partial f}{\partial t}}_{①} + \underbrace{\int \mathrm{d}^3\vec{v} \, v_j \vec{v} \frac{\partial f}{\partial \vec{x}}}_{②} - \underbrace{\int \mathrm{d}^3\vec{v} \, v_j \frac{\partial \phi}{\partial \vec{x}} \frac{\partial f}{\partial \vec{v}}}_{③} = 0 \end{aligned} \tag{176}$$

$$
\begin{aligned}
\textcircled{1}: \int \mathrm{d}^3\vec{v}\, v_j \frac{\partial f}{\partial t} &= \frac{\partial}{\partial t}\int \mathrm{d}^3\vec{v}\, v_j f = \frac{\partial}{\partial t}\left(n\bar{v}_j\right) = \frac{\partial n}{\partial t}\bar{v}_j + n\frac{\partial \bar{v}_j}{\partial t} \\
&= -\bar{v}_j \sum_i \frac{\partial}{\partial x_i}\left(n\bar{v}_i\right) + n\frac{\partial \bar{v}_j}{\partial t} = n\frac{\partial \bar{v}_j}{\partial t} - \bar{v}_j \sum_i \frac{\partial}{\partial x_i}\left(n\bar{v}_i\right) \\
&\text{using the continuity equation } \frac{\partial n}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\left(n\bar{v}_i\right) \\
&\text{to go from the first line to the second}
\end{aligned}
$$

$$
\begin{aligned}
\textcircled{2}: \int \mathrm{d}^3\vec{v}\, v_j \vec{v}\frac{\partial f}{\partial \vec{x}} &= \int \mathrm{d}^3\vec{v}\, v_j \sum_i v_i \frac{\partial f}{\partial x_i} = \sum_i \frac{\partial}{\partial x_i}\underbrace{\int \mathrm{d}^3\vec{v}\, v_j v_i f}_{= n\overline{v_j v_i} = n\left(\sigma_{ij}^2 + \bar{v}_i\bar{v}_j\right)} \\
&= \sum_i \frac{\partial}{\partial x_i}\left(n\left(\sigma_{ij}^2 + \bar{v}_i\bar{v}_j\right)\right) \\
\textcircled{3}: \int \mathrm{d}^3\vec{v}\, v_j \frac{\partial \phi}{\partial \vec{x}}\frac{\partial f}{\partial \vec{v}} &= \int \mathrm{d}^3\vec{v}\, v_j \sum_i \frac{\partial \phi}{\partial x_i}\frac{\partial f}{\partial v_i} = \sum_i \frac{\partial \phi}{\partial x_i}\int \mathrm{d}^3\vec{v}\, v_j \frac{\partial f}{\partial v_i} \qquad (177) \\
&\quad ((k,l,i) \text{ are permutations of } (1,2,3)) \\
&= \sum_i \frac{\partial \phi}{\partial x_i}\int \mathrm{d}v_k \int \mathrm{d}v_l \underbrace{\int \mathrm{d}v_i \left(v_j \frac{\partial f}{\partial v_i}\right)}_{= [v_j f]_{v_i=-\infty}^{v_i=+\infty} - \int \mathrm{d}v_i \frac{\partial v_j}{\partial v_i} f} \\
&= \int \mathrm{d}v_i \delta_{ij} f \\
&= -\sum_i \frac{\partial \phi}{\partial x_i}\int \mathrm{d}v_k \int \mathrm{d}v_l \int \mathrm{d}v_i \delta_{ij} f \\
&= -\sum_i \frac{\partial \phi}{\partial x_i}\int \mathrm{d}^3\vec{v}\delta_{ij} f \\
&= -n\frac{\partial \phi}{\partial x_j}
\end{aligned}
$$

Plugging each term back in, we get

$$
n\frac{\partial \bar{v}_j}{\partial t} - \bar{v}_j \sum_i \frac{\partial}{\partial x_i}(n\bar{v}_i) + \sum_i \frac{\partial}{\partial x_i}\left[n\left(\sigma_{ij}^2 + \bar{v}_i\bar{v}_j\right)\right] + n\frac{\partial \phi}{\partial x_i} = 0 \qquad (178)
$$

which we can rewrite

$$n\frac{\partial \bar{v}_j}{\partial t} - \underline{\bar{v}_j \sum_i \frac{\partial}{\partial x_i}(n\bar{v}_i)} + \sum_i \frac{\partial}{\partial x_i}(n\sigma_{ij}^2) + \underbrace{\sum_i \frac{\partial}{\partial x_i}(n\bar{v}_i\bar{v}_j)}_{= \sum_i (n\bar{v}_i)\frac{\partial}{\partial x_i}\bar{v}_j + \underline{\sum_i \bar{v}_j \frac{\partial}{\partial x_i}(n\bar{v}_i)}} + n\frac{\partial \phi}{\partial x_j} = 0 \tag{179}$$

where the two underlined terms cancel. This gives us

$$n\frac{\partial \bar{v}_j}{\partial t} + \sum_i (n\bar{v}_i)\frac{\partial}{\partial x_i}\bar{v}_j + \sum_i \frac{\partial}{\partial x_i}(n\sigma_{ij}^2) + n\frac{\partial \phi}{\partial x_j} = 0\ . \tag{180}$$

This is the *Jeans equation*, often written

$$\boxed{\frac{\partial \bar{v}_j}{\partial t} + \sum_i \bar{v}_i \frac{\partial \bar{v}_j}{\partial x_i} = -\frac{1}{n}\sum_i \frac{\partial (n\sigma_{ij}^2)}{\partial x_i} - \frac{\partial \phi}{\partial x_j}} \tag{181}$$

Each term can be physically interpreted:

$$\begin{aligned} \frac{\partial \bar{v}_j}{\partial t} &: \text{acceleration of fluid} \\ \sum_i \bar{v}_i \frac{\partial \bar{v}_j}{\partial x_i} &: \text{kinematic viscosity/shear} \\ -\frac{1}{n}\sum_i \frac{\partial (n\sigma_{ij}^2)}{\partial x_i} &: \text{pressure} \\ -\frac{\partial \phi}{\partial x_j} &: \text{gravity} \end{aligned} \tag{182}$$

Jeans equations in spherical systems:
We can convert to spherical coordinates and take velocity moments to give us the Jeans equations in spherical coordinates. This is complicated!

To simplify, we take the radial Jeans equation and focus on steady-state symmetric systems.

Implications:

- $\frac{\partial}{\partial t} = 0$ since we have steady state
- $\bar{v}_r = 0$ otherwise we have net radial motion
- $\bar{v}_\theta = \bar{v}_\phi = 0$ or the symmetry is broken
- $\sigma_{r\phi}^2 = \sigma_{r\theta}^2 = 0$ or the symmetry is broken
- $\sigma_{\phi\phi}^2 = \sigma_{\theta\theta}^2 \equiv \sigma_t^2$ or the symmetry is broken.

The simplified Jeans equation is:

$$\frac{1}{n}\frac{\partial}{\partial r}(n\sigma_{rr}^2) + \frac{2(\sigma_{rr}^2 - \sigma_t^2)}{r} = -\frac{\partial\phi}{\partial r} = -\frac{GM(<r)}{r^2} \tag{183}$$

where we've plugged in gravity as the force.

We have three limits we can look at:

- $\sigma_{rr}^2 \ll \sigma_t^2$: nearly circular orbits
- $\sigma_{rr}^2 \gg \sigma_t^2$: nearly radial orbits
- $\sigma_{rr}^2 = \sigma_t^2$: isotropic orbits

We define the *anisotropy parameter*:

$$\beta = 1 - \frac{\sigma_t^2}{\sigma_{rr}^2} \tag{184}$$

which gives us a useful form of the Jeans equation for observations:

$$\boxed{\frac{1}{n}\frac{\partial}{\partial r}(n\sigma_{rr}^2) + \frac{2\beta\sigma_{rr}^2}{r} = -\frac{GM(<r)}{r^2}}. \tag{185}$$

This depends only on radial components with uncertainty from β, assuming spherical symmetry and a steady-state system.

This can be simplified further to get mass estimates:

$$\begin{aligned} M(<r) &= -\frac{r^2}{G}\left(\frac{1}{n}\frac{\partial}{\partial r}(n\sigma_{rr}^2) + \frac{2\beta\sigma_{rr}^2}{r}\right) \\ &= -\frac{r\sigma_{rr}^2}{G}\left(\frac{r}{n\sigma_{rr}^2}\frac{\partial}{\partial r}(n\sigma_{rr}^2) + 2\beta\right) \\ &= -\frac{r\sigma_{rr}^2}{G}\left(\frac{r}{n}\frac{\mathrm{d}n}{\mathrm{d}r} + \frac{r}{\sigma_{rr}^2}\frac{\mathrm{d}\sigma_{rr}^2}{\mathrm{d}r} + 2\beta\right) \\ &= -\frac{r\sigma_{rr}^2}{G}\left(\frac{\mathrm{d}\ln n}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln\sigma_{rr}^2}{\mathrm{d}\ln r} + 2\beta\right) \end{aligned} \tag{186}$$

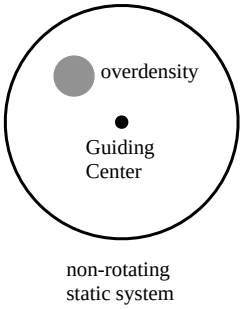
where the last line can be measured with observations.

3.D Stability of stellar systems

The existence of equilibrium solutions to the collisionless Boltzmann equation does not assure stability. Real stellar systems are subject to perturbations. What is important for stability?

Small scales: Jeans instability and random motions

Consider a nearly uniform distribution of stars with perturbations with respect to a static uniform background. We can study the stability of this configuration by inspecting the continuity and the Jeans equations.



We first rewrite and simplify the Jeans equations:

$$\frac{\partial \bar{v}_j}{\partial t} + \sum_i \bar{v}_i \frac{\partial \bar{v} j}{\partial x_i} = -\frac{1}{n} \sum_i \frac{\partial (n\sigma_{ij})}{\partial x_i} - \frac{\partial \phi}{\partial x_j} \, . \tag{187}$$

We can rewrite the number density n using $\rho = mn$ and assume that σ_{ij} is isotropic so the pressure is $P = \rho\sigma_{ij}^2 = \rho\sigma_{ij} = mn\sigma_{ij}^2$. Then we can rewrite the Jeans equations as:

$$\frac{\partial \vec{v}}{\partial t} + \left(\vec{v} \cdot \vec{\nabla}\right) \vec{v} = -\vec{\nabla}\phi - \frac{1}{\rho}\vec{\nabla} P \, . \tag{188}$$

Similarly, the continuity equation becomes:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0 \, . \tag{189}$$

Note that we have dropped the $^-$'s (average value symbols) for simplicity in our equations and $\vec{v}$ is referring to the average velocities at $(\vec{x}, t)$. We will continue with this convention in the following calculations.

Small perturbations:
For a small perturbation in a static uniform background, we have

$$\begin{aligned} \rho &= \rho_0 + \epsilon\rho_1(\vec{x}, t) \\ \vec{v} &= \vec{v}_0 + \epsilon\vec{v}_1(\vec{x}, t) \\ P &= P_0 + \epsilon P_1(\vec{x}, t) \\ \phi &= \phi_0 + \epsilon\phi_1(\vec{x}, t) \, . \end{aligned} \tag{190}$$

We can choose $\phi_0 = 0$ and, since the background is static, $\vec{v}_0 = \vec{0}$. ρ_0 and P_0 are both nonzero constants. Note that this is not a physical set of conditions since Poisson's equation gives $\nabla^2\phi_0 = 4\pi G\rho_0$ so $\phi_0 = 0$ implies $\rho_0 = 0$, but we continue with our calculations ignoring this. This is known as the *Jeans swindle*.

We can plug this into the continuity equation:

$$\frac{\partial}{\partial t}\rho_0 + \epsilon\frac{\partial}{\partial t}\rho_1 + \vec{\nabla}\left(\rho_0\vec{v}_0 + \epsilon\rho_1\vec{v}_0 + \epsilon\rho_0\vec{v}_1 + \epsilon^2\rho_1\vec{v}_1\right) = 0 \, . \tag{191}$$

Performing derivatives on constants and neglecting terms of order ϵ^2, this becomes:

$$\begin{aligned} \epsilon\frac{\partial}{\partial t}\rho_1 + \vec{\nabla}\cdot(\epsilon\rho_0\vec{v}_1) = 0 \\ \Rightarrow \frac{\partial\rho_1}{\partial t} + \rho_0\vec{\nabla}\cdot\vec{v}_1 = 0\ . \end{aligned} \tag{192}$$

We then plug this into the Jeans equation:

$$\left(\frac{\partial\vec{v}_0}{\partial t} + \epsilon\frac{\partial\vec{v}_1}{\partial t}\right) + \Big((\vec{v}_0+\epsilon\vec{v}_1)\cdot\vec{\nabla}\Big)(\vec{v}_0+\epsilon\vec{v}_1) = -\vec{\nabla}(\phi_0+\epsilon\phi_1) - \frac{1}{\rho_0+\epsilon\rho_1}\vec{\nabla}(P_0+\epsilon P_1) \tag{193}$$

then

$$\begin{aligned} \epsilon\frac{\partial\vec{v}_1}{\partial t} = -\epsilon\vec{\nabla}\phi_1 - \underbrace{\frac{1}{\rho_0+\epsilon\rho_1}\vec{\nabla}(\epsilon P_1)}\ . \\ \approx \epsilon\frac{\vec{\nabla}P_1}{\rho_0} \end{aligned} \tag{194}$$

We can write

$$\vec{\nabla}P_1 = \frac{\partial P_1}{\partial\vec{x}} = \underbrace{\left(\frac{\partial P_1}{\partial\rho}\right)\Big|_{\rho_0}}_{v_s^2}\frac{\partial\rho_1}{\partial\vec{x}} \tag{195}$$

where v_s is the sound speed, or the speed at which perturbations can propagate. Returning to the previous equation, this gives us:

$$\begin{aligned} \epsilon\frac{\partial\vec{v}_1}{\partial t} = -\epsilon\vec{\nabla}\phi_1 - \epsilon\frac{v_s^2}{\rho_0}\vec{\nabla}\rho_1 \\ \Rightarrow \frac{\partial\vec{v}_1}{\partial t} = -\vec{\nabla}\phi_1 - \frac{v_s^2}{\rho_0}\vec{\nabla}\rho_1\ . \end{aligned} \tag{196}$$

We now combine the time derivative of the continuity with the Jeans equation:

$$\begin{aligned} \frac{\partial^2\rho_1}{\partial t^2} + \frac{\partial}{\partial t}\left(\rho_0\vec{\nabla}\cdot\vec{v}_1\right) = 0 \\ \Rightarrow \frac{\partial^2\rho_1}{\partial t^2} + \rho_0\underbrace{\vec{\nabla}\cdot\left(\frac{\partial\vec{v}_1}{\partial t}\right)} = 0 \\ = -\vec{\nabla}^2\phi_1 - \frac{v_s^2}{\rho_0}\vec{\nabla}^2\rho_1 \end{aligned} \tag{197}$$

so

$$\begin{aligned} \frac{\partial^2\rho_1}{\partial t^2} + \rho_0\left(-\vec{\nabla}^2\phi_1 - \frac{v_s^2}{\rho_0}\underbrace{\vec{\nabla}^2\rho_1}\right) = 0\ . \\ = 4\pi G\rho_1 \text{ (from Poisson}'\text{s Equation)} \end{aligned} \tag{198}$$

Finally, we get a wave equation for ρ_1:

$$\boxed{\frac{\partial^2\rho_1}{\partial t^2} - 4\pi G\rho_0\rho_1 - v_s^2\vec{\nabla}^2\rho_1 = 0}\ . \tag{199}$$

We use an ansatz for the solution of the form

$$\rho_1 = C \exp^{i(\vec{k}\cdot\vec{c}-\omega t)} \tag{200}$$

which gives the time evolution of perturbations. We plug this into the wave equation and get

$$w^2 = v_s^2 k^2 - 4\pi G\rho_0 \, . \tag{201}$$

We have two solutions:

$w^2 > 0$: the exponent is imaginary, so we get stable oscillating modes

$w^2 < 0$: the exponent is real, so we get unstable growing or decaying modes

If $w = 0$:

$$\lambda_J^2 = \left(\frac{2\pi}{k}\right)^2 = \frac{\pi v_s^2}{G\rho_0} \, . \tag{202}$$

Jeans length and mass:
The *Jeans length* λ_J is the maximum size a perturbation can be to remain stable. The *Jeans mass* M_J is the corresponding mass enclosed within the Jeans length of a given substance.

$$\boxed{\begin{aligned} \lambda_J^2 &= \frac{\pi v_s^2}{G\rho_0} = \frac{\pi\sigma^2}{G\rho_0} \\ M_J &= \frac{4}{3}\pi\rho_0\lambda_J^3 \end{aligned}} \, . \tag{203}$$

So we have stability for $\lambda < \lambda_J$ and $M < M_J$. Note that for collisional gas, the Jeans length is determined by the sound speed v_s and for collisionless dark matter and stars, the Jeans length is determined by the pressure from the velocity dispersion σ.

Meaning of the Jeans length:
If perturbations can be crossed before collapse, pressure can stabilize the collapse.

The free-fall time is

$$t_{\text{ff}} \sim \frac{1}{\sqrt{G\rho}} \tag{204}$$

and the perturbation crossing time is

$$t_{\text{cross}} \sim \frac{r}{v_s} \, . \tag{205}$$

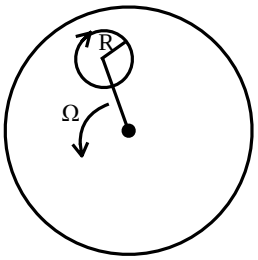
Then we get collapse if

$$\begin{aligned} t_{\text{cross}} &> t_{\text{ff}} \\ \frac{r}{v_s} &> \frac{1}{\sqrt{G\rho}} \\ \Rightarrow r^2 &> \frac{v_s^2}{G\rho} \end{aligned} \tag{206}$$

which is similar to the Jeans length result, differing only by a factor of π. So, random motion and pressure can stabilize perturbations on small scales.

Large scales: Toomre instability and rotational motion.

Consider a rotating stellar disk where radial perturbations can occur. We study the stability of this configuration by inspecting the centripetal and acceleration forces. Note that mass and angular momentum are conserved during the perturbation: $\dot{m} = \dot{L} = 0$.



rotating
non-static system

During the perturbation, $R \to R'$ with $R' = R - \mathrm{d}r$. We want to know when this will lead to collapse and when it will be stable. This is a competition between centripetal and gravitational forces.

The change in gravitational acceleration is

$$\begin{aligned} a_g &= \frac{G\pi R^2 \Sigma}{R'^2}, \text{ and } \pi R^2\Sigma \text{ is mass } (\Sigma \text{ is surface density}) \\ \Rightarrow \frac{da_g}{dR'} &= \frac{-2G\pi R^2\Sigma}{R'^3} \,. \end{aligned} \tag{207}$$

The change in centripetal acceleration, with rotational frequency of the patch Ω, is

$$\begin{aligned} L = \Omega R^2 &= \Omega' R'^2 \text{ (since } \dot{L} = 0) \\ \Rightarrow \Omega' &= \Omega \left(\frac{R}{R'}\right)^2 \,. \end{aligned} \tag{208}$$

So

$$\begin{aligned} a_c &= \frac{R'^2\Omega'^2}{R'} = R'\Omega'^2 = \Omega^2 \frac{R^4}{R'^3} \\ \Rightarrow \frac{\mathrm{d}a_c}{\mathrm{d}R'} &= \frac{-3\Omega^2 R^4}{R'^4} \,. \end{aligned} \tag{209}$$

Stability: the system is stable if $|\mathrm{d}a_g| < |\mathrm{d}a_c|$. So we need

$$\begin{aligned} \frac{2\pi G R^2\Sigma}{R'^3} &< \frac{3\Omega^2 R^4}{R'^4} \\ \Rightarrow \frac{2\pi G\Sigma}{3\Omega^2} &< R \underbrace{\frac{R}{R'}}_{\approx 1} \end{aligned} \tag{210}$$

and the disk is stable if

$$\boxed{R_{\mathrm{rot}} > \frac{2\pi G\Sigma}{3\Omega^3}} \,. \tag{211}$$

Full stability criterion: On small scales, we have stability if $R < \lambda_J$ and on large scales, we have stability if $R > R_{\rm rot}$. Small scales are stabilized by random motion and large scales are stabilized by rotational motion. The system is unstable if $\lambda_J < R < R_{\rm rot}$. We can combine the two criteria and get full stability when $\lambda_J \geq R_{\rm rot}$. This gives us (adapting λ_J from an arbitrary 3D potential to a 2D disk):

$$\begin{aligned} \frac{\pi}{8}\frac{\sigma^2}{G\Sigma} &\geq \frac{2\pi G\Sigma}{3\Omega^2} \\ \Rightarrow \sigma_{\rm crit} &\geq \frac{4}{\sqrt{3}}\frac{G\Sigma}{\Omega}\,. \end{aligned} \tag{212}$$

Note that the angular speed of the patch is only approximately Ω. It actually rotates with epicyclic frequency κ, which is not too far off from Ω for real galaxies. We can relate κ to Ω for a typical galactic disk:

$$\kappa^2(R_g) = \left(r\frac{{\rm d}\Omega^2}{{\rm d}r} + 4\Omega^2\right)\Big|_{R_g} \tag{213}$$

and in galaxies with circular velocity that is approximately constant:

$$\Omega = \frac{v_c}{r} \Rightarrow \kappa^2 = 2\Omega^2 \Rightarrow \kappa = \sqrt{2}\Omega\,. \tag{214}$$

So for galaxies:

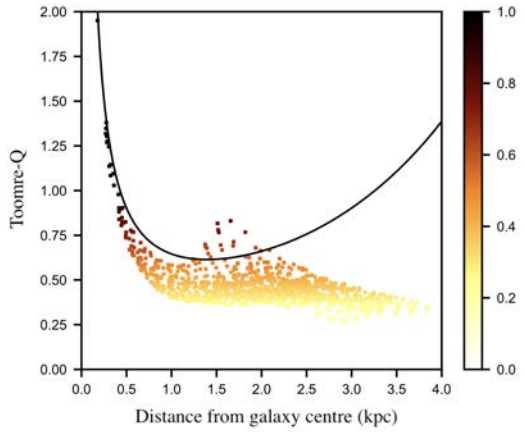
$$\sigma_{\rm crit} \geq \frac{4}{\sqrt{3}}\frac{G\Sigma}{\kappa^2/\sqrt{2}} = \underbrace{\sqrt{\frac{32}{3}}}_{3.26}\frac{G\Sigma}{\kappa}\,. \tag{215}$$

Toomre finds

$$\sigma_{\rm crit} = 3.26\frac{G\Sigma}{\kappa} \tag{216}$$

Toomre criterion Q:
We can write the stability criterion Q for rotating disks:

$$\boxed{Q = \frac{\sigma}{\sigma_{\rm crit}}\begin{cases} > 1 : \text{stable} \\ < 1 : \text{unstable} \end{cases}} \tag{217}$$



Here we show Q as a function of radius from the galactic center for the galaxy DLA0817 (the Wolfe Disk) from Neelemen et al. 2020. The solid line shows Q assuming the gas density falls off exponentially. The points show observed data, which underestimates Q likely due to beam smearing which increases measured surface density.

Extended Data Figure 7b. Neeleman, M., Prochaska, J.X., Kanekar, N. et al. A cold, massive, rotating disk galaxy 1.5 billion years after the Big Bang. Nature 581, 269–272 (2020). https://doi.org/10.1038/s41586-020-2276-y

Code used to generate kinematic models: mneeleman, and J. Xavier Prochaska. "Mneeleman/qubefit: Small Documentation Updates". Zenodo, February 11, 2021. https://doi.org/10.5281/zenodo.4534407.

3.E Stellar population synthesis

So far, we have only viewed stars as massive particles without other features. But:

- Stars are constantly born at a star formation rate (SFR) $\psi(t)$
- Stars are born with a certain mass spectrum. This is the initial stellar mass function (IMF) $\phi(m)$
- Stars emit light with flux at different wavelengths F_λ.

The galactic spectrum is a superposition of stellar spectra. Adding the stellar spectra taking into account $\psi(t)$ and $\phi(m)$ allows us to constrain the initial mass function and star formation rate of galaxies. We can use this to learn about the stellar population and galaxy evolution.

Star formation rate:
The units of star formation rate are usually $[\psi] = \mathrm{M}_\odot/\mathrm{yr}$. For the Milky Way, $\psi(t) \sim 3\mathrm{M}_\odot/\mathrm{yr}$

$$(t) = \frac{\mathrm{d}m}{\mathrm{d}t} \, . \tag{218}$$

There are a few observational indications for the star formation rate:

- Far infrared (FIR) emission from dust around young stars:

$$\frac{\mathrm{SFR}_{\mathrm{FIR}}}{\mathrm{M}_\odot/\mathrm{yr}} \sim \frac{L_{\mathrm{FIR}}}{5.8 \times 10^9 \mathrm{L}_\odot} \tag{219}$$

- H_α emission from HII regions around young stars:

$$\frac{\mathrm{SFR}_{\mathrm{H}_\alpha}}{\mathrm{M}_\odot/\mathrm{yr}} \sim \frac{L_{H_\alpha}}{1.3 \times 10^{41} \mathrm{erg/s}} \tag{220}$$

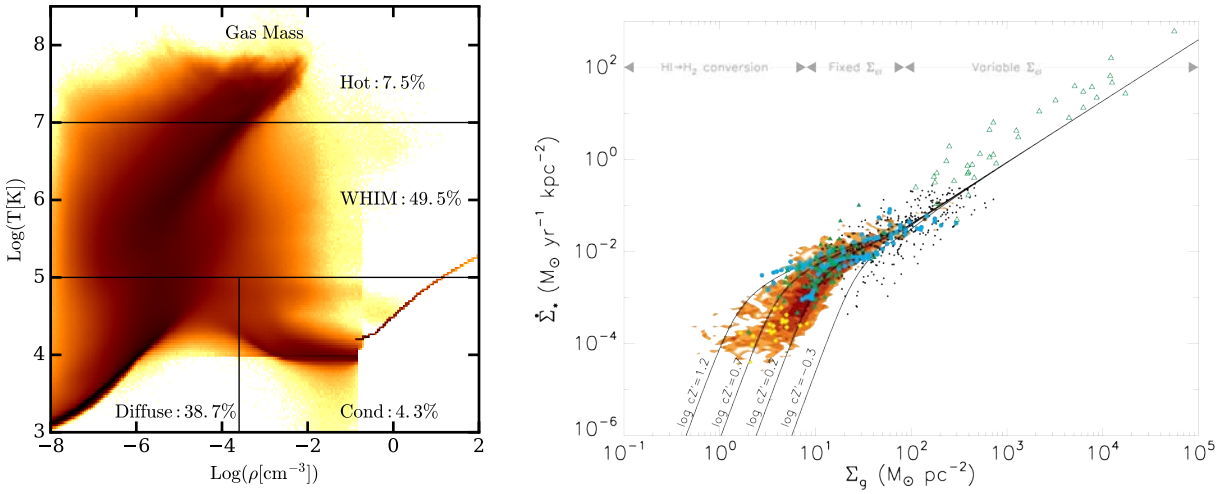
- Ultraviolet (UV) radiation from young stars:

$$\frac{\mathrm{SFR}_{\mathrm{UV}}}{\mathrm{M}_\odot/\mathrm{yr}} \sim \frac{L_{\mathrm{UV}}}{7.2 \times 10^{27} \mathrm{erg/s}} \tag{221}$$

We also have theoretical models, for example, the exponential model

$$(t) \propto e^{-t/\tau} \, . \tag{222}$$

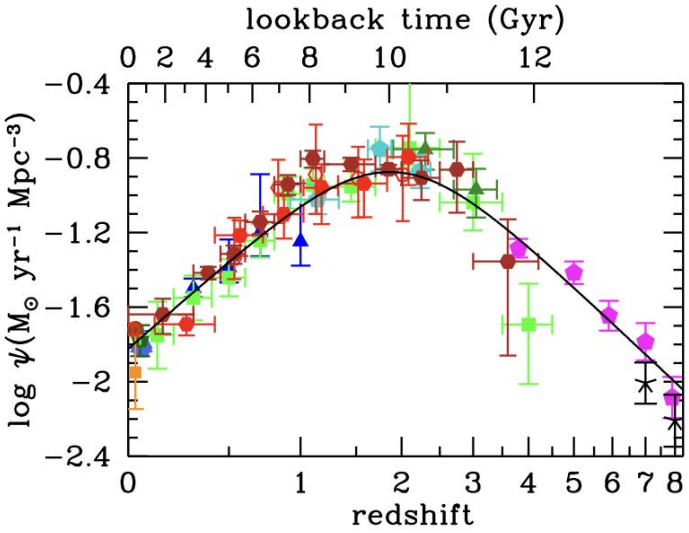
For a given galaxy, the star formation rate depends on the density and temperature of the gas. When gas is cold and dense, it is able to collapse into stars. The star formation rate can be roughly approximated by dividing the gas mass by the free-fall time.



The above plot on the left is from Torrey et al. 2019 shows a phase diagram of gas in the IllustrisTNG simulations at $z = 0$ and is split into various phases of the ISM. Darker regions show a higher gas mass, and the percentages show the total fraction of gas mass in each phase. The condensed material is in the lower right corner, and stars form along the thin line. The plot to the right is from Krumholz et al. 2009 shows the star formation rate surface density as a function of gas surface density. Each point is a different galaxy, compiled from several sources (different colors).

The star formation rate of a galaxy depends primarily on the molecular gas within a galaxy rather than the total gas. However, it can be difficult to predict what fraction of a galaxy's gas is in the molecular phase. This fraction depends on the total gas density, metallicity, and clumping on small scales. To get more precise predictions for star formation, it is also necessary to consider events such as supernovae and shocks.

Over cosmic history, the star formation rate (across all galaxies) started low and increased, peaked at $z \approx 2$, and has been decreasing since.



Initial stellar mass function:
$\phi(m)\mathrm{d}m$ is the relative number of stars born with masses in $(m, m + \mathrm{d}m)$. Note that the units are $[\phi] = \text{mass}^{-2}$, the number of stars formed per mass interval per total mass. This

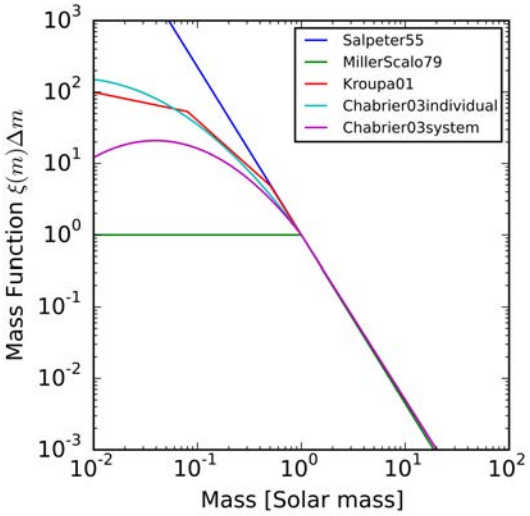
is normalized so

$$\int_{m_l}^{m_h} m\phi(m)\mathrm{d}m = 1\mathrm{M}_\odot \tag{223}$$

$m_l \sim 0.08\mathrm{M}_\odot$ since hydrogen fusion can't occur in stars lower than this and $m_h \sim 100\mathrm{M}_\odot$ since the Eddington limit prevents stars larger than this.

Example:
M_* is the total mass of newly formed stars. Then the total number $\mathrm{d}N(m)$ and total mass $\mathrm{d}M(m)$ of stars born in $(m, m+\mathrm{d}m)$ are

$$\begin{aligned} \mathrm{d}N(m) &= \frac{M_*}{\mathrm{M}_\odot}\phi(m)\mathrm{d}m \\ \mathrm{d}M(m) &= \frac{M_*}{\mathrm{M}_\odot}m\phi(m)\mathrm{d}m\ . \end{aligned} \tag{224}$$



The initial mass function is often assumed to follow the Salpeter mass function:

$$\phi(m) \propto m^{-(1+x)},\ x = 1.35\ . \tag{225}$$

There are other forms, like the Chabrier function, although the form of the initial mass function is uncertain since it depends on relating luminosity and mass and we observe the present day mass function, not the initial mass function.

Stellar spectra:
The stellar spectrum of a star is given by its luminosity L, effective temperature T_{eff}, and chemical composition z. The evolution of a star in the (L, T_{eff}) plane (stellar evolutionary or HR diagram track) only depends on the initial mass and initial metallicity. Once the initial mass and metallicity are known, one can calculate a stellar spectrum.

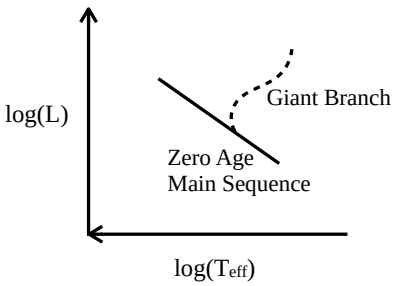
From the Stefan-Boltzmann law $L = 4\pi R^2\sigma T_{\mathrm{eff}}^4$, we can see that L and T are linearly related on the (logarithmic) HR diagram through R:

$$\log(L) = 2\log(R) + 4\log(T)\ . \tag{226}$$

Stars along the $R = 1$ line form the main sequence.

Stars move in the HR diagram as they age and go through the stellar stages (main sequence, red giant, white dwarf, etc.).
Time spend on the main sequence is

$$\boxed{\tau_{\rm MS} \propto M^{-3}} \,. \tag{227}$$



This comes from assuming that the lifetime of a star depends on how much fuel it has (its mass) and how fast it burns that fuel (rate of energy burning, luminosity)

$$\tau_{\rm MS} \propto \frac{M}{L} \,. \tag{228}$$

From the observed mass luminosity relationship, $L \propto M^4$ so

$$\tau_{\rm MS} \propto \frac{M}{L} \propto \frac{M}{M^4} \propto M^{-3} \,. \tag{229}$$

For low mass stars ($M < 0.7\,{\rm M}_\odot$), $L \propto M^3$ so $\tau_{\rm MS} \propto M^{-2}$.

Population synthesis:
A galaxy spectrum is a superposition of stellar spectrum:

$$L_\lambda = \int_0^t \mathcal{L}_\lambda^{\rm cp}(t - t', Z(t'))\psi(t'){\rm d}t' \,. \tag{230}$$

We can measure time from t' that the stars formed so $\tau = t - t'$ and $\tau_0 = t'$. $\psi(t')$ is the star formation rate at t'. $\mathcal{L}_\lambda$ is the luminosity at λ per unit stellar mass of all stars of a coeval population of age τ with initial metallicity $Z(\tau_0)$:

$$\mathcal{L}_\lambda^{\rm cp}(\tau, Z(\tau_0)) = \int \mathcal{L}_\lambda(m, Z(\tau_0), \tau)\frac{\phi(m)}{M_\odot}{\rm d}m \tag{231}$$

where $\mathcal{L}_\lambda$ is the luminosity at wavelength λ of a star with initial mass m and initial metallicity $Z(\tau_0)$ at time τ.

A few notes:

- $L_\lambda(t)$ is a convolution of ϕ, ψ, and $\mathcal{L}_\lambda$.
- ϕ and ψ are not known precisely.
- There are sophisticated codes available to numerically iterate to figure out ϕ and ψ:
  - assume an initial mass function ϕ
  - impose a star formation rate
  - run convolution
  - compare with data (can break some degeneracies with spectral features)
  - adjust SFR and IMF and repeat.

3.F Chemical evolution of galaxies

We have used stellar population synthesis to constrain the initial mass function and the star formation rate. Chemical evolution can also be used to learn about the baryonic history of a galaxy. We use heavy elements as a chronometer. The general chemical evolution follows:

- $t = 0$: no heavy elements
- stellar nucleosynthesis generates heavy elements
- supernovae eject heavy elements into the interstellar medium
- heavy elements are incorporated into new stars.

The *metallicity* Z of a star is

$$Z = \frac{\text{mass of heavy elements}}{\text{total mass}} \tag{232}$$

and is often quoted as a fraction of the solar metallicity $Z/\mathrm{Z}_\odot$ with $\mathrm{Z}_\odot \approx 0.02$.

The abundance $\left[\frac{X}{Y}\right]$ of elements is a comparison between two elements X and Y, e.g. $\left[\frac{\mathrm{Fe}}{\mathrm{H}}\right]$. We report it as the fraction of the log of the solar abundance:

$$\left[\frac{X}{Y}\right] = \log\left(\frac{n_x/n_y}{(n_x/n_y)_\odot}\right) \tag{233}$$

so $\left[\frac{X}{Y}\right] = 0$ means the star has the same abundance as the sun, $\left[\frac{X}{Y}\right] = -1$ means the star has $\frac{1}{10}$ of the solar abundance, $\left[\frac{X}{Y}\right] = -2$ means the star has $\frac{1}{100}$ of the solar abundance, etc.

Note that the metallicity measures by mass and abundance measures by number.

Modelling chemical evolution:

M is the total mass

M_s is the mass in stars

M_g is the mass in gas

f is the gas inflow rate

e is the gas outflow rate

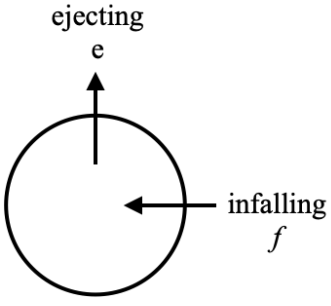
ψ is the star formation rate

E is the gas ejection rate

E_Z is the ejection rate of metals from stars, supernovae, etc.

$Z_f f$ is the infalling metals per time

ZM_g is the mass of metals in gas



$$\begin{aligned} M &= M_s + M_g \\ \frac{\mathrm{d}M}{\mathrm{d}t} &= f - e \\ \frac{\mathrm{d}M_g}{\mathrm{d}t} &= \psi - E \\ \frac{\mathrm{d}M_g}{\mathrm{d}t} &= -\psi + E + f - e \\ \frac{\mathrm{d}(ZM_g)}{\mathrm{d}t} &= -Z\psi + E_Z + Z_f f - Ze \end{aligned} \tag{234}$$

This forms a complete chemical model. We can figure out individual terms and then solve. We will use some approximations for an analytical solution.

$$E(t) = \int_{m_t}^{\infty} (m - w_m)\psi\left(t - \tau_{\rm MS}(m)\right)\phi(m)\mathrm{d}m \tag{235}$$

m_t: Main sequence turnoff mass. This is the lowest mass of stars dying at time t.

$m - w_m$: The ejected mass; w_m is the remnant mass.

$\tau_m(m)$: Main sequence lifetime at mass m.

${}_{t-\tau_m(m)}\phi(m)$: Birth rate of stars of mass m at time $t-\tau_m(m)$, which is the death rate at time t.

$$E_z(t) = \int_{m_t}^{\infty} \left[(m - w_m)Z(t - \tau_{\rm MS}(m)) + m\rho_{Zm}\right]\psi(t - \tau_{\rm MS}(m))\phi(m)\mathrm{d}m \tag{236}$$

$(m - w_m)z(t - \tau_{\rm MS}(m))$: mass of metals that at time $t - \tau_{\rm MS}(m)$ were locked in a star of mass m and are now ejected with the envelop at time t.

$m\rho_{Zm}$: new metals produced by a star of mass m. (Note: some elements get destroyed, for example lithium has a $\rho_{zm} < 0$.)

Instantaneous recycling approximation: (IRA)
We assume that the mass and elements of stars are returned to the interstellar medium without delay and the ejecta are fully mixed immediately. This only works for massive enough stars, $m > m_{\rm lim}$.

$$\psi(t - \tau_m(m)) \approx \psi(t) \tag{237}$$

then

$$E(t) \approx \psi(t)\int_{m_{\rm lim}}^{\infty} \left[m - w_m\right]\phi(m)\mathrm{d}m = \psi(t)R\,. \tag{238}$$

Stars below $m_{\rm lim}$ never lose mass while stars greater than $m_{\rm lim}$ immediately lose mass. This is because massive stars get off the main sequence so quickly ($\tau_{\rm MS} \propto M^{-3}$) so this process is essentially instantaneous, $\tau_{\rm MS} \approx 0$. R is the returned mass per star formed:

$$R = \int_{m_{\rm lim}}^{\infty} (m - w_m)\phi(m)\mathrm{d}m\,. \tag{239}$$

Then

$$\begin{aligned} E_z(t) &\approx \int_{m_{\rm lim}}^{\infty} \left[(m - w_m)Z(t) + m\rho_{Zm}\right]\phi(m)\mathrm{d}m \\ &= \psi(t)Z(t)R + \psi(t)\int_{m_{\rm lim}}^{\infty} m\rho_{Zm}\phi(m)\mathrm{d}m \\ &= \psi(t)Z(t)R + (1 - R)y\psi(t) \end{aligned} \tag{240}$$

where y is the mass of produced metals per remnant mass (white dwarfs, neutron stars, etc.)

$$y = \frac{1}{1-R}\int_{m_{\rm lim}}^{\infty} m\rho_{Zm}\phi(m)\mathrm{d}m \ . \tag{241}$$

This gives us the equations of chemical evolution in the instantaneous recycling approximation:

$$\begin{aligned} M &= M_s + M_g \\ \frac{\mathrm{d}M}{\mathrm{d}t} &= f - e \\ \frac{\mathrm{d}M_s}{\mathrm{d}t} &= (1-R)\psi(t) \\ \frac{\mathrm{d}M_g}{\mathrm{d}t} &= -(1-R)\psi(t) + f - e \\ \frac{\mathrm{d}(ZM_g)}{\mathrm{d}t} &= -z\psi + RZ(t)\psi(t) + (1-R)y\psi(t) + Z_f f - Ze \\ &= (1-R)(-Z+y)\psi + Z_f f - Ze \end{aligned} \tag{242}$$

We can combine the last two equations for $\frac{\mathrm{d}M_g}{\mathrm{d}t}$ and $\frac{\mathrm{d}(ZM_g)}{\mathrm{d}t}$ to get

$$M_g\frac{\mathrm{d}Z}{\mathrm{d}t} = (1-R)y\psi(t) + (Z_f - Z)f + Ze \ . \tag{243}$$

Closed-box model:
The simplest evolution model is to assume a closed box ($f = e = 0$) containing only gas ($M_g(0) = M, M_s(0) = 0$ with zero metallicity ($Z(0) = 0$)). The equations then simplify:

$$\begin{aligned} M &= M_s + M_g \\ \frac{\mathrm{d}M}{\mathrm{d}t} &= 0 \\ \frac{\mathrm{d}M_s}{\mathrm{d}t} &= (1-R)\psi(t) \\ \frac{\mathrm{d}M_g}{\mathrm{d}t} &= -(1-R)\psi(t) \\ \frac{\mathrm{d}(ZM_g)}{\mathrm{d}t} &= (1-R)(-Z+y)\psi \\ M_g\frac{\mathrm{d}Z}{\mathrm{d}t} &= (1-R)y\psi(t) \ . \end{aligned} \tag{244}$$

We can divide $\frac{\mathrm{d}M_g}{\mathrm{d}t}$ by $M_g\frac{\mathrm{d}Z}{\mathrm{d}t}$ to get

$$\frac{1}{M_g}\frac{\mathrm{d}M_g}{\mathrm{d}Z} = -\frac{1}{y} \tag{245}$$

and integrate

$$\begin{aligned} \ln(M_g)\big|_M^{M_g(t)} = \ln\left(\frac{M_g(t)}{M}\right) &= \int_0^{Z(t)} -\frac{\mathrm{d}Z}{y} = -\frac{Z(t)}{y} \\ \Rightarrow Z(t) &= y\ln\left(\frac{M}{M_g(t)}\right) \end{aligned} \tag{246}$$

so we have:

$$\boxed{Z(t) = y \ln\left(\frac{M_g(t=0)}{M_g(t)}\right)} \tag{247}$$

which is the metallicity of gas as a function of only $M_g(t)$.

Metallicity of stars:
In the closed box model, stars and gas must contain all the metals ever produced.

$$Z_s M_s + Z M_g = \int_0^t \int_0^\infty m \rho_{Zm} \psi(t') \phi(m) \mathrm{d}m \mathrm{d}t' \tag{248}$$

where Z_s is the average metallicity of stars and the right side of the equation represents the mass of all metals injected into the interstellar medium until time t. Then

$$Z_s M_s + Z M_g = \int_0^t (1-R) y \psi(t') \mathrm{d}t' \approx (1-R) y \bar{\psi}(t) t \ . \tag{249}$$

We can integrate our equation

$$\frac{\mathrm{d}M_s}{\mathrm{d}t} = (1-R)\psi(t) \tag{250}$$

to get the mass

$$M_s = (1-R)\bar{\psi}(t)t \tag{251}$$

which matches the second equality in equation above. This makes sense since $\bar{\psi}(t)$ is the total stellar mass and $(1-R)$ subtracts the remnants. We can substitute $u = M_g/M$ to get

$$Z_s = y - Z\frac{M_g}{M - M_g} = y - Z\frac{u}{1-u} \ . \tag{252}$$

So we finally get

$$\boxed{\begin{aligned} \text{gas} : Z(s) &= y \ln\left(\frac{1}{u(t)}\right) \\ \text{stars} : Z_s(t) &= y - Z(t)\frac{u(t)}{1-u(t)} \end{aligned}} \tag{253}$$

where u is the gas fraction M_g/M and M is constant for the closed box model. As $u \to 0, Z_s \to y$, which gives the typical metallicity of stars. This must be less than or equal to the typical yield.

G-dwarf problem:
We want to measure the metallicity distribution of G stars. These stars have not evolved much and are still on the main sequence. Their age is so high that they formed from a very low metallicity gas, since $Z(0) = 0$. We can use the closed box result to predict their metallicity distribution. However, we cannot use an average Z_s since we are looking for a distribution.

We apply the closed-box model:

$$M_s(\geq u) = (1-u)M \Rightarrow \frac{M_s(\geq u)}{M_{s,0}} = \frac{1-u}{1-u_0} \tag{254}$$

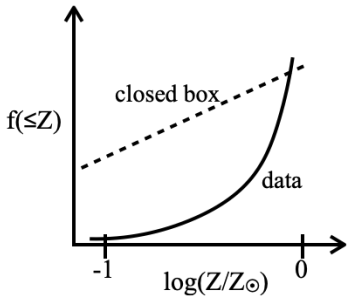
where $M_s(\geq u)$ is the mass of stars formed while the gas fraction was $\geq u$ and $...{}_0$ refer to present-day values.

Then the stellar mass fraction $M_s(\geq u)/M$ was made from gas with $Z \leq y\ln(1/u)$ and we can rewrite the stellar mass fraction using $u = e^{-Z/y}$ and $u_0 = e^{-Z_0/y}$:

$$\frac{M_s(\leq Z)}{M_{s,0}} = \frac{1-e^{-Z/y}}{1-u_0} = \frac{1-u_0^{Z/Z_0}}{1-u_0}\ . \tag{255}$$

Then we can get the fraction of stellar mass with metallicity $\leq Z$:

$$f(\leq Z) = \frac{M_s(\leq Z)}{M_{s,0}} = \frac{1-u_0^{Z/Z_0}}{1-u_0}\ . \tag{256}$$



However, the model does not agree well with the data because the model is incomplete (infalls, variations in the IMF, etc.).

3.G Active galaxies (AGN)

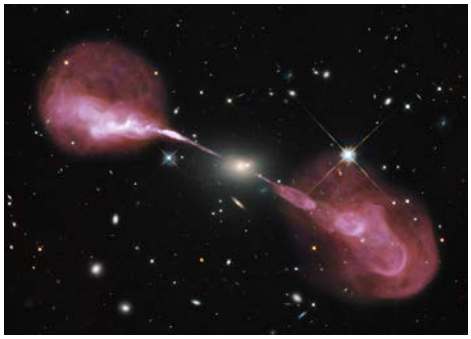
AGN is Active Galactic Nucleus.
Definition:

- Galaxies whose total luminosity is dominated by radiation not produced in stars. Stars produce near-UV, optical, and near-IR light in blackbodies. Other sources may emit radio or X-ray light.
- The energy generation is associated with a point-like source at the nucleus of the galaxy ($\sim$ black hole with mass $10^6 - 10^9 M_\odot$).

AGN types:

- Radio galaxies:

  - high radio luminosity $L_{\text{radio}} \geq 10^{18} L_{\odot}$
  - radio emission from two external regions (radio lobes)
  - energized by jets (particle acceleration $E_e \sim 10^{12}$ eV
  - $\sim 50\%$ E0/S0 galaxies, $\sim 50\%$ quasars
  - synchrotron emission of electrons



Hercules A galaxy

NASA, ESA, S. Baum and C. O'Dea (RIT), R. Perley and W. Cotton (NRAO/AUI/NSF), and the Hubble Heritage Team (STScI/AURA)

- Quasars/QSO:
  - Quasar (quasi-stellar radio source): optical point source with radio jet
  - QSO (quasi-stellar object): like a quasar but no radio emission
  - Quasars and QSO's are similar phenomena, 90% of optically found QSO's are radio quiet, $\sim 10\%$ are radio loud
  - mostly found in elliptical galaxies
  - $L_{\text{quasar}} \sim 10^{45-48}$ erg/s
  - synchrotron jets between 0.1 pc-1 Mpc
  - maximum space density $\sim z = 2 - 3$

Einstein Cross gravitational lens

J.Rhoads, S.Malhotra, I.Dell'Antonio (NOAO)/WIYN/ NOIRLab/NSF.

- BL Lac objects:
  - quasar with enhanced continuum emission
  - highly variable
  - extremely luminous
  - highly polarized
  - jet pointing towards observer

Markarian 501 galaxy

Sloan Digital Sky Survey on Wikimedia Commons. License CC-BY.

- Seyfert galaxies:

- spiral galaxies
- bright unresolved nuclei (less luminous than quasars)
- $L \approx 10^{42} - 10^{45}$ erg/s

Spanish Dancer galaxy

Structure of AGN physics:
Sizes: changes of state of the emission region propagate at maximum speed c. Variability means state change:

$$\Delta t_{\text{variable}} c \sim r_{\text{emission}} \tag{257}$$

In the radio/optical band:

$$\Delta t_{\text{variable}} \sim 1 - 10 \text{ days}$$
$$r_{\text{emission}} \sim 10^{-3} - 10^{-2} \text{ pc}$$

At TeV energies:

$$\Delta t_{\text{variable}} \sim 1 \text{ day}$$
$$r_{\text{emission}} \sim 10^{-3} \text{ pc}$$

We can compare this to the Schwarzschild radius for a black hole with mass $M_\bullet$:

$$R_s = \frac{2GM_\bullet}{c^2} \tag{258}$$

which gives a size for various masses:

$$\begin{aligned} M_\bullet &= 10^6 M_\odot \rightarrow R_s = 10^{-7} \text{ pc} \\ M_\bullet &= 10^7 M_\odot \rightarrow R_s = 10^{-6} \text{ pc} \\ M_\bullet &= 10^9 M_\odot \rightarrow R_s = 10^{-4} \text{ pc} \end{aligned} \tag{259}$$

which gives the variability in the vicinity of a supermassive black hole.

We can consider various possible energy sources for the observed variability:

- Stars: $N_* = 3 \times 10^8$ O-type stars in the central region to get the necessary luminosity (O stars have luminosity of $\sim 10^{5.5} L_\odot$), but this leads to a stellar density that is too high and would be unstable.
- Supernovae: the energy of a supernova is $E_{\text{SN}} \sim 10^{52}$ erg, so we would need 10^{10} supernovae within 10^{-3} pc in 10^7 years. This would require producing 10^{10} stars continually, which has the same problem as the source being stars (too dense and unstable).
- We need accretion onto a supermassive black hole to create the luminosity.

Accretion onto supermassive black holes (SMBH):
Idea: a supermassive black hole ($M_\bullet \sim 10^6 - 10^{9.5}\mathrm{M}_\odot$) accretes $10^{-4} - 10\,\mathrm{M}_\odot/\,\mathrm{yr}$. Jets and nonthermal radiation are created by the accretion disk (gravitational energy is converted to thermal energy and radiation).

The radiative efficiency of this process is $\frac{1}{16}$, so the luminosity of the accretion disk is approximately given by:

$$L_{\mathrm{acc}} \approx \frac{1}{16}\dot{m}c^2 \tag{260}$$

which means that 1 g of material produces approximately 10^6 kwh. We can compare the efficiency of an accretion disk ($\frac{1}{16}$) to the efficiency of hydrogen burning, which is 0.007 (so $L_{\mathrm{H-burning}} \approx 0.007\dot{m}c^2$).

The *Eddington luminosity* is the maximum possible AGN luminosity, which is reached when the radiation pressure exceeds the gravitational acceleration per area. This comes from processes like Thomson scattering. The radiation pressure is given by

$$P_\gamma = \frac{E}{c} = \frac{h\nu}{c}\ . \tag{261}$$

We can write the momentum per time (equivalent to force) as L/c, so the pressure is force per area

$$P_{\mathrm{total}} = \frac{L}{4\pi r^2 c}\ . \tag{262}$$

We find where the radiative force on a fully ionized plasma (i.e. the force on an e^-) exceeds the gravitational force on a proton for a black hole of mass $M_\bullet$:

$$\begin{aligned} F_{\mathrm{rad}} &> F_{\mathrm{grav}} \\ \frac{L}{4\pi r^2 c}\sigma_T &> \frac{GM_\bullet m_p}{r^2} \end{aligned} \tag{263}$$

where σ_T is the Thomson cross section, so the radiative force on an electron is $F_{\mathrm{rad}} = \sigma_T P_{\mathrm{total}}$. This gives the Eddington luminosity

$$\boxed{L_{\mathrm{edd}} = \frac{4\pi c G M_\bullet m_p}{\sigma_T} = 1.3\times 10^{38}\frac{M_\bullet}{M_\odot}\mathrm{erg/s}}\ . \tag{264}$$

To achieve AGN luminosities, the SMBH must be massive enough to to be blown apart.

This leads to the *Eddington accretion rate*, which is the maximal possible accretion rate possible for an accretion disk. This is reached for when L_{acc} exceeds L_{edd}:

$$\begin{aligned} L_{\mathrm{acc}} &> L_{\mathrm{edd}} \\ \frac{1}{16}\dot{m}c^2 &> 1.3\times 10^{38}\frac{M_\bullet}{M_\odot}\mathrm{erg/s} \end{aligned} \tag{265}$$

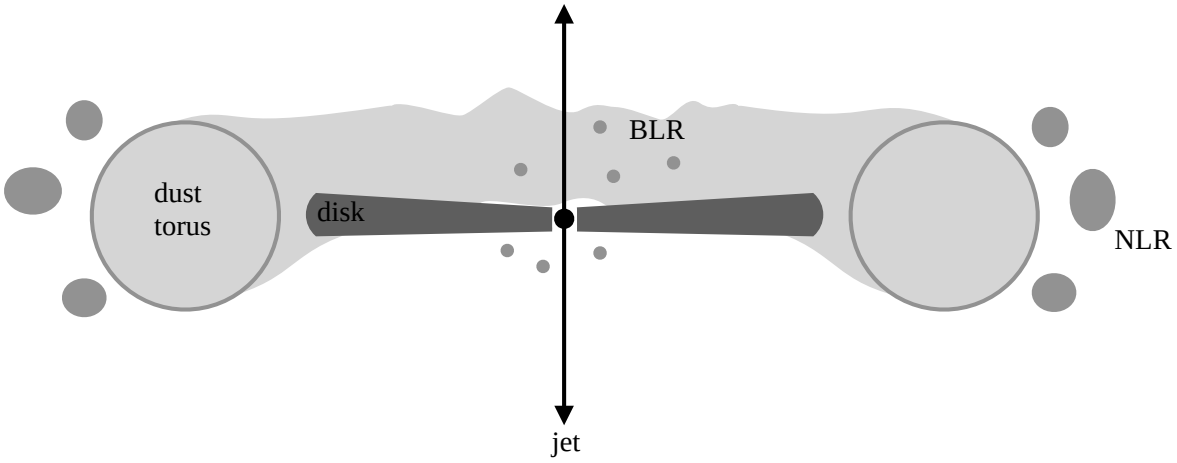
so $\dot{m}_{\mathrm{edd}}$ occurs when the two are equal:

$$\dot{m}_{\mathrm{edd}} = 5\times 10^{-10}\frac{M_\bullet}{M_\odot}\frac{M_\odot}{\mathrm{yrs}}\ . \tag{266}$$

An accretion disk is formed when gas spirals in from large distances until the innermost stable orbit (ISCO). Viscous processes in the disk lead to heating to temperatures $T \sim 10^8\,\mathrm{K}$. This is highly efficient in releasing energy.

Unified model of AGN:
Different AGN types are manifestations of the same phenomenon:

- SMBH at the center with $M_\bullet \sim 10^6 - 10^{10}\mathrm{M}_\odot$
- An accretion disk extending to $\sim 100 - 1000 r_s$ emits the X-ray, UV, optical, and TeV radiation
- Jets are made of radio synchrotron radiation from strong magnetic fields
- A dust torus from $\sim 1\,\mathrm{pc}$ to $\sim 50 - 100\,\mathrm{pc}$ produces IR emission
- Broad line region (BLR) formed from clouds of thick gas within $\sim 0.1 - 1\,\mathrm{pc}$ (velocities are faster near the black hole, $v \sim 10^4\,\mathrm{km/s}$
- Narrow line region (NLR) formed from clouds of thin gas within ∼few pc (farther away from the black hole, $v \sim 100 - 1000\,\mathrm{km/s}$. Slower velocities leads to less broadening of lines scattered in the clouds, hence narrow line region).



The observed manifestation depends on the viewing angle and the accretion rate. For example, BL Lac objects have a line of sight directly down the jet and a high accretion rate.

Part II

Cosmology and Structure Formation

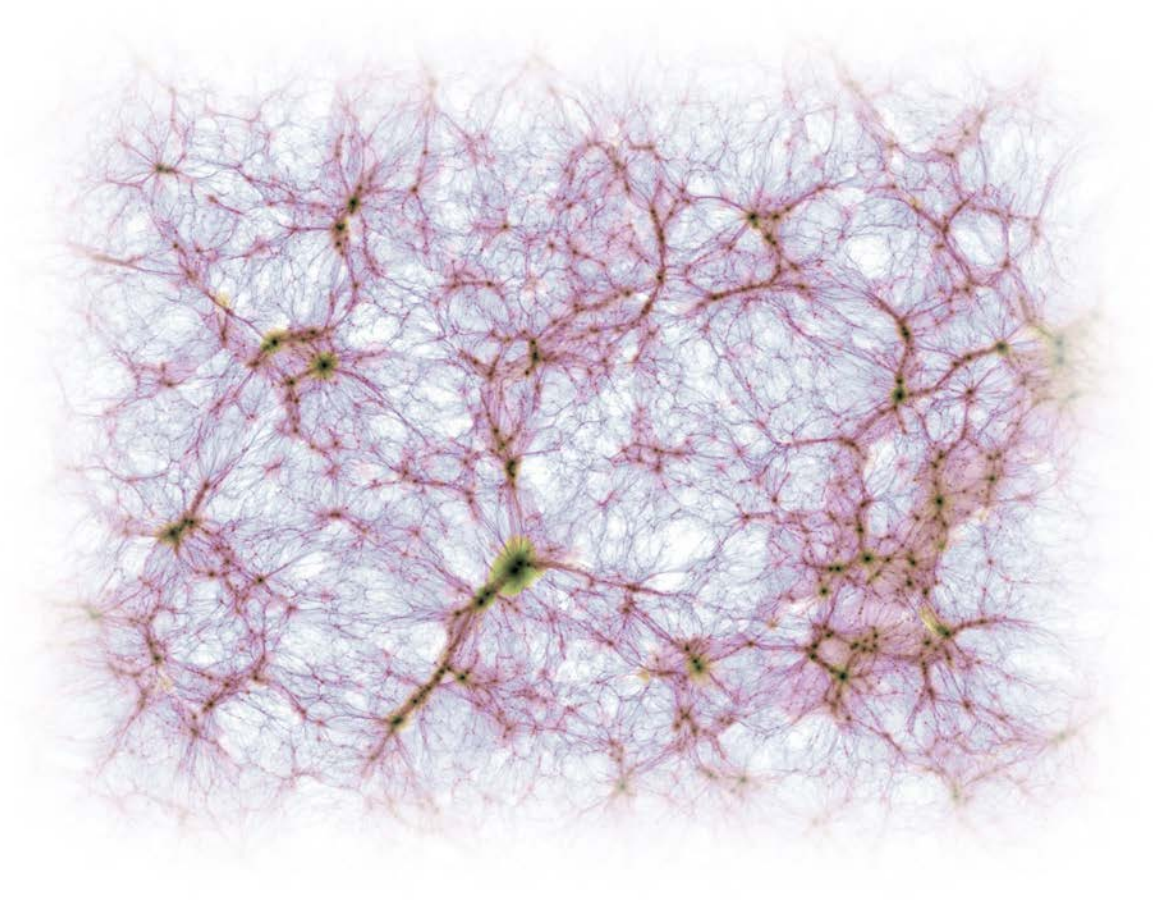


Image: TNG Simulations

1 Cosmology

Cosmology is the study of dynamics of the entire Universe as a single dynamical system.

1.A Cosmological Principle and dynamics

- The Universe is homogeneous: it is uniform on large scales.
- The Universe is isotropic: it looks the same for all observers on large scales.

This implies that the space-time metric is the same everywhere, which generates symmetries and simplifies the solutions to general relativity equations.

Hubble Law:
We observe that

$$\vec{v} = H_0 \vec{r} \tag{267}$$

where $H_0 \sim 70\,\mathrm{km/s/Mpc}$ refers to the present-day Hubble factor. The specific form of this law can be derived from the cosmological principles:

- Linearity (follows from isotropy):
  Suppose $\vec{v} = f(\vec{r})$.
  Then from Observer A's perspective

  $$\vec{v}_1 = f(\vec{r}_1) \text{ and } \vec{v}_2 = f(\vec{r}_2)$$

  $$\text{and } \vec{v}_1 - \vec{v}_2 = f(\vec{r}_1) - f(\vec{r}_2)\ .$$

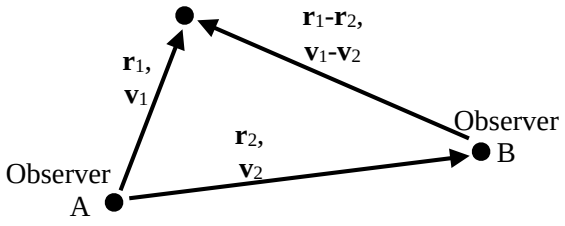
  From Observer B's perspective

  $$\vec{v}_1 - \vec{v}_2 = f(\vec{r}_1 - \vec{r}_2)$$

  so we find that

  $$f(\vec{r}_1 - \vec{r}_2) = f(\vec{r}_1) - f(\vec{r}_2)\ .$$

  This implies that f is linear since isotropy requires that each observer sees the same Hubble law.



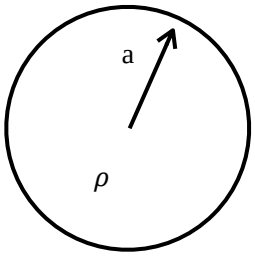
- Uniqueness (follows from homogeneity):
  $f(...)$ is linear, so

  $$f(\vec{r}) = \underline{\underline{\mathbf{H}}}\vec{r} \tag{268}$$

  where $\underline{\underline{\mathbf{H}}}$ is a matrix.
  We assume that it is non-diagonal, otherwise a special direction would be preferred (since it introduces an axis), and $\underline{\underline{\mathbf{H}}} = H_0 \mathbb{1}$. Then

  $$f(\vec{r}) = H_0 \vec{r}\,. \tag{269}$$

Dynamics of cosmological expansion:
Hubble's law implies cosmological expansion. General relativity allows a detailed derivation of the dynamics, but here we use Birkhoff's theorem to get initial insight. Birkhoff's theorem states that the dynamics of a uniform expanding self-gravitating sphere is equivalent to a section of the Universe as a whole.



$$\begin{aligned} M(<a) &= \frac{4}{3}\pi a^3 \rho \\ \Rightarrow \ddot{a} &= -\frac{GM(<a)}{a^2} = -\frac{4\pi G\rho}{a^2}\cdot\frac{1}{3}a^3 \end{aligned} \tag{270}$$

We multiply each side by $\dot{a}$:

$$\begin{aligned} \ddot{a}\dot{a} &= -\frac{4\pi G\rho}{3}a\dot{a} \\ &= -\frac{4\pi G\rho_0}{3}a_0^3 a^{-2}\dot{a} \ \ (\text{since } \rho = \rho_0\frac{\mathrm{a}_0^3}{\mathrm{a}^3}) \ . \end{aligned} \tag{271}$$

Since $\ddot{a}\dot{a} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\dot{a}^2\right)$ and $a^{-2}\dot{a} = \frac{\mathrm{d}}{\mathrm{d}t}\left(-\frac{1}{a}\right)$, we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\dot{a}^2\right) = -\frac{4\pi G\rho_0 a_0^3}{3}\frac{\mathrm{d}}{\mathrm{d}t}\left(-\frac{1}{a}\right) \ . \tag{272}$$

Integrate $\int dt$ on both sides:

$$\frac{1}{2}\dot{a}^2 = \frac{4\pi G\rho_0 a_0^3}{3}\frac{1}{a} + \tilde{\kappa} \ . \tag{273}$$

Here, $\tilde{\kappa}$ is the integration constant, and we can use the density expression $\rho_0 a_0^3 = \rho a^3$ to simplify our equation:

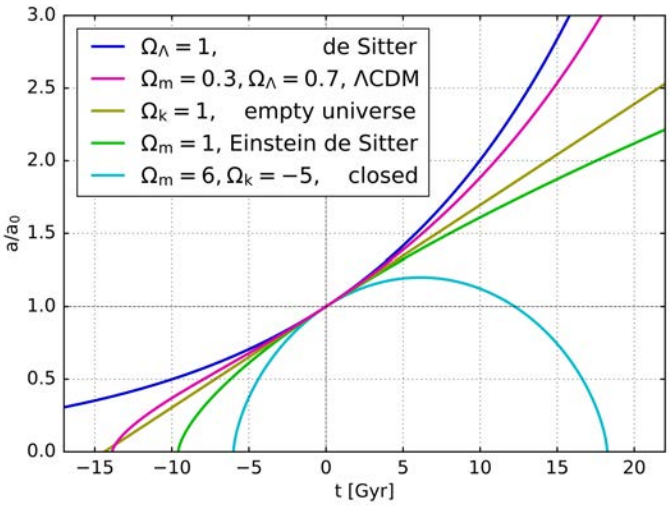
$$\frac{1}{2}\dot{a}^2 = \frac{4\pi G}{3}\rho a^2 + \tilde{\kappa} \ . \tag{274}$$

So the dynamics is given by

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\tilde{k}}{a^2} \ . \tag{275}$$

This is the Friedmann equation for $\Lambda = 0$. With a cosmological constant Λ, we have

$$\boxed{\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\tilde{k}}{a^2} + \frac{\Lambda}{3}} \ . \tag{276}$$



Here, a is the scale factor such that length $l = l_0 \frac{a}{a_0}$. Also note that:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\dot{r}}{r} = \frac{v}{r} = H \quad \left(H(t) \equiv \frac{\dot{a}}{a}\right)$$

$\frac{8\pi G}{3}\rho$ is the matter/radiation density

$\frac{\tilde{\kappa}}{a^2}$ is the curvature

$\frac{\Lambda}{3}$ is the cosmological constant

Since volume grows with the length cubed and the total mass in the universe is constant, the matter density is proportional to a^{-3}. The radiation density also decreases due to the increasing volume but also decreases as the wavelengths are stretched, so radiation density is proportional to a^{-4}. The dark energy Λ is constant.

The plot above shows how the scale factor grows with time for several different types of universes. Our current understanding of our universe is that it is described by the ΛCDM model, where roughly 30% of the energy budget is matter, 70% is dark energy, and there is a very small amount of radiation and no curvature. If there were positive or negative curvature, we would get an open ore closed universe. There are also several toy universes that are often useful to think about. A flat, dark energy-only universe is the de Sitter model and a flat, matter-only universe is the Einstein-de Sitter model An empty universe has only a curvature term, and is an open universe.

Dynamical evolution of the Universe:
Different terms in the Friedmann equation dominate at different times.

- radiation term $\propto a^{-4} \implies$ dominates at very early times
- matter term $\propto a^{-3} \implies$ dominates at early times
- curvature term $\propto a^{-2} \implies$ dominates at medium times
- Λ term $\propto$ constant $\implies$ dominates at late times

Therefore, from the Friedmann equation, we can derive different regimes of the Universe:

- radiation regime

$$\begin{array}{l} \dot{a}^2 \propto a^{-2} \\ \dot{a} \propto a^{-1} \\ a\mathrm{d}a \propto \mathrm{d}t \end{array} \quad \Longrightarrow \quad \begin{array}{l} a \propto t^{\frac{1}{2}} \\ H(t) = \frac{\dot{a}}{a} = \frac{1}{2t} \end{array} \quad \Longrightarrow \quad t_0 = \frac{1}{2}\frac{1}{H_0}$$

- matter regime

$$\dot{a}^2 \propto a^{-1} \quad \dot{a} \propto a^{-\frac{1}{2}} \quad \sqrt{a}\mathrm{d}a \propto \mathrm{d}t \implies a \propto t^{\frac{2}{3}} \quad H(t) = \frac{\dot{a}}{a} = \frac{2}{3t} \implies t_0 = \frac{2}{3}\frac{1}{H_0}$$

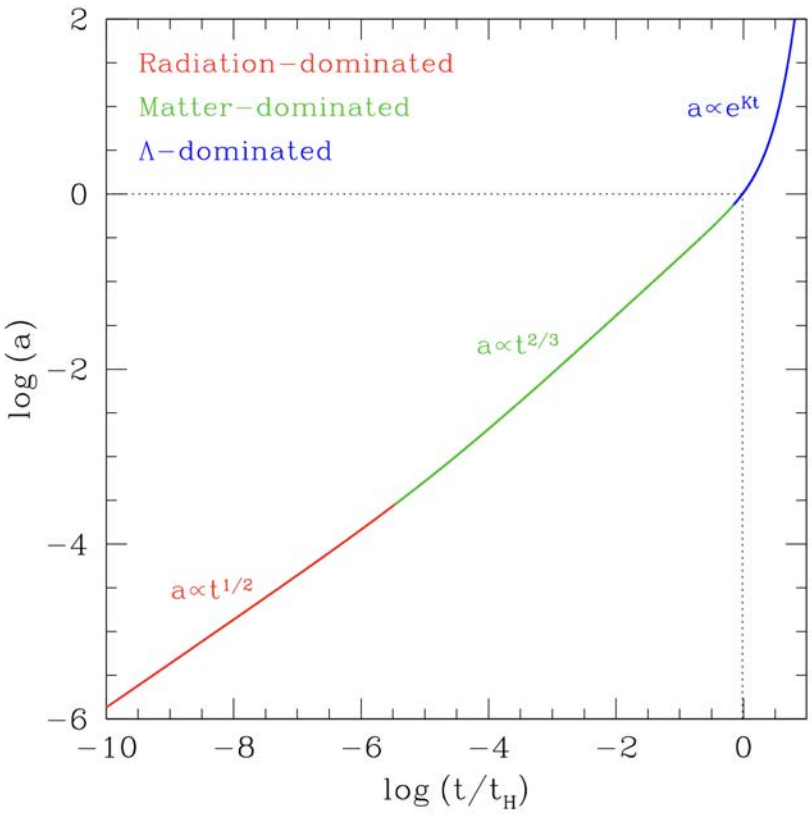
- curvature regime

$$\dot{a}^2 \propto \text{constant} \quad \dot{a} \propto \text{constant} \quad \mathrm{d}a \propto \mathrm{d}t \implies a \propto t \quad H(t) = \frac{\dot{a}}{a} = \frac{1}{t} \implies t_0 = \frac{1}{H_0}$$

- Λ regime

$$\dot{a}^2 \propto a^2\frac{\Lambda}{3} \quad \dot{a} \propto a\sqrt{\frac{\Lambda}{3}} \quad \frac{\mathrm{d}a}{a} \propto \sqrt{\frac{\Lambda}{3}}\mathrm{d}t \implies a \propto e^{\sqrt{\frac{\Lambda}{3}}t} \implies \text{exponential growth}$$

As the universe evolves, it expands at different rates depending on the regime (radiation/matter/Λ).



1.B Dynamics derived with general relativity

Goal: use the field equation

$$G_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} - \Lambda g_{\mu\nu} \tag{277}$$

to derive the Friedmann equation.
$G_{\mu\nu}$: Einstein tensor; 1st and 2nd derivatives of the metric

$T_{\mu\nu}$: stress-energy tensor
$g_{\mu\nu}$: metric (similar to Poisson's equation $\bigtriangledown^2\Phi = 4\pi G\rho + \frac{\Lambda}{3}$ with Φ replaced with curvature)

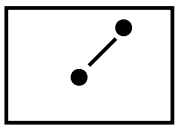
First we need to specify $g_{\mu\nu}$ and $T_{\mu\nu}$.

Metrics:
The space-time interval is

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \,. \tag{278}$$

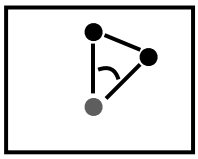
Some examples of spatial metrics:

- 2D-flat space in Cartesian coordinates:



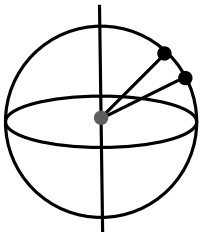
$$\mathrm{d}s^2 = \begin{pmatrix}\mathrm{d}x & \mathrm{d}y\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mathrm{d}x\\ \mathrm{d}y\end{pmatrix} = \mathrm{d}x^2 + \mathrm{d}y^2 \tag{279}$$

- 2D-flat space polar coordinates:



$$\mathrm{d}s^2 = \begin{pmatrix}\mathrm{d}r & \mathrm{d}\theta\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & r^2\end{pmatrix}\begin{pmatrix}\mathrm{d}r\\ \mathrm{d}\theta\end{pmatrix} = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 \tag{280}$$

- 2D-curved space:



$$\mathrm{d}s^2 = \begin{pmatrix}\mathrm{d}\theta & \mathrm{d}\varphi\end{pmatrix}\begin{pmatrix}R^2 & 0\\ 0 & R^2\sin^2\theta\end{pmatrix}\begin{pmatrix}\mathrm{d}\theta\\ \mathrm{d}\varphi\end{pmatrix} = R^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi) \tag{281}$$

We can rewrite $\chi = R\theta$, so

$$\mathrm{d}s^2 = \mathrm{d}\chi^2 + R^2\sin^2\frac{\chi}{R}\mathrm{d}\varphi^2 \,. \tag{282}$$

When R goes to infinity, we have

$$\sin\left(\frac{\chi}{R}\right) \approx \frac{\chi}{R} \tag{283}$$

$$\Rightarrow \mathrm{d}s^2 = \mathrm{d}\chi^2 + \chi^2\mathrm{d}\varphi^2 \tag{284}$$

which is gives us flat space!

- 4D space time:

$$g_{\mu\nu} = \begin{pmatrix}-1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix} \tag{285}$$

and

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \,. \tag{286}$$

Robertson-Walker metric:
The metric form follows from homogeneity and isotropy, and the field equations give us the time evolution:

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a(t)\left[\mathrm{d}\chi^2 + f_k^2(\chi)\right]\left(\mathrm{d}\theta^2 + sin^2\theta\mathrm{d}\varphi^2\right) \tag{287}$$

and

$$f_k(\chi) = \begin{cases} k^{-1/2}\sin\left(k^{1/2}\chi\right), & \text{closed k} > 0 \\ \chi, & \text{flat k} = 0 \\ |k|^{-1/2}\sinh\left(|k|^{1/2}\chi\right), & \text{open k} < 0 \end{cases} \tag{288}$$

with the units for k: $[k] = \frac{1}{L^2}$.

Derivation of the Friedmann equations:
We have the stress energy tensor:

$$T_{\mu\nu} = \begin{pmatrix} T_{00} \cong \text{energy density} & T_{0j} \cong \text{energy flux} \\ T_{j0} \cong \text{momentum density} & T_{ik} \cong \text{stress tensor} \end{pmatrix} \tag{289}$$

The stress tensor T_{ik} is force per unit area:

$$\begin{pmatrix} T_{ii} \cong \text{pressure} & T_{ik} \cong \text{shear} \end{pmatrix} \tag{290}$$

$T_{\mu\nu}$ has to be a perfect fluid with no shear or isotropic pressure:

$$T_{\mu\nu} = (\rho c^2 + p)u_\mu u_\nu - \frac{p}{c^2}g_{\mu\nu} \tag{291}$$

In the rest frame of a comoving observer:

$$T_{\mu\nu} = \begin{pmatrix} -\rho c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \tag{292}$$

To evaluate $G_{\mu\nu}$, we take the derivative of the metric. We then plug this into the Einstein field equations, which gives two independent equations. This leads to the Friedmann equations:

$$\boxed{\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \end{aligned}}. \tag{293}$$

For relativistic bosons and fermions $p = \rho c^2/3$, and for non-relativistic particles $p = 0$.

Critical density and density parameters:
The *critical density* ρ_{crit} is the density that gives a flat universe ($k = 0$) and is given by

$$\rho_{\text{crit}} = \frac{3H^2(t)}{8\pi G} \tag{294}$$

with present-day value

$$\rho_{\mathrm{crit},0} = \frac{3H_0^2}{8\pi G} \approx 1.8 \times 10^{-29} h^2 \ \mathrm{g/cm^3} \ . \tag{295}$$

For a sphere with radius a filled with the critical density, the gravitational potential is equal to the specific kinetic energy:

$$\frac{G\frac{4}{3}\pi\rho_{\mathrm{crit}}a^3}{a} = \frac{\dot{a}^2}{2} \ . \tag{296}$$

This is the limiting case between an open and closed universe and leads to eternal expansion.

We define the *cosmological density parameters* in terms of the critical density:

$$\begin{aligned}
\Omega_m(t) &= \frac{\rho_m(t)}{\rho_{\mathrm{crit}}(t)} \\
\Omega_r(t) &= \frac{\rho_r(t)}{\rho_{\mathrm{crit}}(t)} \\
\Omega_k(t) &= -\frac{kc^2}{H^2} \\
\Omega_\Lambda(t) &= \frac{\Lambda c^2}{3H^2} = \frac{\rho_\Lambda(t)}{\rho_{\mathrm{crit}}(t)}, \ \rho_\Lambda(t) = \frac{\Lambda c^2}{8\pi G} \\
\Omega(t) &= \Omega_m(t) + \Omega_r(t)
\end{aligned} \tag{297}$$

The present-day values are:

$$\begin{aligned}
\Omega_{m,0} &= \frac{\rho_{m,0}}{\rho_{\mathrm{crit},0}} \\
\Omega_{r,0} &= \frac{\rho_{r,0}}{\rho_{\mathrm{crit},0}} \\
\Omega_{k,0} &= -\frac{kc^2}{H_0^2} \\
\Omega_{\Lambda,0} &= \frac{\rho_{\Lambda,0}}{\rho_{\mathrm{crit},0}} \\
\Omega_0 &= \frac{\rho_0}{\rho_{\mathrm{crit},0}}
\end{aligned} \tag{298}$$

so

$$\begin{aligned}
\rho_m &= \Omega_{m,0}\rho_{\mathrm{crit},0}a^{-3} \\
\rho_r &= \Omega_{r,0}\rho_{\mathrm{crit},0}a^{-4} \\
\rho_k &= \Omega_{k,0}\rho_{\mathrm{crit},0}a^{-2} \\
\rho_\Lambda &= \Omega_{\Lambda,0}\rho_{\mathrm{crit},0} \ .
\end{aligned} \tag{299}$$

We also often consider the baryon density parameter Ω_b separately from the total matter density, so the total matter density is the sum of the baryon and dark matter densities $\Omega_m = \Omega_{\mathrm{dm}} + \Omega_b$.

Our current measurements of these values are (from the Planck 2018 results)

$$\begin{aligned}\Omega_{m,0} &= 0.315 \pm 0.007 \\ \Omega_{\mathrm{dm},0} &= 0.264 \pm 0.003 \\ \Omega_{b,0} &= 0.0493 \pm 0.0003 \\ \Omega_{k,0} &= 0.0007 \pm 0.0019 \\ \Omega_{\Lambda,0} &= 0.6847 \pm 0.0073\end{aligned} \tag{300}$$

with $H_0 = 67.4 \pm 0.5\,\mathrm{km/s/Mpc}$. The radiation parameter $\Omega_{r,0}$ can be derived from the measured temperature of the CMB and relating the photon and neutrino density to get $\Omega_{r,0} \approx 10^{-4}$. We discuss how to obtain these values from observations in Part III.

We can rewrite first Friedmann equation:

$$\begin{aligned}H^2(t) &= \frac{8\pi G}{3}(\rho_m + \rho_r + \rho_\Lambda) - \frac{kc^2}{a^2} \\ &= \frac{8\pi G}{3}\rho_{\mathrm{crit},0}[\Omega_{m,0}a^{-3} + \Omega_{r,0}a^{-4} + \Omega_{\Lambda,0}] - \frac{kc^2}{a^2} \\ &(\rho_{\mathrm{crit},0} = \frac{3{H_0}^2}{8\pi G}) \\ &= {H_0}^2\left[\Omega_{m,0}a^{-3} + \Omega_{r,0}a^{-4} + \Omega_{\Lambda,0} - \frac{kc^2}{a^2{H_0}^2}\right] \\ &(-\frac{kc^2}{{H_0}^2} = \Omega_{k,0} = 1 - \Omega_{r,0} - \Omega_{m,0} - \Omega_{\Lambda,0}) \\ &= {H_0}^2\left[\Omega_{r,0}a^{-4} + \Omega_{m,0}a^{-3} + \Omega_{k,0}a^{-2} + \Omega_{\Lambda,0}\right]\end{aligned} \tag{301}$$

So we find:

$$\boxed{\begin{aligned}&H^2(a) = {H_0}^2E^2(a) \\ &E^2(a) = \Omega_{r,0}a^{-4} + \Omega_{m,0}a^{-3} + \Omega_{k,0}a^{-2} + \Omega_{\Lambda,0}\end{aligned}} \tag{302}$$

which is a useful form of the Friedmann equation.

Notes:

- Radiation dominates in early times, then matter, then the cosmological constant.
- Matter-Λ equality occurs when $\Omega_\Lambda = \Omega_m$:

$$\begin{aligned}\Omega_{\Lambda,0} &= \frac{\Omega_{m,0}}{a^3} \\ \Longrightarrow a &\approx \frac{1}{2.3}, z \approx 1.3\ (z \approx 1 \text{ is } 6-7\,\mathrm{Gyr} \text{ after the Big Bang})\ .\end{aligned} \tag{303}$$

- Matter-radiation equality occurs when $\Omega_r = \Omega_m$:

$$\Omega_{r,0}a^{-4} = \Omega_{m,0}a^{-3} \tag{304}$$

$$\Longrightarrow a \approx \frac{1}{3700},\ z = 3700\ . \tag{305}$$

- Observationally, $\Omega_{k,0} \approx 0$.

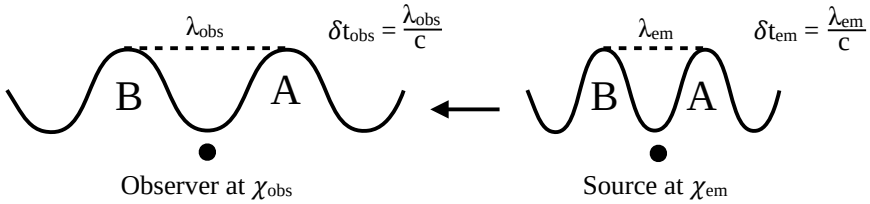
1.C Observational cosmology

Goal: relate the cosmological parameters to observations.

Redshift:
Redshift z is defined by the difference in observed wavelength and emitted wavelength of light:

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm em}} = 1 + \frac{\lambda_{\rm obs} - \lambda_{\rm em}}{\lambda_{\rm em}} \equiv 1 + z \tag{306}$$

In cosmology, this is due to the expansion of space. Light travels from the source at $(t_{\rm em}, a_{\rm em}, z_{\rm em})$ to the observer at $(t_{\rm obs}, a_{\rm obs}, z_{\rm obs})$.



The spatial hypersurface can shrink or expand depending on $a(t)$, so $\lambda_{\rm obs}$ is not necessarily equal to $\lambda_{\rm em}$. Photons always travel along the shortest path, so for light we have:

$$\mathrm{d}s = 0 \Rightarrow c\mathrm{d}t = a(t)\mathrm{d}\chi \tag{307}$$

$$\begin{aligned}
\text{Pulse A}: \int_{\chi_{\rm em}}^{\chi_{\rm obs}} \mathrm{d}\chi &= \int_{t_{\rm em}}^{t_{\rm obs}} \frac{c\mathrm{d}t}{a(t)} \\
\text{Pulse B}: \int_{\chi_{\rm em}}^{\chi_{\rm obs}} \mathrm{d}\chi &= \int_{t_{\rm em}+\delta t_{\rm obs}}^{t_{\rm obs}+\delta t_{\rm obs}} \frac{c\mathrm{d}t}{a(t)} = \int_{t_{\rm em}}^{t_{\rm obs}} ... + \int_{t_{\rm obs}}^{t_{\rm obs}\delta_{t_{\rm obs}}} ... - \int_{t_{\rm em}}^{t_{\rm em}+\delta_{t_{\rm em}}} \\
&\approx \int_{t_{\rm em}}^{t_{\rm obs}} \frac{c\mathrm{d}t}{a(t)} + \frac{c\delta t_{\rm obs}}{a(t_{\rm obs})} - \frac{c\delta t_{\rm em}}{a(t_{\rm em})}
\end{aligned} \tag{308}$$

then

$$\frac{c\delta t_{\rm obs}}{a(t_{\rm obs})} = \frac{c\delta t_{\rm em}}{a(t_{\rm em})} \Rightarrow \frac{\lambda_{\rm obs}}{\lambda_{\rm em}} = \frac{a_{\rm obs}}{a_{\rm em}} \tag{309}$$

so

$$\frac{a_{\rm obs}}{a_{\rm em}} = 1 + z\ . \tag{310}$$

For $a_0 = 1$ (observing today) we have:

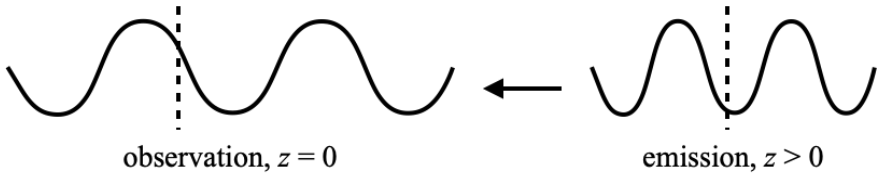
$$\begin{aligned}
\frac{1}{a} &= 1 + z \\
\boxed{a = \frac{1}{1+z}}\ .
\end{aligned} \tag{311}$$

Note: the change of luminosity $L = \frac{\text{energy of photons}}{\text{time}}$ is affected "twice" by expansion since

$$\begin{aligned} \frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} &= \frac{\delta t_{\text{obs}}}{\delta t_{\text{em}}} = \frac{a_{\text{obs}}}{a_{\text{em}}} \\ \Rightarrow L_{\text{obs}} &= \frac{h\nu_{\text{obs}}}{\delta t_{\text{obs}}} = \frac{h\nu_{\text{em}}}{\delta t_{\text{em}}} \left(\frac{a_{\text{em}}}{a_{\text{obs}}}\right)^2 = L_{\text{em}} \left(\frac{a_{\text{em}}}{a_{\text{obs}}}\right)^2 . \end{aligned} \tag{312}$$

If $a_{\text{obs}} = 1, z_{\text{obs}} = 0$ and $a_{\text{em}} = a, z_{\text{em}} = z$, we have

$$\begin{aligned} L_{\text{obs}} &= L_{\text{em}} \left(\frac{1}{1+z}\right)^2 \\ \Rightarrow \boxed{L_{\text{obs}} &= \frac{L_{\text{em}}}{(1+z)^2}} . \end{aligned} \tag{313}$$



Distance measures:
Question: what is the distance between a source at (z, t, a) and an observer?
In static Euclidean space, we can measure a unique distance in different ways. For a source with luminosity L and size l, we have:

- luminosity distance D_L: $F = \frac{L}{4\pi D_L^2}$
- angular diameter distance D_A: $\varphi = \frac{l}{D_A}$

Note that $D_L \neq D_A$ in expanding space!

Comoving distance:
Comoving coordinates move with space as it expands.

$$\begin{aligned} \chi(t_{\text{em}}, t_{\text{obs}}) &= \int_{t_{\text{em}}}^{t_{\text{obs}}} \frac{c\mathrm{d}t}{a(t)} \\ &\quad (\text{for light, } \mathrm{d}s = 0 \Rightarrow a\mathrm{d}\chi = c\mathrm{d}t) \\ &= \chi_{\text{em}}^{\text{obs}} = c\int_{a_{\text{em}}}^{a_{\text{obs}}} \frac{\mathrm{d}a}{\dot{a}a} = \frac{c}{H_0}\int_{a_{\text{em}}}^{a_{\text{obs}}} \frac{\mathrm{d}a}{a^2 E(a)} \end{aligned} \tag{314}$$

The comoving distance is not measurable through observations, but it is useful theoretically.

We also define a function that depends on the curvature k of space that is helpful in writing the metrics:

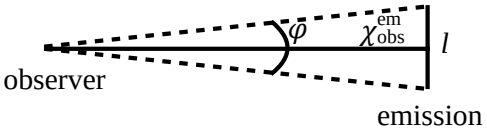
$$f_k(\chi) = \begin{cases} \frac{1}{\sqrt{k}} \sin\left(\chi\sqrt{k}\right), & k > 0 \\ \chi, & k = 0 \\ \frac{1}{\sqrt{|k|}} \sinh\left(\chi\sqrt{|k|}\right), & k < 0 \end{cases} \tag{315}$$

Angular diameter distance:
The *angular diameter distance* D_A is defined such that

$$\varphi = \frac{l}{D_A} \tag{316}$$

for an object that has angular size φ. Then the endpoints of l have the same (χ, θ, t):

$$\begin{aligned} l &= a_{\text{em}} f_k\left(\chi_{\text{em}}^{\text{obs}}\right) \varphi \\ &= \frac{1}{1+z} f_k\left(\chi_{\text{em}}^{\text{obs}}\right) \varphi \end{aligned} \tag{317}$$

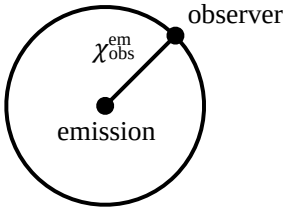


so

$$\boxed{D_A = \frac{1}{1+z} f_k\left(\chi_{\text{em}}^{\text{obs}}\right)} \,. \tag{318}$$

Luminosity distance:
The *luminosity distance* D_L is defined such that

$$F = \frac{L}{4\pi D_L^2} \,. \tag{319}$$



Then the observed surface, for $a_{\text{obs}} = 1$, is

$$4\pi a_{\text{obs}}^2 f_k^2\left(\chi_{\text{em}}^{\text{obs}}\right) = 4\pi f_k\left(\chi_{\text{em}}^{\text{obs}}\right) \,. \tag{320}$$

Furthermore, we can relate the observed and emitted luminosities

$$\begin{aligned} L_{\text{obs}} &= \frac{1}{(1+z)^2} L_{\text{em}} = \frac{1}{(1+z)^2} L \\ \Rightarrow F &= \frac{L}{4\pi f_k^2\left(\chi_{\text{obs}}^{em}\right)(1+z)^2} = \frac{L}{4\pi\left[f_k^2\left(\chi_{\text{obs}}^{em}\right)(1+z)\right]^2} \end{aligned} \tag{321}$$

so

$$\boxed{D_L = (1+z) f_k\left(\chi_{\rm em}^{\rm obs}\right)}. \tag{322}$$

This also gives us the relation

$$D_A = \frac{1}{(1+z)^2} D_L \tag{323}$$

so $D_A \approx D_L$ if $z \ll 1$.

Notes:

- The simplest Einstein-de Sitter case is:

$$\begin{aligned} \Omega_{\Lambda,0} = \Omega_{\kappa,0} = \Omega_{r,0} = 0,\ \Omega_{m,0} = 1 \\ \Rightarrow D_A = \frac{2c}{H_0}\frac{1}{1+z}\left(1 - \frac{1}{\sqrt{1+z}}\right) \\ D_L = \frac{2c}{H_0}(1+z)\left(1 - \frac{1}{\sqrt{1+z}}\right) \end{aligned} \tag{324}$$

  For $z \ll 1$:

$$D_A \approx D_L \approx \frac{2c}{H_0}\left(1 - \left(1 - \frac{1}{2}z\right)\right) \tag{325}$$

$$= \frac{c}{H_0} z \tag{326}$$

  Furthermore:

$$\chi_{\rm em}^{\rm obs} = \frac{c}{H_0}\int_{a_{\rm em}}^{a_{\rm obs}} \frac{{\rm d}a}{a^2 E(a)} = \frac{c}{H_0}\int_{a_{\rm em}}^{1} \frac{{\rm d}a}{a^2 E(a)} \tag{327}$$

$$\tag{328}$$

  For $E(a) \approx 1$ and $a \approx 1 - z$:

$$\chi_{\rm em}^{\rm obs} \approx \frac{c}{H_0}\frac{1}{(1-z)^2}\int_{a_{\rm em}}^{1} {\rm d}a \tag{329}$$

$$= \frac{c}{H_0}\frac{1}{(1-z)^2}(1-a) \tag{330}$$

$$\approx \frac{c}{H_0}\frac{1}{(1-z)^2} z \tag{331}$$

$$\approx \frac{c}{H_0} z \tag{332}$$

  So we get $D_A = D_L = \chi_{\rm em}^{\rm obs}$ for $z \ll 1$, i.e. agreement for low z.

- The general flat case is: $k = 0 \Rightarrow f_k(\chi) = \chi$.

$$\begin{aligned} D_A &= \frac{1}{1+z}\chi_{\rm em}^{\rm obs} \\ &= \frac{c}{H_0}\frac{1}{1+z}\int_{a_{\rm em}}^{a_{\rm obs}} \frac{{\rm d}a}{a^2 E(a)} \\ &= \frac{c}{H_0}\frac{1}{1+z}\int_{a_{\rm em}}^{1} \frac{{\rm d}a}{a^2 E(a)} \end{aligned} \tag{333}$$

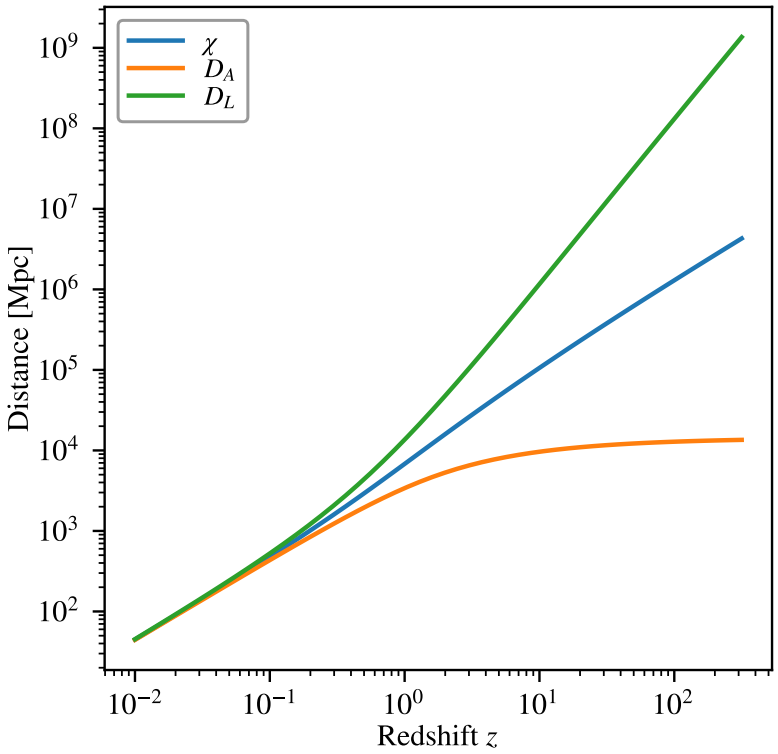
Since $a = \frac{1}{1+z}$ and $\mathrm{d}a = -\frac{1}{1+z}^2 \mathrm{d}z$, then

$$D_A = \frac{c}{H_0}\frac{1}{1+z}\int_0^z [\Omega_{m,0}(1+z)^3 + \Omega_{r,0}(1+z)^4 + \Omega_{\Lambda,0}]^{-\frac{1}{2}}\mathrm{d}z \tag{334}$$

So generally for $\Omega_{\kappa,0} = 0$ (i.e. flat):

$$D_A = \frac{c}{H_0}\frac{1}{1+z}\int_0^z [\Omega_{m,0}(1+z)^3 + \Omega_{r,0}(1+z)^4 + \Omega_{\Lambda,0}]^{-\frac{1}{2}}\mathrm{d}z \tag{335}$$

$$D_L = \frac{c}{H_0}(1+z)\int_0^z [\Omega_{m,0}(1+z)^3 + \Omega_{r,0}(1+z)^4 + \Omega_{\Lambda,0}]^{-\frac{1}{2}}\mathrm{d}z\,. \tag{336}$$



To the left is a plot comparing the angular diameter, luminosity, and comoving distances for a flat universe with $\Omega_m = 0.315$ and $H_0 = 67.4\,\mathrm{km/s/Mpc}$. By $z \sim 1$, there is a significant difference between the distance measures.

Comoving volume element:
We want to measure the number of objects in a given angle $\mathrm{d}\Omega$ on the sky in a redshift range $(z, z + \mathrm{d}z)$. What is the comoving number density of those objects, i.e. what is the corresponding comoving volume?

In general:

$$V = \mathrm{d}A\mathrm{d}r \tag{337}$$

where $\mathrm{d}A$ is area and $\mathrm{d}r$ is depth. So

$$\mathrm{d}V_\chi = \underbrace{\underbrace{(D_A{}^2\mathrm{d}\Omega)}_{\text{proper area}}(1+z)^2}_{\text{comoving area}} \cdot \underbrace{\mathrm{d}\chi}_{\text{comoving depth}} \tag{338}$$

$$A_{\text{proper}} = a^2 A_{\text{comoving}} \tag{339}$$

$$\Rightarrow A_{\text{comoving}} = \frac{1}{a^2} A_{\text{proper}} = (1+z)^2 A_{\text{proper}} \tag{340}$$

$\mathrm{d}\chi = \frac{c}{H_0}\frac{\mathrm{d}z}{E(z)}$, so

$$\mathrm{d}V_\chi = (D_A{}^2\mathrm{d}\Omega)(1+z)^2 \frac{c}{H_0}\frac{1}{E(z)}dz \tag{341}$$

and finally:

$$\mathrm{d}V_\chi = \frac{c}{H_0}\frac{(1+z)^2 D_A{}^2}{E(z)}\mathrm{d}\Omega\mathrm{d}z\ . \tag{342}$$

Plug in $D_A = \frac{1}{1+z} f_\kappa(\chi)$:

$$\mathrm{d}V_\chi = \frac{c}{H_0}\frac{f_\kappa^2(\chi)}{E(z)}\mathrm{d}z\mathrm{d}\Omega \tag{343}$$

$$= f_k^2(\chi) r d\Omega \frac{\mathrm{d}\chi}{\mathrm{d}z}\mathrm{d}z\ . \tag{344}$$

1.D Inflation

So far, dynamics have been described by the Friedmann equations with some mass-energy content of the Universe: $\Omega_m, \Omega_r, \Omega_k, \Omega_\Lambda$. Is this sufficient to explain all data?

Problems:

- Horizon problem:

  $$\rho_r \propto (1+z)^4 \text{ and } \rho_r \propto T^4 \Rightarrow T \propto (1+z) \tag{345}$$

  The Universe cools and at some z_{recomb}, it consists of neutral hydrogen atoms (recombination). We get the balancing equation

  $$H^+ + e \rightleftharpoons H^0 + \chi \tag{346}$$

  where $\chi = 13.6\,\text{eV}$ is the ionization energy. We also have:

  $$\begin{aligned} x &= \frac{\text{number density of free e}^-}{\text{number density of protons}} \\ \eta &= \frac{n_b}{n_\gamma} = \frac{\text{baryon number density}}{\text{photon number density}} \approx 5\times 10^{-10}\left(\frac{\Omega_{b,0}h^2}{0.01}\right) \end{aligned} \tag{347}$$

which we use in the Saha equation:

$$\frac{1-x}{x^2} \approx 3.84\eta \left(\frac{k_B T}{m_e c^2}\right)^{3/2} e^{-\frac{\chi}{k_B T}} \tag{348}$$

With $\chi = 13.6\,\text{eV}$ corresponding to $\sim 10^5\,\text{K}$ ($1\,\text{eV} \approx 10^4\,\text{K}$), we would expect $x < 1$ for $T < 10^5\,\text{K}$. However, there are many more photons than baryons, which leads to $x < 1$ only for $T \approx 3000\,\text{K}$. This gives $z_{\text{recomb}} \approx 1090$ (for $\Omega_{b,0} = 0.045, T_0 = 2.73\,\text{K}$).

After this time, photons can escape or free stream, and we can observe them as the Cosmic Microwave Background. The CMB is very uniform: $\frac{\Delta T}{T} \leq 10^{-5}$ (note that CMB maps are typically logarithmic).

Why is this a problem?
Horizons are the largest causally connected regions by light rays.

The comoving horizon size is:

$$\mathrm{d}s = 0 \text{ (light)} \implies c\mathrm{d}t = a(t)\mathrm{d}\chi \implies \chi_{\text{horizon}} = \int_0^t \frac{c\mathrm{d}t}{a(t)} \tag{349}$$

So we have

$$\chi_{\text{horizon}}(z) = \int_0^{a=(1+z)^{-1}} \frac{c\mathrm{d}a}{a^2 H(a)}\,. \tag{350}$$

For a flat radiation dominated universe:

$$l_{\text{horizon}} = a\chi_{\text{horizon}} = \frac{c}{H(z)} = \frac{c}{H_0\sqrt{\Omega_{r,0}}}\frac{1}{1+z} \tag{351}$$

flat matter dominated:

$$\begin{aligned} l_{\text{horizon}} = & \underbrace{a\Big(\int_0^{(1+z_{eg})^{-1}} \frac{c\mathrm{d}a}{a^2 H(a)} + \int_{(1+z_{eq})^{-1}}^{(1+z)^{-1}} \frac{c\mathrm{d}a}{a^2 H(a)}\Big)}_{\text{largest contribution comes from matter dominated phase}} \\ & \approx \frac{2c}{H(z)} = \frac{2c}{H_0\sqrt{\Omega_{m,0}}}\frac{a}{\sqrt{1+z}} = \frac{2c}{H_0\sqrt{\Omega_{m,0}}}\frac{1}{(1+z)^{\frac{3}{2}}} \end{aligned} \tag{352}$$

Apply this to z_{recomb}:

$$\begin{aligned} & \frac{\rho_{r,0}(1+z_{\text{recomb}})^4}{\rho_{m,0}(1+z_{\text{recomb}})^3} \sim 5\times 10^{-2} \Rightarrow \text{matter dominated regime} \\ \Rightarrow & l_{\text{horizon}} = \frac{2c}{H_0}\Omega_{m,0}^{-\frac{1}{2}}(1+z)^{-\frac{3}{2}} \end{aligned} \tag{353}$$

Angular size of horizon:

$$\varphi_{\text{horizon}} = \frac{l_{\text{horizon}}}{D_A} \qquad D_A : \text{Angular diameter distance} \tag{354}$$

$$
\begin{aligned}
D_A &= \frac{1}{1+z} f_\kappa(\chi_{\text{em}}^{\text{obs}}) = \frac{1}{1+z}\chi_{\text{em}}^{\text{obs}} \qquad \text{(for flat universe)} \\
&= \frac{c}{1+z}\frac{1}{H_0}\int_0^z \left[\Omega_{m,0}(1+z)^3\right]^{-\frac{1}{2}} \mathrm{d}z \\
&= \left(\frac{2c}{H_0(1+z)\Omega_m 0^{\frac{1}{2}}}\left[-\frac{1}{\sqrt{1+z}}\right]\right)_0^z \\
&= \frac{2c}{H_0 \underbrace{(1+z)}_{\approx z, z\gg 1}\Omega_{m0}^{\frac{1}{2}}} \underbrace{\left[1-\frac{1}{\sqrt{1+z}}\right]}_{\approx 1, z\gg 1} \\
&\approx \frac{2c}{H_0}\frac{1}{\Omega_{m0}^{\frac{1}{2}} z}
\end{aligned}
\tag{355}
$$

This gives us the angular size of the horizon at recombination:

$$
\varphi_{\text{horizon,recomb}} \approx \sqrt{\frac{1}{z_{\text{recomb}}}} \sim 1.7^\circ \tag{356}
$$

Or more generally:

$$
\varphi_{\text{horizon,recomb}} \approx 1.7^\circ \sqrt{\Omega_{m,0}}\,. \tag{357}
$$

This is much smaller than the full sky, so how can the CMB be so uniform?

- Flatness problem:
  At high z, Λ is irrelevant in the Friedmann equations, so:

$$
H^2(a) = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} = H^2(a)\left[\Omega(a) - \frac{kc^2}{a^2H^2(a)}\right], \qquad \rho = \rho_m + \rho_r \tag{358}
$$

  Thus, deviation from flatness $\Omega(a) = 1$ is:

$$
|\Omega(a) - 1| = \frac{kc^2}{a^2H^2(a)}\,. \tag{359}
$$

  Since $a \propto t^{2/3}$ in matter dominated times and $a \propto t^{1/2}$ during radiation dominated times, we have:

$$
|\Omega(t) - 1| \propto \begin{cases} t, & \text{radiation dominated} \\ t^{2/3}, & \text{matter dominated} \end{cases} \tag{360}
$$

  Thus, any small deviation $\Omega(t_{\text{early}}) \neq 1$ at early times quickly blows up! $\Omega(t_{\text{early}})$ must therefore be very close to 1, which leads to a "fine-tuning problem."

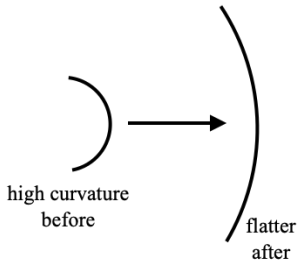
- Monopole problem:
  General unified theories predict many magnetic monopoles, but this is not observed. The number density must decrease.

- Seeds of structure formation problem:
  What seeds the perturbations that become the large structures we observe?

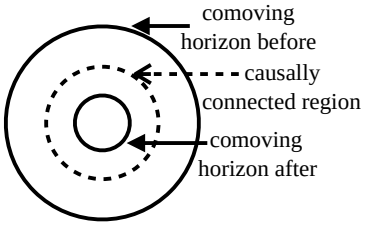
Inflation: basic ideas:

- Flatness problem:

  If $\frac{kc^2}{a^2H^2(a)}$ decreased with time for a short period, then $\Omega(a)$ would be driven towards $\Omega(a) = 1$.



- Horizon problem:
  If $\frac{kc^2}{a^2H^2(a)}$ shrinks, then $\chi \propto \frac{c}{aH(a)}$ also shrinks.



$\Rightarrow$ can explain smoothness within the observable universe.

So decreasing $\frac{1}{aH(a)}$ seems to solve two problems! The conditions for a shrinking comoving horizon:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{c}{aH}\right) &< 0 \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{c}{\dot{a}}\right) &< 0 \\ -\frac{c\ddot{a}}{\dot{a}^2} &< 0 \\ \Rightarrow \ddot{a} &> 0 \end{aligned} \tag{361}$$

We need some period of accelerated expansion. We can look at the second Friedmann equation (e.g. for acceleration):

$$\begin{aligned} \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}(\rho + \frac{3p}{c^2}) + \frac{\Lambda c^2}{3} \\ &\text{(at early times, } \Lambda = 0) \\ &= -\frac{4\pi G}{3}(\rho + \frac{3p}{c^2}) \\ &\Rightarrow p < -\frac{\rho c^2}{3} \leftarrow \text{we need sufficiently negative pressure} \\ &\Rightarrow \frac{p}{\rho c^2} < -\frac{1}{3} \end{aligned} \tag{362}$$

This also solves the monopole and seed problem! Rapid expansion would decrease the density of monopoles and blow up tiny perturbations. All problems are then solved.

Note that $\frac{\Lambda c^2}{3}$ actually corresponds to a negative pressure term. To see this more clearly, we combine both Friedmann equations to derive the energy conservation equation:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}(\rho c^2 a^3) + p\frac{\mathrm{d}}{\mathrm{d}t}(a^3) = 0 \\ &\Rightarrow \dot{\rho} = -3H(a)(\rho + \frac{p}{c^2}) \end{aligned} \tag{363}$$

And for Λ with $\rho_\Lambda = \text{constant } (= \rho)$:

$$\rho + \frac{p}{c^2} = 0 \ \Rightarrow p = -\rho c^2 \tag{364}$$

So the equation of state parameter is

$$w = \frac{p}{\rho c^2} = -1 < -\frac{1}{3} \tag{365}$$

where $w = -1/3$ is needed for accelerated expansion as shown above. Thus, Λ leads to accelerated expansion and therefore a shrinking comoving horizon. Once Λ dominates in the Friedmann equation:

$$\begin{aligned} &H^2(a) = H_0^2\Omega_\Lambda = \left(\frac{\dot{a}}{a}\right)^2 \\ \Rightarrow &a \propto e^{\sqrt{\Omega_\Lambda}H_0 t} \end{aligned} \tag{366}$$

which is exponential growth.

Inflation:
Λ has all the features we want, but it:

- acts too late
- is constant, i.e. even if it acted early enough, it would not stop inflation!

How do we get all this in the early universe? We look at a homogeneous scalar field (inflation):

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi) \tag{367}$$

which leads to the energy-momentum tensor:

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\mathcal{L} \ \Rightarrow \begin{aligned} T_{00} &= \rho c^2 = \frac{1}{2}\dot{\phi}^2 + V(\phi) \\ T_{ii} &= p = \frac{1}{2}\dot{\phi}^2 - V(\phi) \end{aligned} \tag{368}$$

To get $w < -1/3$, we require:

$$\begin{aligned} \frac{1}{2}\dot{\phi} - V(\phi) &< -\frac{1}{3}\left(\frac{1}{2}\dot{\phi} + V(\phi)\right) \\ p &< -\frac{\rho c^2}{3} \\ \Rightarrow \dot{\phi}^2 &< V(\phi) \end{aligned} \tag{369}$$

i.e. the field must be moving slowly during inflation. Thus, the potential should be flat and have a minimum to stop inflation. Furthermore:

$$\begin{aligned}
&\text{Friedmann equation}: H^2 = \frac{8\pi G}{3}\left[\frac{1}{2}\dot{\phi}^2 + V(\phi)\right] \\
&\text{Energy conservation}: \dot{\rho} = -3H(a)\left(\rho + \frac{p}{c^2}\right) \\
&\qquad \dot{\rho}c^2 = \dot{\phi}\ddot{\phi} + \dot{\phi}\frac{\mathrm{d}V}{\mathrm{d}\phi} \\
&\qquad \text{with } \rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \text{ and } \frac{p}{c^2} = \frac{1}{2}\dot{\phi} - V(\phi) \\
&\qquad \Rightarrow \dot{\phi}\ddot{\phi} + \dot{\phi}\frac{dV}{d\phi} = -3H(a)\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right) - 3H(a)\left(\frac{1}{2}\dot{\phi}^2 - V(\phi)\right) \\
&\qquad \Rightarrow \ddot{\phi} + \frac{\mathrm{d}V}{\mathrm{d}\phi} = -3H(a)\dot{\phi} \\
&\qquad \Rightarrow \boxed{\ddot{\phi} + \underbrace{3H(a)}_{\text{Hubble drag}}\dot{\phi} = -\frac{\mathrm{d}V}{\mathrm{d}\phi}}
\end{aligned} \tag{370}$$

and we get the *field evolution equation*. In a static universe, $H = 0$, and there is no Hubble drag. $\frac{\mathrm{d}V}{\mathrm{d}\phi}$ is how fast energy is extracted from inflation.

Slow roll conditions:
We approximate

$$H^2 \approx \frac{8\pi G}{3}V(\phi) \tag{371}$$

which is $\approx V_0$ and roughly constant during the slow roll, leading to exponential growth. We also have

$$3H\dot{\phi} \approx -\frac{\mathrm{d}V}{\mathrm{d}\phi} \text{ (with } \ddot{\phi} \approx 0) \tag{372}$$

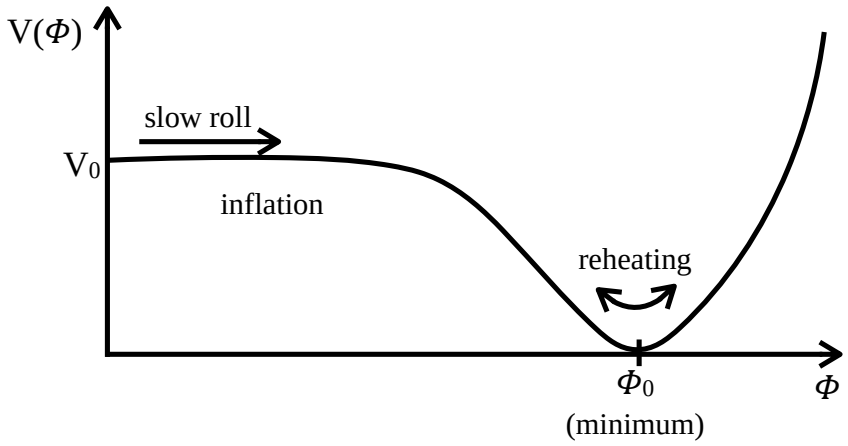
which is equivalent to:

$$\begin{gathered}
\dot{\phi}^2 \ll V \\
\text{and} \\
\frac{\mathrm{d}}{\mathrm{d}t}\dot{\phi}^2 \ll \frac{\mathrm{d}V}{\mathrm{d}t} \Rightarrow \ddot{\phi} \ll \frac{\mathrm{d}V}{\mathrm{d}\phi}
\end{gathered} \tag{373}$$

This can be rewritten in slow roll parameters:

$$\begin{aligned}
\epsilon &:= \frac{1}{24\pi G}\left(\frac{V'}{V}\right)^2 \ll 1 \\
\eta &:= \frac{1}{8\pi G}\left(\frac{V''}{V}\right) \ll 1
\end{aligned} \tag{374}$$

As long as these conditions are valid, inflation will go on. The slow roll potential is:



During reheating, the inflation field decays through coupling to ordinary matter ("reheat universe").

Since

$$H^2 = \frac{8\pi G}{3} V(\phi) \approx \frac{8\pi G}{3} V_0 \tag{375}$$

during inflation, large values of ϕ_0 and V_0 lead to more inflation (longer slow roll).

1.E Basic story of cosmology

Main ingredients:

- metric (geometry)
- Friedmann equations (dynamics)
- distances (connection to observations)
- horizons (evidence for inflation)

Emerging story

a) $t = 0$: Big Bang

b) $t \sim 10^{-34}$ s: inflation

c) T decreases as $T \propto (1 + z)$

d) $z \approx 3200$: transition from radiation to matter domination

e) $z \approx 1100$: recombination

f) Structure formation is nonlinear. First stars and galaxies...

g) $z \approx 0.33$: transition from matter to Λ domination

The first five stages here are optically thick to photons, while later is optically thin and potentially observable.

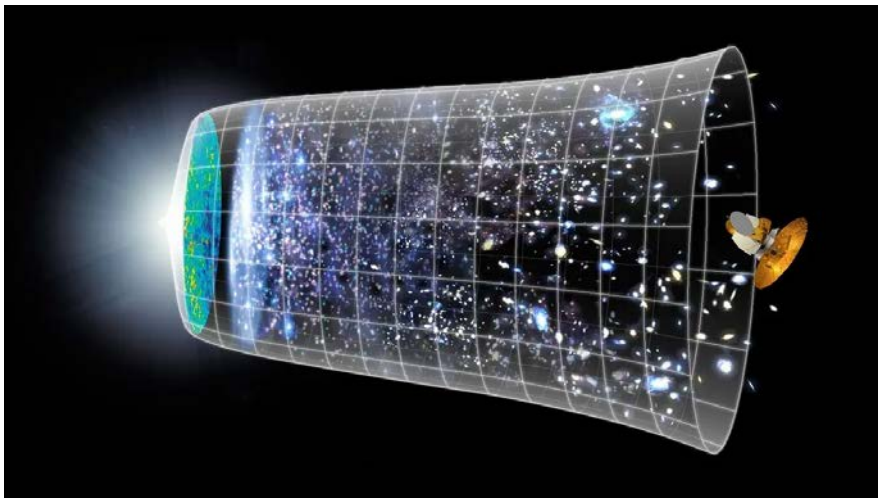


Image: NASA / WMAP Science Team. Image is in the public domain.

2 Structure formation

So far, we have assumed a uniform cosmology. We now add perturbations to study the growth of structure.

2.A Linear perturbation theory

There are small perturbations at early times. The Universe consists of matter (dark matter and baryons) and radiation. Λ and curvature are unimportant early on.

Basic equations:

- non-relativistic matter (dark matter, baryons) is important in the matter-dominated regime:

$$\begin{aligned} \text{continuity equation}: \frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0 \\ \text{momentum equation}: \frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla})\vec{v} = -\frac{\vec{\nabla} p}{\rho} + \vec{\nabla}\phi \\ \text{Poisson's equation}: \vec{\nabla}^2 \phi = 4\pi G \rho \end{aligned} \tag{376}$$

- relativistic matter (radiation)

$$\begin{aligned} \text{continuity equation}: \frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \left((\rho + \frac{p}{c^2})\vec{v} \right) = 0 \\ \text{momentum equation}: \frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla})\vec{v} = -\frac{\vec{\nabla} p}{\rho + \frac{p}{c^2}} + \vec{\nabla}\phi \\ \text{Poisson's equation}: \vec{\nabla}^2 \phi = 4\pi G (\rho + \frac{3p}{c^2}) \end{aligned} \tag{377}$$

Notes:

- non-relativistic
  - Dark matter follows the collisionless Boltzmann equation; $1^{\text{st}}/2^{\text{nd}}$ equations only hold for moments (Jeans equation).
  - For dark matter, there is no well-defined velocity field $\vec{v}(\vec{x},t)$ due to multistream. $\vec{v}(\vec{x},t)$ is just an average.
  - Nevertheless, it recovers the correct growth rate for large scales $> \lambda_J$ when pressure can be neglected.
- relativistic
  - gravitational source terms include pressure terms.
  - For pure radiation: $p = \frac{\rho c^2}{3}$.

This leads to the perturbation equation where some small perturbation δ evolves in a smooth background density $\bar{\rho}$:

$$\delta = \frac{\Delta\rho}{\bar{\rho}} = \frac{\rho - \bar{\rho}}{\bar{\rho}} \,. \tag{378}$$

Perturbation equations:

$$\begin{aligned} \text{non-relativistic: } \ddot{\delta} + 2H\dot{\delta} &= \left(4\pi G\bar{\rho}\delta + \frac{v_s^2\vec{\nabla}^2\delta}{a^2}\right) \\ \text{relativistic: } \ddot{\delta} + 2H\dot{\delta} &= \left(\frac{32}{3}\pi G\bar{\rho}\delta + \frac{v_s^2\vec{\nabla}^2\delta}{a^2}\right) \end{aligned} \tag{379}$$

where

$$\begin{aligned} \delta &= \delta(\vec{x},t) \\ \vec{x} &: \text{comoving coordinates} \\ \vec{r} &: a\vec{x} \text{ physical coordinates} \\ \vec{\nabla} &= \frac{\partial}{\partial\vec{x}} \\ v_s &= \begin{cases} c_s, & \text{non-relativistic baryons} \\ \sigma, & \text{non-relativistic dark matter} \\ \dfrac{c}{\sqrt{3}}, & \text{relativistic radiation} \end{cases} \end{aligned} \tag{380}$$

Fourier representation:

For comoving coordinate $\vec{x}$ and comoving wave number $\vec{k}$:

$$\begin{aligned}\delta(\vec{x},t) &= \int \frac{d^3k}{(2\pi)^3}\hat{\delta}(\vec{k},\delta)e^{-i\vec{k}\cdot\vec{x}} \\ \hat{\delta}(\vec{k},t) &= \int d^3x\delta(\vec{x},t)e^{+i\vec{k}\cdot\vec{x}}\end{aligned} \tag{381}$$

and we get

$$\begin{aligned}\text{non-relativistic: } \ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} &= \hat{\delta}\left(4\pi G\bar{\rho} - \frac{v_s^2k^2}{a^2}\right) \\ \text{relativistic: } \ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} &= \hat{\delta}\left(\frac{32}{3}\pi G\bar{\rho} - \frac{v_s^2k^2}{a^2}\right) .\end{aligned} \tag{382}$$

Growing modes:
For a static background, $H = 0$ and:

$$\ddot{\hat{\delta}} + w_0^2\hat{\delta} = 0 \tag{383}$$

with

$$w_0^2 = \begin{cases} \dfrac{v_s^2k^2}{a^2} - 4\pi G\bar{\rho}, & \text{non-relativistic} \\ \dfrac{v_s^2k^2}{a^2} - \dfrac{32}{3}\pi G\bar{\rho}, & \text{relativistic} \end{cases} \tag{384}$$

For physical wave number $\tilde{k} = k/a$, we get oscillation for:

$$\tilde{k} \geq \tilde{k}_J = \begin{cases} \dfrac{2\sqrt{\pi G\bar{\rho}}}{v_s}, & \text{non-relativistic} \\ \dfrac{\sqrt{\frac{32}{3}}\sqrt{\pi G\bar{\rho}}}{v_s}, & \text{relativistic} \end{cases} \tag{385}$$

and growth (no oscillations) for modes with lengths $l = \frac{2\pi}{\tilde{k}}$ greater than λ_J:

$$l \geq \lambda_J = \frac{2\pi}{\tilde{k}_J} \propto \frac{v_s}{\sqrt{\pi G\bar{\rho}}} \tag{386}$$

where $\tilde{k} = k/a$ is in physical units. We can make this more general for $H \neq 0$ and neglecting the pressure terms for $l \geq \lambda_J$ and we get:

$$\begin{aligned}\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} &= 4\pi G\bar{\rho}\hat{\delta}, \qquad \text{non-relativistic} \\ \ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} &= \frac{32}{3}\pi G\bar{\rho}\hat{\delta}, \qquad \text{relativistic}\end{aligned} \tag{387}$$

Now for $\Omega = 1$, we have the critical density as the background density for the radiation and matter dominated regime:

$$\bar{\rho} = \rho_{\text{crit}} = \frac{3H^2}{8\pi G} \tag{388}$$

such that

$$\begin{aligned} \ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} &= \frac{3}{2}H^2\hat{\delta}, \text{ matter dominated } \Omega = 1 \\ \ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} &= 4H^2\hat{\delta}, \text{ radiation dominated } \Omega = 1 \end{aligned} \tag{389}$$

Now:

$$H = \frac{\dot{a}}{a} = \begin{cases} \dfrac{2}{3t}, & \text{matter dominated} \\ \dfrac{1}{2t}, & \text{radiation dominated} \end{cases} \tag{390}$$

We now assume $\hat{\delta}(\vec{k}, t) \propto t^n$. Then:

$$\begin{aligned} n^2 + \frac{n}{3} - \frac{2}{3} &= 0, \qquad \text{matter dominated} \\ n^2 - 1 &= 0, \qquad \text{radiation dominated} \end{aligned} \tag{391}$$

$$\begin{aligned} \Rightarrow n &= -1, \frac{2}{3}, \qquad \text{matter dominated} \\ n &= -1, +1, \quad \text{radiation dominated} \end{aligned} \tag{392}$$

that correspond to negative decaying modes, which are unimportant since the perturbations vanish, and positive growing modes. This gives us:

$$\boxed{\hat{\delta} \propto \begin{cases} a, & \text{matter dominated} \\ a^2, & \text{radiation dominated} \end{cases}} \tag{393}$$

In general, we write $\delta = D\delta_0$ or $\hat{\delta} = D\hat{\delta}_0$ where D is the growth factor such that $D(z = 0) = 1$.

Growth of baryons vs. cold dark matter:
Growth of perturbations occurs for $\lambda > \lambda_J$ or $M > M_J = \frac{4}{3}\pi\bar{\rho}\lambda_J^3$.
Baryons:

- Until recombination, there is strong coupling between photons and electrons.
- Before recombination

$$\begin{aligned} c_s^2 = \frac{\partial p}{\partial \rho}, \; p = \frac{1}{3}\rho c^2, \; \text{(radiation)} \\ \Rightarrow c_s = \frac{c}{\sqrt{3}}\,. \end{aligned} \tag{394}$$

- After recombination

$$\begin{aligned} c_s^2 = \frac{\partial p}{\partial \rho}, \; p = \frac{\rho}{mkT}, \; T = 2.71 \text{ K}(1 + \text{z}), \; \text{(ideal gas)} \\ \Rightarrow c_s = \sqrt{\frac{kT}{m}} \approx 5 \text{ km/s} \end{aligned} \tag{395}$$

- So
$$\begin{aligned} z \gg 1100: \;\; M_J &\geq 10^{16}\,\mathrm{M}_\odot \\ z < 1100: \;\; M_J &\lesssim 10^{5}\,\mathrm{M}_\odot \end{aligned} \tag{396}$$
since after recombination, the photon pressure support is removed.
- Structure can only form after $z \sim 1000$.
- if structure can only grow from $z \sim 1000$, δ will be amplified by $\sim 10^3$ (since matter dominated growth is $\propto a$). BUT: the CMB has $\delta \sim 10^{-5}$ and $10^{-5} \times 10^3 \sim 10^{-2}$ today, which is much less than what we observe in the low redshift universe. This theory of structure growth is not sufficient.
- Solution: dark matter must have clumped before and baryons fall into dark matter wells.

Cold dark matter:

- CDM is very cold, so it has a tiny velocity dispersion σ. This means that M_J is tiny and collapse on all scales is possible.
- CDM does not interact with radiation, so it can grow before recombination.

Cold dark matter is needed to make structure formation work!

2.B Growth of linear perturbations

Full general relativity treatment can be used to study growth beyond the horizon scale. Modes outside the horizon can always grow (no causal contact):

$$\delta \propto \begin{cases} a \propto t^{2/3}, & \text{matter dominated} \\ a^2 \propto t, & \text{radiation dominated} \end{cases} \tag{397}$$

Once a mode enters the horizon, its growth changes (note that we have perturbations on different length scales).

Baryons:

Baryons have a finite Jeans length before recombination:

$$\lambda_J = \frac{c}{\sqrt{G\bar{\rho}}}\sqrt{\frac{\pi}{3}}, \;\; (c_s = \frac{c}{\sqrt{3}}) \tag{398}$$

so modes with $l < \lambda_J$ have no growth. However, this is only if l is also within the horizon.

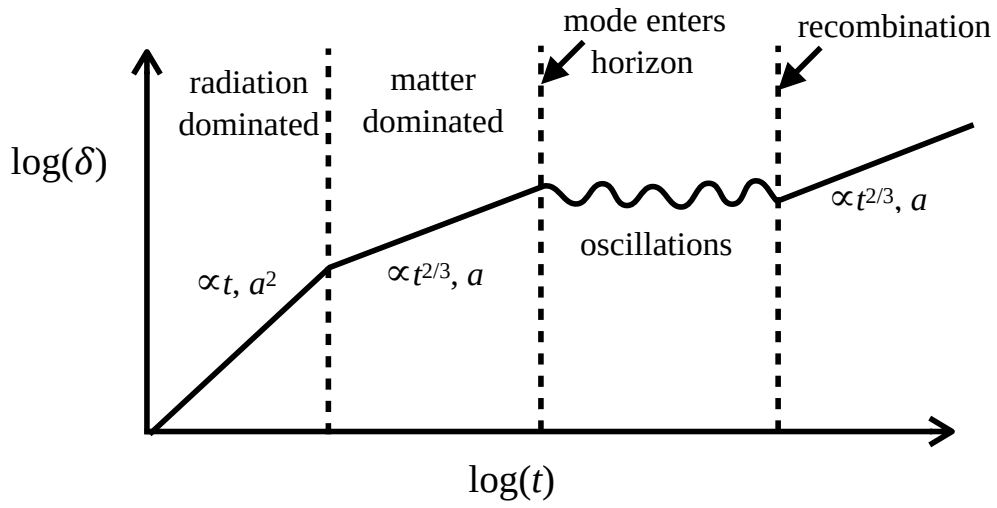
The growth of the physical horizon is (for $\Omega = 1$):

$$\begin{aligned} l_{\text{horizon}} = a\int_0^t \frac{c\mathrm{d}t}{a(t)} &= \begin{cases} 2ct, & \text{radiation dominated; } a \propto t^{1/2} \\ 3ct, & \text{matter dominated; } a \propto t^{2/3} \end{cases} \\ &= \begin{cases} \dfrac{c}{\sqrt{G\bar{\rho}}}\sqrt{\dfrac{3}{8\pi}}, & \text{radiation dominated} \\ \dfrac{c}{\sqrt{G\bar{\rho}}}\sqrt{\dfrac{3}{2\pi}}, & \text{matter dominated} \end{cases} \end{aligned} \tag{399}$$

where we used

$$H^2 = \frac{8\pi G\rho}{3} \text{ (for } \rho = \rho_{\rm crit} = \frac{3H^2}{8\pi G}) = \begin{cases} \frac{1}{4t^2}, & \text{radiation} \\ \frac{4}{at^2}, & \text{matter} \end{cases} \quad (400)$$

$l_{\rm horizon} < \lambda_J$: as soon as mode length l enters the horizon, it will oscillate! So before recombination, perturbations can grow if $l > l_{\rm horizon}$, otherwise they will oscillate.



The horizon and Jeans mass grow as we have seen before: $M_J \sim 10^{16}\mathrm{M}_\odot$ at $z \sim 1000$, so all modes smaller than $10^{16}\mathrm{M}_\odot$ entered the horizon before recombination and therefore start to oscillate and stop growing.

There is also another problem for those modes: *Silk damping*! Before decoupling, photons do not free stream because of Thomson scattering off free electrons. The mean free path gets large towards recombination. So:

- $M < M_J \sim 10^{16}\mathrm{M}_\odot$ perturbations oscillate due to photon pressure.
- Photons can diffuse out of potential wells and take baryons with them (electrons through Thomson scattering and protons through Coulomb interactions), which erases perturbations.

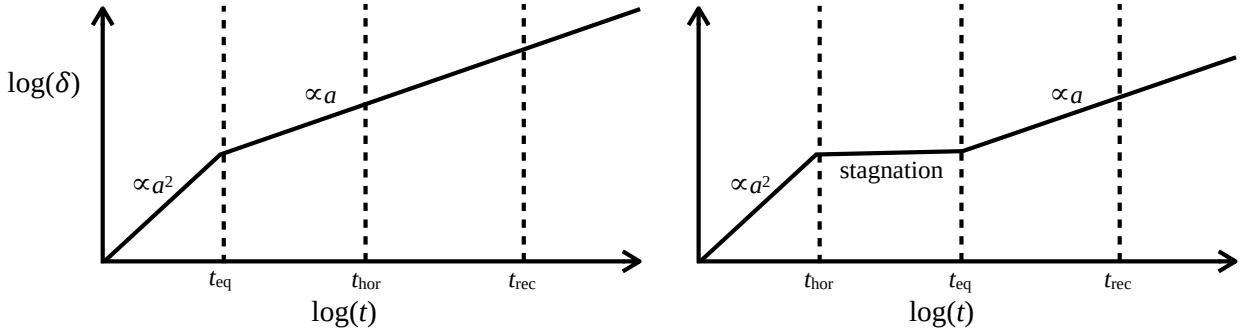
The net effect is that all perturbations $\sim 10^{12}\mathrm{M}_\odot$ (Silk mass) are damped and erased!

Cold dark matter:
Cold dark matter has essentially zero Jeans mass, so all modes can already grow. However, for subhorizon modes in the radiation dominated epoch, $\delta \sim$constant (stagnation). Because

the expansion rate is higher than the growth rate, we get:

$$\begin{aligned} \text{expansion timescale: } \tau_{\text{Hubble}} &\approx \frac{1}{\sqrt{G\rho_r}} \\ \text{collapse timescale: } \tau_{\text{Jeans}} &\approx \frac{1}{\sqrt{G\rho_m}} \end{aligned} \tag{401}$$

and $\tau_{\text{Hubble}} \ll \tau_{\text{Jeans}}$ if $\rho_r \gg \rho_m$. So modes entering the horizon during the radiation dominated phase are frozen (but not damped through something like Silk damping). After recombination, baryons can fall into CDM wells and grow.



A mode that enters the horizon at t_{hor} after matter-radiation equality at t_{eq} will always grow. Modes that enter the horizon during the radiation dominated regime will stagnate until matter domination.

Cold dark matter is then the main driver of structure formation since it there is time for CDM perturbations to grow large enough. Without CDM, structure formation is not possible.

2.C Statistical measures of structure

We see structure on different scales. We can use the power spectrum $P(k)$ to describe this. Reminder:

$$\begin{aligned} \delta(\vec{x}) &= \int \frac{d^3k}{(2\pi)^3} \hat{\delta}(\vec{k}) e^{-i\vec{k}\cdot\vec{x}} \\ \hat{\delta}(\vec{k}) &= \int d^3x \delta(\vec{x}) e^{+i\vec{k}\cdot\vec{x}} \end{aligned} \tag{402}$$

Variance and the power spectrum:
average:

$$\langle \delta \rangle = \int d^3x \delta(\vec{x}) = 0 \tag{403}$$

variance:

$$\begin{aligned} \sigma^2 &= \left\langle \delta^2 \right\rangle - \langle \delta \rangle^2 = \left\langle \delta^2 \right\rangle > 0 \\ \left\langle \delta^2 \right\rangle &= \int \mathrm{d}^3x \delta^2(\vec{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^3} |\hat{\delta}(\vec{k})|^2 \end{aligned} \tag{404}$$

If we assume homogeneity and isotropy, $\vec{k} \to k = |\vec{k}|$ and $\mathrm{d}^3\vec{k} = 4\pi k^2 \mathrm{d}k$. Then we get:

$$\begin{aligned} \sigma^2 &= \frac{1}{2\pi^2}\int |\hat{\delta}(k)|^2 k^2 \mathrm{d}k \\ &=: \frac{1}{2\pi^2}\int P(k) k^2 \mathrm{d}k \\ \text{with } P(k) &= |\hat{\delta}(k)|^2 \end{aligned} \tag{405}$$

Notes:

- $P(k)$ and σ are functions of time since $\hat{\delta}(k)$ grows ($\sigma = D\sigma_0$).
- The initial power spectrum is the primordial power spectrum set at the end of inflation. The general form is
$$P(k) = Ak^n \tag{406}$$
which is a power law and is scale-free. According to predictions from inflation, $n \approx 1$.

Measuring $P(k)$ and galaxy clustering:
If we assume galaxies trace the mass perturbations, what is the probability dP that we find two galaxies in volumes dV_1 and dV_2 at a distance r from each other?

$$\begin{aligned} \mathrm{d}P &= n_0(1+\delta(\vec{x}))\mathrm{d}V_1 \cdot n_0(1+\delta(\vec{x}+\vec{r}))\mathrm{d}V_2 \\ &= n_0^2(1+\underbrace{\delta(\vec{x})}_{=0}+\underbrace{\delta(\vec{x}+\vec{r})}_{=0}+\delta(\vec{x})\delta(\vec{x}+\vec{r}))\mathrm{d}V_1\mathrm{d}V_2 \\ &= n_0^2(1+\xi(r))\mathrm{d}V_1\mathrm{d}V_2 \end{aligned} \tag{407}$$

where $\delta(\vec{x})$ and $\delta(\vec{x}+\vec{r})$ are zero on average and $\vec{r} \to r$ due to isotropy. ξ is the *two-point correlation function* and is related to $P(k)$:

$$\begin{aligned} \xi(r) &= \int \mathrm{d}^3\vec{x}\delta(\vec{x})\delta(\vec{x}+\vec{r}) \\ &= \int \mathrm{d}^3\vec{x}\int \frac{\mathrm{d}^3k}{(2\pi)^3}\hat{\delta}(\vec{k})e^{-i\vec{k}\cdot\vec{x}} \underbrace{\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\hat{\delta}(\vec{k}')e^{-i\vec{k}'\cdot(\vec{x}+\vec{r})}}_{\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\hat{\delta}(\vec{k}')e^{+i\vec{k}'\cdot(\vec{x}+\vec{r})}\ (\delta \text{ real})} \\ &= \left(\frac{1}{(2\pi)^3}\right)^2 \int \mathrm{d}^3\vec{x}\int\int \mathrm{d}^3k\, d^3k'\, \hat{\delta}(\vec{k})\hat{\delta}(\vec{k}')e^{-i(\vec{k}-\vec{k}')\cdot\vec{x}}e^{-i\vec{k}\vec{r}} \\ &\quad \text{using } \frac{1}{(2\pi)^3}\int \mathrm{d}^3x e^{i(\vec{k}-\vec{k}')\cdot\vec{x}} = \delta(\vec{k}-\vec{k}') \\ &= \frac{1}{(2\pi)^3}\int |\hat{\delta}(\vec{k})|^2 e^{i\vec{k}\cdot\vec{r}}\mathrm{d}^3k \\ &= \frac{1}{(2\pi)^3}\int |\hat{\delta}(k)|^2 e^{i\vec{k}\cdot\vec{r}}\mathrm{d}^3k \quad \text{where } \vec{k} \to k \text{ from isotropy} \\ &= \frac{1}{(2\pi)^3}\int P(k)e^{i\vec{k}\cdot\vec{r}}\mathrm{d}^3k \\ \Rightarrow &\boxed{\xi(r) = \frac{1}{(2\pi)^3}\int P(k)e^{i\vec{k}\cdot\vec{r}}\mathrm{d}^3k}\,. \end{aligned} \tag{408}$$

Observationally:

$$\xi(r) \approx \left(\frac{r}{r_0}\right)^{-1.8} \tag{409}$$

with $r_0 \approx 5\ h^{-1}$ Mpc for galaxies. Different objects have a different r_0, and more massive objects are more clustered, e.g. the cluster-cluster correlation function differs from the galaxy-galaxy correlation function: $\xi_{cc} \approx 20\xi_{gg}$.

2.D Form of the primordial power spectrum

There is no scale in the power spectrum $P(k) = Ak^n$. We want to know what n and A are. Initially, fluctuations on different scales should have the same amplitude on different scales.

Power spectrum index:
Fluctuations on certain mass or length scales are (0 is large scale, $k_{\max}$ is the smallest scale):

$$\begin{aligned} \sigma^2 &\approx \int_0^{k_{\max}} P(k)k^2 \mathrm{d}k \\ &= \int_0^{k_{\max}} Ak^{n+2} \mathrm{d}k \propto k_{\max}^{n+3} \\ \Rightarrow \sigma &\propto k_{\max}^{\frac{1}{2}(n+3)} \text{ or } \sigma \propto k^{\frac{1}{2}(n+3)} \end{aligned} \tag{410}$$

For mass, we get:

$$\begin{aligned} M \propto R^3 \propto k^{-3} \Rightarrow k \propto M^{-1/3} \\ \Rightarrow \sigma \propto M^{-\frac{1}{6}(n+3)} \end{aligned} \tag{411}$$

So:

$$\sigma \propto \begin{cases} k^{\frac{1}{2}(n+3)} \\ M^{-\frac{1}{6}(n+3)} \end{cases} \tag{412}$$

Does this tell us something about n? Modes can always grow outside the horizon, but we do not want "special" modes. All modes should therefore have the same σ, i.e. the same strength/fluctuation amplitude, when they enter the horizon.

The horizon mass, i.e. the mass within the horizon, is:

$$\begin{aligned} M_h &\propto \rho_m r_h^3 \\ \rho_m &\propto (1+z)^3 \end{aligned} \tag{413}$$

and

$$\begin{aligned} r_h &\propto \begin{cases} a^2 = (1+z)^{-2}, & \text{radiation dominated} \\ a^{3/2} = (1+z)^{-3/2}, & \text{matter dominated} \end{cases} \\ \Rightarrow M_h &\propto \begin{cases} (1+z_h)^{-3}, & \text{radiation dominated} \\ (1+z_h)^{-3/2}, & \text{matter dominated} \end{cases} \end{aligned} \tag{414}$$

σ grows:

$$\begin{aligned} \sigma &\propto \delta \\ \sigma &\propto \begin{cases} a^2 = (1+z)^{-2}, \text{ radiation dominated} \\ a^{3/2} = (1+z)^{-1}, \text{ matter dominated} \end{cases} \end{aligned} \tag{415}$$

We now find σ of the horizon mass, i.e. the fluctuation strength once this mode enters.

- radiation dominated case:

$$\sigma(z_h) = \sigma(z_p)\left(\frac{1+z_p}{1+z_h}\right)^2 \propto \sigma(z_p)(1+z_h)^{-2} \tag{416}$$

  where z_h is the redshift once mass M is within the horizon, and z_p is the redshift at the end of inflation. Then

$$\begin{aligned} \sigma(z_p) &\propto M^{-\frac{1}{6}(n+3)} \\ M_h = M &\propto (1+z_h)^{-3} \Rightarrow (1+z_h)^{-2} \propto M^{2/3} \end{aligned} \tag{417}$$

  so we find:

$$\sigma(z_h) \propto M^{-\frac{1}{6}(n+3)} M^{2/3} = M^{-(\frac{1}{2}+\frac{n-4}{6})} \tag{418}$$

- matter dominated case
  This follows the same calculation, so we get the same result and the fluctuation of a mode once it enters the horizon is:

$$\sigma(z_h) \propto M^{-(\frac{1}{2}+\frac{n-4}{6})} \tag{419}$$

Now, we do not want "special" modes, so $\sigma(z_h)$ should not depend on n! We get $n \approx 1$ according to the Harrison-Zel'dovich spectrum.

Power spectrum amplitude:
n can be calculated with theory from inflation, but the amplitude comes from observations. We measure the number of fluctuations in galaxy surveys within a sphere of $8\,\mathrm{Mpc}/h$, or σ_8. The fluctuations in galaxies are not exactly the fluctuations in mass:

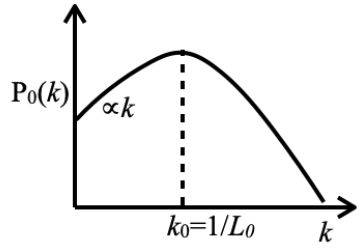
$$\sigma_{8,\mathrm{gal}} = b\sigma_{8,\mathrm{mass}} \tag{420}$$

where b is the *bias* of the galaxy clustering compared to the mass fluctuations. Observationally, $\sigma_{8,\mathrm{gal}} \approx 1$. From WMAP and SDSS, we have:

$$\begin{aligned} n &= 0.953 \pm 0.016 \\ \sigma_8 &= 0.756 \pm 0.035 \end{aligned} \tag{421}$$

Transfer function:
We found that modes entering the horizon during the radiation dominated phase do not grow

(stagnation). The primordial power spectrum is therefore modified by the transfer function:



$$P_0(k) = (Ak)T^2(k),$$
$$T(K) \approx \begin{cases} 1, \frac{1}{k} \gg L_0 \\ \frac{1}{k^2}, \frac{1}{k} \ll L_0 \end{cases} \tag{422}$$

where L_0 is the comoving horizon at z_{equality}.

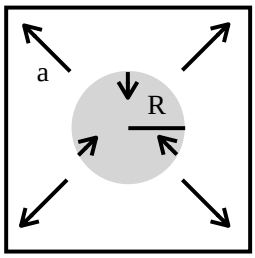
2.E Nonlinear evolution: spherical collapse

For $\delta \ll 1$, we can use linear perturbation theory, but for $\delta \sim 1$, nonlinear evolution begins and halos form. This requires simulations.

Halos:

- A distribution of dark matter as a collection of nearly spherical overdense clouds to form halos.
- We study the dynamics of spherical, homogeneous overdensities for a basic understanding. This is the spherical collapse model.

Spherical collapse model:
We consider an overdense sphere in an Einstein-de Sitter cosmology. The overdensity will eventually reach a maximum radius and then collapse to a virialized halo because the gravity within the overdensity is stronger.



$$\begin{aligned} H &= H_0 a^{-3/2} && \text{Friedmann equation for Einsten-de Sitter} \\ x &= \frac{a}{a_{\text{ta}}} && a_{\text{ta}} \text{ is the scale factor at maximum expansion} \\ y &= \frac{R}{R_{\text{ta}}} && \text{radius in units of maximum radius} \end{aligned} \tag{423}$$

We can simplify:

$$\begin{aligned} \tau &= H_{\text{ta}} t \quad (\text{with } H_{\text{ta}} = H_0 a_{\text{ta}}^{-3/2}) \\ \Rightarrow x' &= \frac{\mathrm{d}x}{\mathrm{d}\tau} = \frac{1}{H_{\text{ta}}} \frac{\dot{a}}{a_{\text{ta}}} = \frac{H}{H_{\text{ta}}} x = x^{-1/2} \\ &\left(\text{using } \frac{H}{H_{\text{ta}}} = \frac{H_0 a^{-3/2}}{H_0 a_{\text{ta}}^{-3/2}} = \frac{a^{-3/2}}{a_{\text{ta}}^{-3/2}} = x^{-3/2} \text{ for the final equality}\right) \end{aligned} \tag{424}$$

So

$$\boxed{x' = x^{-1/2}} \tag{425}$$

We use the Newtonian equation of motion for the radius R:

$$\ddot{R} = -\frac{GM}{R^2} = -\underbrace{\frac{4\pi}{3}\rho_{\rm ta}R_{\rm ta}^3}_{\text{enclosed mass stays the same}}\frac{G}{R^2} \tag{426}$$

We can rewrite this:

$$\rho_{\rm ta} = \frac{3H_{\rm ta}^2}{8\pi G}\xi \tag{427}$$

where ξ is the overdensity parameter, which is the overdensity of the halo with respect to the background at turnaround ($\xi > 1$ for overdensities). Then using τ and y, we have:

$$\boxed{y'' = -\frac{\xi}{2y^2}} \tag{428}$$

with the boundary conditions

$$\begin{aligned} y'|_{x=1} &= 0 \\ y|_{x=0} &= 0 \end{aligned} \tag{429}$$

and we can solve the equations:

$$\begin{aligned} x' &= x^{-1/2} \\ y'' &= -\frac{3}{2y^2} \end{aligned} \tag{430}$$

Then we get an implicit solution for x:

$$x' = x^{-1/2} \Rightarrow \boxed{\tau = \frac{2}{3}x^{3/2}} \tag{431}$$

So

$$\begin{aligned} x &= \left(\frac{3}{2}\right)\tau^{2/3} \\ \frac{{\rm d}x}{{\rm d}t} &= \frac{2}{3}\left(\frac{3}{2}\right)^{2/3}\tau^{-1/3} = x^{-1/2} \end{aligned} \tag{432}$$

We also have

$$y' = \pm\sqrt{\xi}\sqrt{\frac{1}{y} - 1} \tag{433}$$

using the first boundary condition. We also use the + before turnaround and the − after.

Then

$$\begin{aligned}\frac{dy'}{d\tau} &= \pm\sqrt{\xi}y'\frac{\mathrm{d}}{\mathrm{d}y}\left(\left(\frac{1}{y}-1\right)^{1/2}\right)\\ &= \pm\sqrt{\xi}y'\frac{1}{2}\left(\frac{1}{y}-1\right)^{-1/2}(-y^{-2})\\ &= -\frac{\xi}{2y^2}\left(\pm\frac{1}{\sqrt{\xi}}y'\left(\frac{1}{y}-1\right)^{-1/2}\right)\\ &= -\frac{\xi}{2y^2}\underbrace{\left(y'\left(\pm\sqrt{\xi}\sqrt{\frac{1}{y}-1}\right)\right)^{-1}}_{=1}\end{aligned} \tag{434}$$

Integrating before turnaround and using the second boundary condition gives us an implicit solution for y:

$$\boxed{\tau = \frac{1}{\sqrt{\xi}}\left(\frac{1}{2}\arcsin(2y-1)-\sqrt{y-y^2}+\frac{\pi}{4}\right)}. \tag{435}$$

At turnaround:

$$\begin{aligned}x = 1 = y, \tau &= \frac{2}{3}\\ \Rightarrow \frac{2}{3} &= \frac{1}{\sqrt{\xi}}\left(\frac{1}{2}\underbrace{\arcsin(1)}_{\pi/2}+\frac{\pi}{4}\right) = \frac{1}{\sqrt{\xi}}\frac{\pi}{2}\\ \Rightarrow \xi &= \left(\frac{3\pi}{4}\right)^2\end{aligned} \tag{436}$$

so we get the overdensity parameter ξ.

At collapse:
We assume symmetry, so we get collapse at $\tau = \frac{4}{3}$. Then

$$x_c = \left(\frac{3}{2}\right)^{2/3}\tau^{2/3} = \left(\frac{3}{2}\right)^{2/3}\left(\frac{4}{3}\right)^{2/3} = 4^{1/3} \tag{437}$$

Collapse parameters:

- Linearly extrapolated values:
  at early times, $y \ll 1$, so
  $$\tau \approx \frac{8}{9\pi}y^{3/2}\left(1+\frac{3y}{10}\right). \tag{438}$$
  The overdensity inside the halo relative to the background is:
  $$\Delta = \left(\frac{x}{y}\right)\xi. \tag{439}$$

Note here that ξ is the overdensity at turnaround, and we want to find Δ at collapse. The background density is proportional to x^{-3}, and the halo density is proportional to y^{-3}. We also know that $x = y = 1$ at turnaround, where $\Delta = \xi$. Now

$$\begin{aligned}
\tau = \frac{2}{3}x^{3/2} \Rightarrow \left(\frac{2}{3}\right) x^{3/2} &\approx \frac{8}{9\pi} y^{3/2}\left(1+\frac{3y}{10}\right) \\
\Rightarrow \left(\frac{x}{y}\right)^{3/2} &= \frac{3}{2}\frac{8}{9\pi}\left(1+\frac{3y}{10}\right) \\
\Rightarrow \left(\frac{x}{y}\right)^{3} &= \underbrace{\left(\frac{4}{3\pi}\right)^2}_{=1/\xi} \underbrace{\left(1+\frac{3y}{10}\right)^2}_{\approx\left(1+\frac{3y}{5}\right)} \\
\Rightarrow \Delta = \left(\frac{x}{y}\right)^{3} \xi &= 1+\frac{3y}{5}
\end{aligned} \tag{440}$$

linear density contrast (assuming $y \ll 1$):

$$\delta = \Delta - 1 = \frac{3y}{5} \tag{441}$$

– The linearly extrapolated density contrast at turnaround is:

$$\delta_{\rm ta} = \frac{a_{\rm ta}}{a}\delta = \frac{\delta}{x} = \frac{3y}{5x} \tag{442}$$

since linear perturbations δ grow like the scale factor. Now

$$\frac{1}{x} = \left(\frac{3\tau}{2}\right)^{-2/3} \approx \left(\frac{3\pi}{4}\right)^{2/3}\frac{1}{y} \tag{443}$$

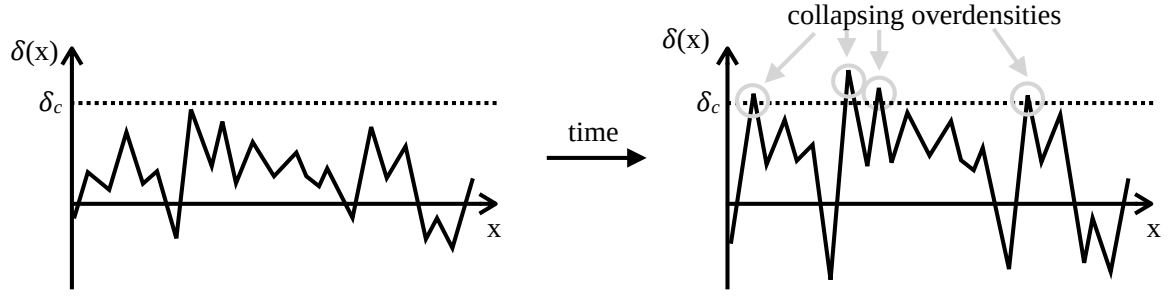
using the lowest order in y. We can then insert this into $\delta_{\rm ta}$ and get:

$$\delta_{\rm ta} = \frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} \approx 1.06 \tag{444}$$

– The linearly extrapolated density contrast at collapse is:

$$\delta_c = \frac{a_c}{a_{\rm ta}}\delta_{\rm ta} = x_c\delta_{\rm ta} = 4^{1/3}\delta_{\rm ra} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \approx 1.69 \tag{445}$$

So the halo can be considered collapsed when its density contrast expected from lineary theory has reached δ_c. If we draw a density field as a function of one-dimensional space, we can identify which overdensities will collapse at a given time:



- Nonlinear values:
  We now look at the potential energy of a halo:

$$\begin{aligned} &\text{at turnaround: } E = V_{\text{ta}} && \text{(no kinetic energy)} \\ &\text{at collapse: } E = T_c + V_c = \frac{1}{2}V_c && \text{(virial theorem: } 2T_c + V_c = 0) \\ &\Rightarrow V_{\text{ra}} = E = \frac{1}{2}V_c \Rightarrow V_c = 2V_{\text{ta}} \end{aligned} \tag{446}$$

  Since potential energy is proportional to $\frac{1}{r}$ and $y = 1$ at turnaround, we know that $y = \frac{1}{2}$ at virialization. Then we get the overdensity at this time:

$$\Delta_V = \left(\frac{x_c}{y}\right)^3 \xi = \left(\frac{4^{1/3}}{\frac{1}{2}}\right)^3 \xi = 32\xi = 32\left(\frac{3\pi}{4}\right)^2 = 18\pi^2 \approx 178 \tag{447}$$

  A halo in virial equilibrium is expected to have a mean density of ~ 178 higher than the background. This is why masses and radii of halos are often quoted as M_{200}, which is the mass enclosed in a sphere of radius R_{200} with an average density 200 times the mean or critical density of the Universe.

2.F Press-Schechter mass function

We want to know the halo mass function, i.e. the number density of a given mass of halos at a given redshift.

Analytic derivation:
We consider a halo of mass M. The characteristic length scale is then $R(M) = R$:

$$\begin{aligned} \frac{4\pi}{3}R^3\rho_c(z)\Omega_m(z) &= M \\ \Rightarrow R(M) &= \left(\frac{3M}{4\pi\rho_c(z)\Omega_m(z)}\right)^{1/3} \end{aligned} \tag{448}$$

Halos of mass M are then forming if the smoothed density field $\bar{\delta}$ crosses $\delta_c = 1.69$:

$$\bar{\delta}(\vec{x}) = \int \mathrm{d}^3y \delta(\vec{x}) W_R(|\vec{x} - \vec{y}|) \tag{449}$$

where W_R is the window function.

The variance on the scale $R(M)$ is:

$$\sigma_R^2 = \frac{1}{2\pi}\int_0^\infty k^2 \mathrm{d}k P(k)\hat{W}_R(k) \; . \tag{450}$$

Inflation produces a Gaussian random field, so the probability of finding a smoothed density contrast $\bar{\delta}(\vec{x})$ at a given point in space $\vec{x}$ is:

$$p(\bar{\delta}(\vec{x}), z) = \frac{1}{\sqrt{2\pi\sigma_R^2(z)}} e^{-\frac{\bar{\delta}^2(\vec{x})}{2\sigma_R^2(z)}} \tag{451}$$

where $\sigma_R(z)$ is the linearly evolved $\sigma_R : \sigma_R(z) = \sigma_R D(z)$ for growth factor D.

The Press-Schechter idea is that the probability of finding the filtered density contrast at or above the linear density contrast for spherical collapse, $\bar{\delta} > \delta_c$, is equal to the fraction of volume filled with halos of mass M:

$$F(M,z) = \int_{\delta_c}^{\infty} d\bar{\delta} p(\bar{\delta}, z) = \frac{1}{2}\mathrm{erfc}\left(\frac{\delta_c}{\sqrt{2}\sigma_R(z)}\right) . \tag{452}$$

The distribution of halos over mass M is simply $\frac{\partial F(M,z)}{\partial M}$. To calculate this, we need:

$$\frac{\partial}{\partial M} - \frac{d\sigma_R(z)}{dM}\frac{\partial}{\partial \sigma_R(z)} = \frac{d\sigma_R}{dM}\frac{\partial}{\partial \sigma_R} . \tag{453}$$

Using $\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{erfc}(x) = -\frac{2}{\sqrt{\pi}}e^{-x^2}$, we get:

$$\begin{aligned}
\frac{\partial F(M,z)}{\partial M} &= \frac{\mathrm{d}\sigma_R}{\mathrm{d}M}\frac{\partial}{\partial \sigma_R}\left(\frac{1}{2}\mathrm{erfc}\left(\frac{\delta_c}{\sqrt{2}D(z)\sigma_R}\right)\right) \\
&= \frac{\mathrm{d}\sigma_R}{\mathrm{d}M}\frac{1}{2}\left(-\frac{\delta_c}{\sqrt{2}D(z)\sigma_R^2}\right)\left(-\frac{2}{\sqrt{\pi}}e^{-\frac{\delta_c^2}{2\sigma_R^2 D^2(z)}}\right) \\
&= \frac{\mathrm{d}\sigma_R}{\mathrm{d}M}\frac{\delta_c}{\sqrt{2\pi}\sigma_R^2 D(z)}e^{-\frac{\delta_c^2}{2\sigma_R^2 D^2(z)}} \\
&= \frac{1}{\sqrt{2\pi}}\frac{\delta_c}{\sigma_R D(z)}\frac{\mathrm{d}\ln(\sigma_R)}{\mathrm{d}M}e^{-\frac{\delta_c^2}{2\sigma_R^2 D^2(z)}}
\end{aligned} \tag{454}$$

so

$$\frac{\partial F(M,z)}{\partial M}\mathrm{d}M = \text{fraction of volume filled with halos of mass } [M, M+\mathrm{d}M]. \tag{455}$$

We must convert $\frac{\partial F}{\partial M}$ to an actual halo mass function. We convert to comoving number density by dividing by the mean volume $M/\rho_c(z)\Omega_M(z)$ occupied by mass M halos:

$$\frac{\partial F(M,z)}{\partial M} = \frac{1}{\sqrt{2\pi}}\frac{\rho_c(z)\Omega_M(z)\delta_c}{\sigma_R D(z)}\frac{\mathrm{d}\ln(\sigma_R)}{\mathrm{d}M}e^{-\frac{\delta_c^2}{2\sigma_R^2 D^2(z)}}\frac{1}{M} \tag{456}$$

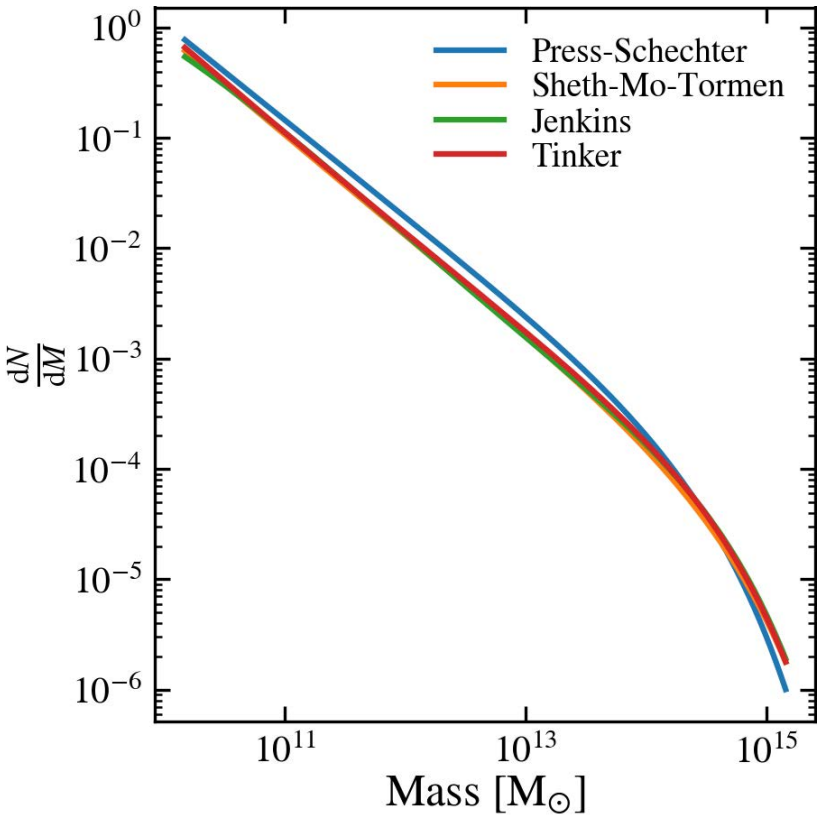
However, we need a fudge factor for the mass function to work. We require

$$\int_0^1 \mathrm{d}M \frac{\partial F(M,z)}{\partial M} = 1 \tag{457}$$

since $\frac{\partial F}{\partial M}$ is a volume fraction. But we get $\frac{1}{2}$ using $\frac{\partial F(M,z)}{\partial M}$ above! We therefore add a factor of two:

$$\boxed{N(M,z) = \sqrt{\frac{2}{\pi}}\frac{\rho_c(z)\Omega_M(z)\delta_c}{\sigma_R D(z)}\frac{\mathrm{d}\ln(\sigma_R)}{\mathrm{d}M}e^{-\frac{\delta_c^2}{2\sigma_R^2 D^2(z)}}\frac{1}{M}} . \tag{458}$$

See Problem Set 6 for a description of the Extended Press-Schechter formalism that explains the fudge factor.



Here we show the mass function for several theoretical models from Press and Schechter 1973, Sheth, Mo, and Tormen 2002, Jenkins et al. 2002, and Tinker et al. 2008. The lines are fairly similar, although the Press-Schechter deviates slightly more from the other models.

Part III

CMB, BBN, and Thermal History of the Universe

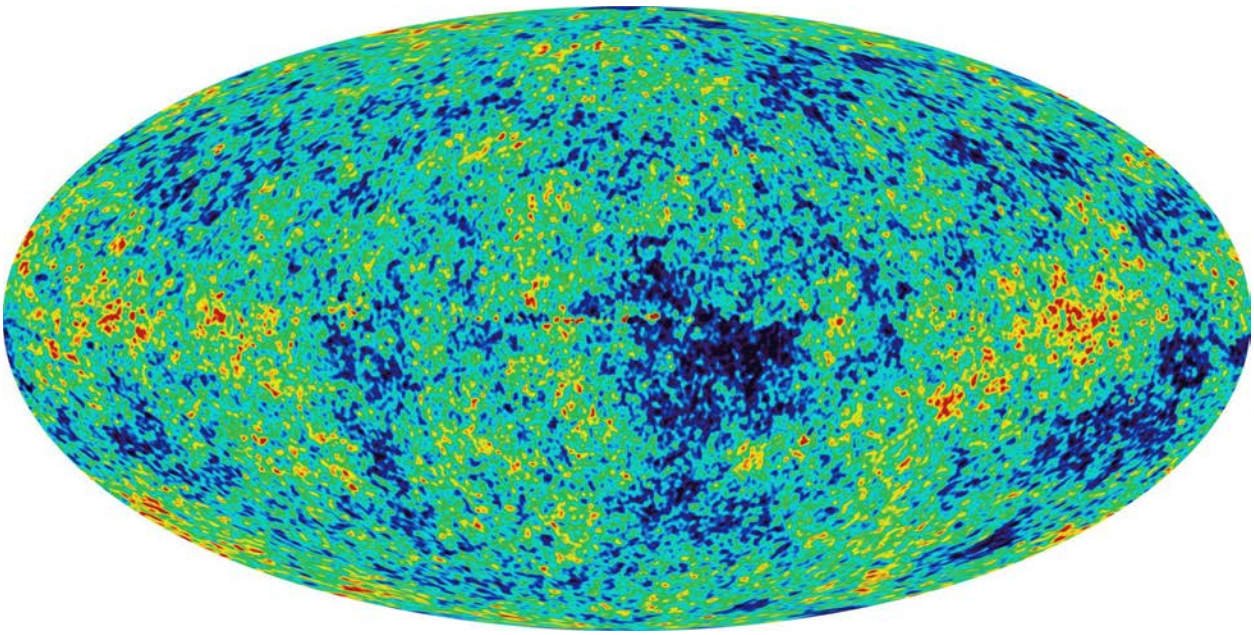


Image: NASA. Image is in the public domain.

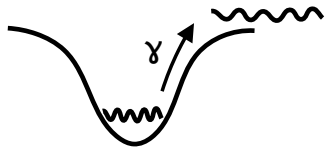
So far, we have mostly discussed the late-time evolution of the Universe (except inflation). We now study the early phases.

1 The Cosmic Microwave Background

1.A Basic picture of the CMB

At $z \sim 1000$, photons decouple from matter (previously coupled due to Thomson scattering). At that time, the dark matter has already formed dark matter potential wells, which leads to perturbations in baryons. This leads to temperature fluctuations in the CMB $\frac{\delta T}{T} \sim 10^{-5}$ (note: without dark matter, we would expect $\frac{\delta T}{T} \sim 10^{-3}$). This leads to anisotropies in the CMB.

Primary anisotropies:
These are anisotropies caused by properties of the CMB.

- Large scales:
  - 

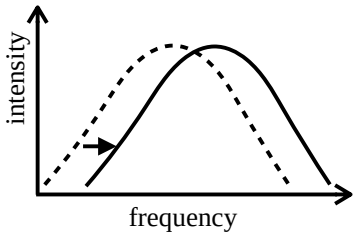
Dark matter potential wells lead to gravitational redshift and gravitational time delay. Photons were scattered earlier (and then delayed), so they were higher temperature.

    Both effects (gravitational redshift and time delay) always happen together. This is the *Sachs-Wolfe effect*.
  - Doppler effect due to the peculiar motion of electrons.
- On scales larger than the horizon, baryons follow dark matter, leading to higher temperatures in dark matter wells.
- On scales smaller than the horizon, baryons feel radiation pressure. This leads to *baryonic acoustic oscillations* (BAO).
- On very small scales, the imperfect coupling between photons and electrons leads to diffusion. Fluctuations are smeared out and damped on scales $\leq 5'$. This is called *Silk damping*.

Secondary anisotropies:
These impact the measurements of the CMB due to effects on photons as the travel from the CMB to us.

- Thomson scattering of CMB photons: the Universe was reionized by the first stars, galaxies, and quasars between $z \sim 1000$ and $z \sim 6$. These photons then experience Thomson scattering with free electrons as they travel through space. The scattering is isotropic, so it results in an overall reduction of CMB anisotropies.

- Integrated Sachs-Wolfe effect: photons experience gravitational potential and time delays as they travel through structures in the Universe.
- Gravitational lensing from structures in the Universe.
- Sunyaev-Zel'dovich (SZ) effect: CMB photons passing through the hot intergalactic medium of galaxies Thomson scatter with electrons. This reduces the intensity for lower frequencies and increases the intensity for large frequencies, resulting in a shift in the spectrum.



1.B Describing anisotropies and the fluctuation spectrum:

We focus now on understanding the primary anisotropies. We have three main effects:

- Large scales: Sachs-Wolfe and Doppler effects roughly compensate each other. The photons then provide an imprint of the dark matter distribution.
- Smaller scales: Baryonic acoustic oscillations of the photon-baryon plasma.
- Smallest scales: Silk damping due to photon diffusion.

We need to quantify the temperature fluctuations on the sky. We decompose the fluctuations into spherical harmonics:

$$T(\vec{\theta}) = \sum_{l,m} a_{lm} Y_l^m(\vec{\theta}) \tag{459}$$

where $\vec{\theta} = (\theta, \varphi)$ and a_{lm} are the complex coefficients

$$a_{lm} = \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\theta \sin\theta T(\theta,\varphi) Y_{l_1}^{m*}(\theta,\phi) \tag{460}$$

since

$$\int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\theta \sin\theta Y_{l_1}^{m_1 *}(\theta,\varphi) Y_{l_2}^{m_2}(\theta,\varphi) = \delta_{l_1 l_2}\delta_{m_1 m_2} \ . \tag{461}$$

We define the *power spectrum*:

$$C_l = \left\langle |a_{lm}|^2 \right\rangle \tag{462}$$

averaging over m. We often plot $l(l+1)C_l$ and define this as the amplitude of fluctuations on the angular scale $\theta \sim \frac{\pi}{l} = \frac{180^\circ}{l}$. $l = 1$ is the dipole anisotropy due to the motion of Earth and $l = 2$ is the quadrupole anisotropy.

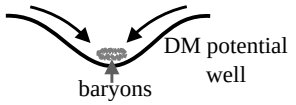
Fluctuations on large scales:
At $z = z_{\mathrm{rec}} \sim 1000$, there is a characteristic scale the horizon with angle:

$$\varphi_{\mathrm{horizon,rec}} \approx 1.7^\circ \sqrt{\Omega_{m,0}} \tag{463}$$

so for a flat universe: $\varphi_{\text{horizon,rec}} \approx 1.7°$. Then for $\theta \gg 1.7°$, large-scale effects dominate (Sachs-Wolfe and Doppler) and there are no baryonic acoustic oscillations. Then C_l reflect the matter power spectrum $P(k)$ on large scales. For $P(k) \propto k$ (Harrison-Zel'dovich spectrum), $l(l+1)C_l$ is approximately constant for $l \ll \frac{180°}{1.7°} \approx 100$.

Fluctuations on small scales:
For $\theta \ll 1.7°$, physical effects can act.

The baryon-photon fluid has a sound speed of $c_s \approx c/\sqrt{3}$. Then the largest wavelengths such that the wave can have half an oscillation (compression) until $z_{\text{rec}}, t_{\text{rec}}$ is:



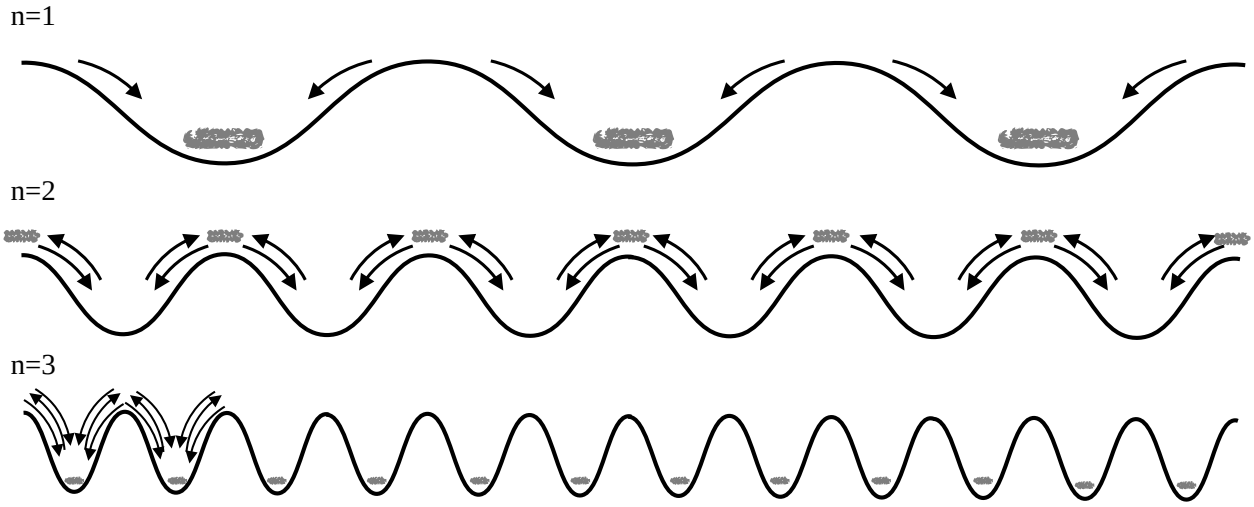
$$\lambda_{\max} = t_{\text{rec}} c_s = t_{\text{rec}} \frac{c}{\sqrt{3}} \,. \tag{464}$$

The sound horizon is $\frac{1}{\sqrt{3}}$ times smaller than the horizon. The angular scale is then

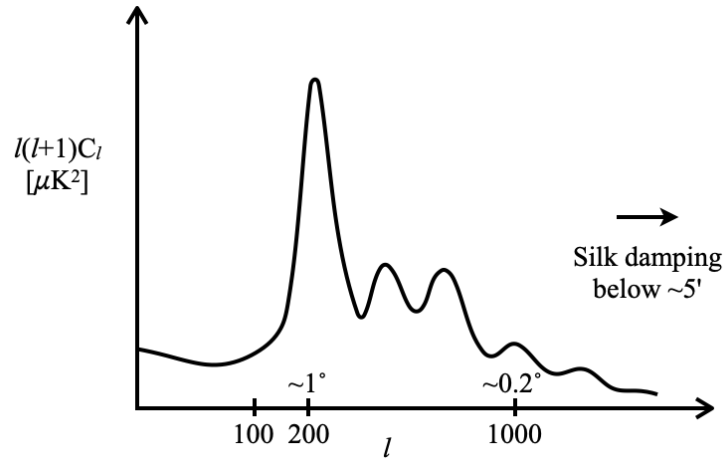
$$\begin{aligned} \theta_1 &\approx \frac{1.7°}{\sqrt{3}} \sim 1° \\ l_1 &\approx 200 \end{aligned} \tag{465}$$

so we can expect the first peak in $l(l+1)C_l$ there since baryons are compressed. Adiabatic compression and the Doppler effect lead to temperature fluctuations on that scale.

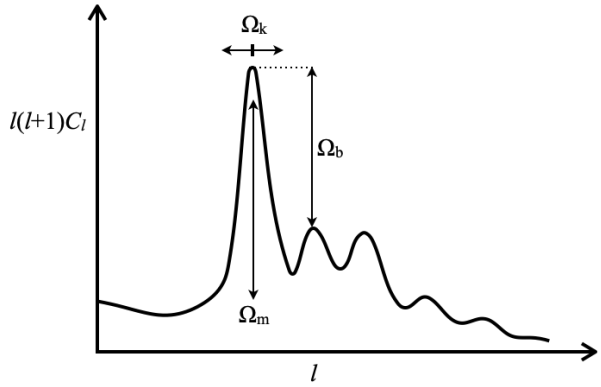
The second peak occurs for scales for which one full oscillation is possible and so forth:



Peaks happen at stationary points of oscillations. We can draw the power spectrum:



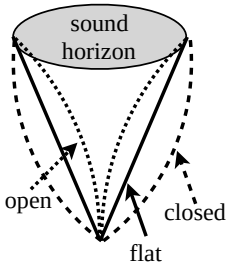
1.C Cosmology with the CMB



We can derive the cosmological parameters Ω_k, Ω_m, and Ω_b from the first and second peaks of the power spectrum.

- Derive Ω_k from the first peak:
  The actual extent of the first peak does not strongly depend on Ω_k, but the angular scale/angular diameter distance is sensitive to Ω_k.

  In an open universe, the angular size of the sound horizon will appear smaller, and the first peak will move to larger l. In a closed universe, it will appear larger, so l will be lower for the first peak.



- Derive Ω_m from the first peak:
  A naive assumption might lead us to believe that more matter means more gravity and so bigger peaks. However, the timing effect is more important!

  Consider two values of Ω_m ($\Omega_1 > \Omega_2$). For larger Ω_m, the universe is younger for a given redshift (e.g., $z_{\rm rec}$).

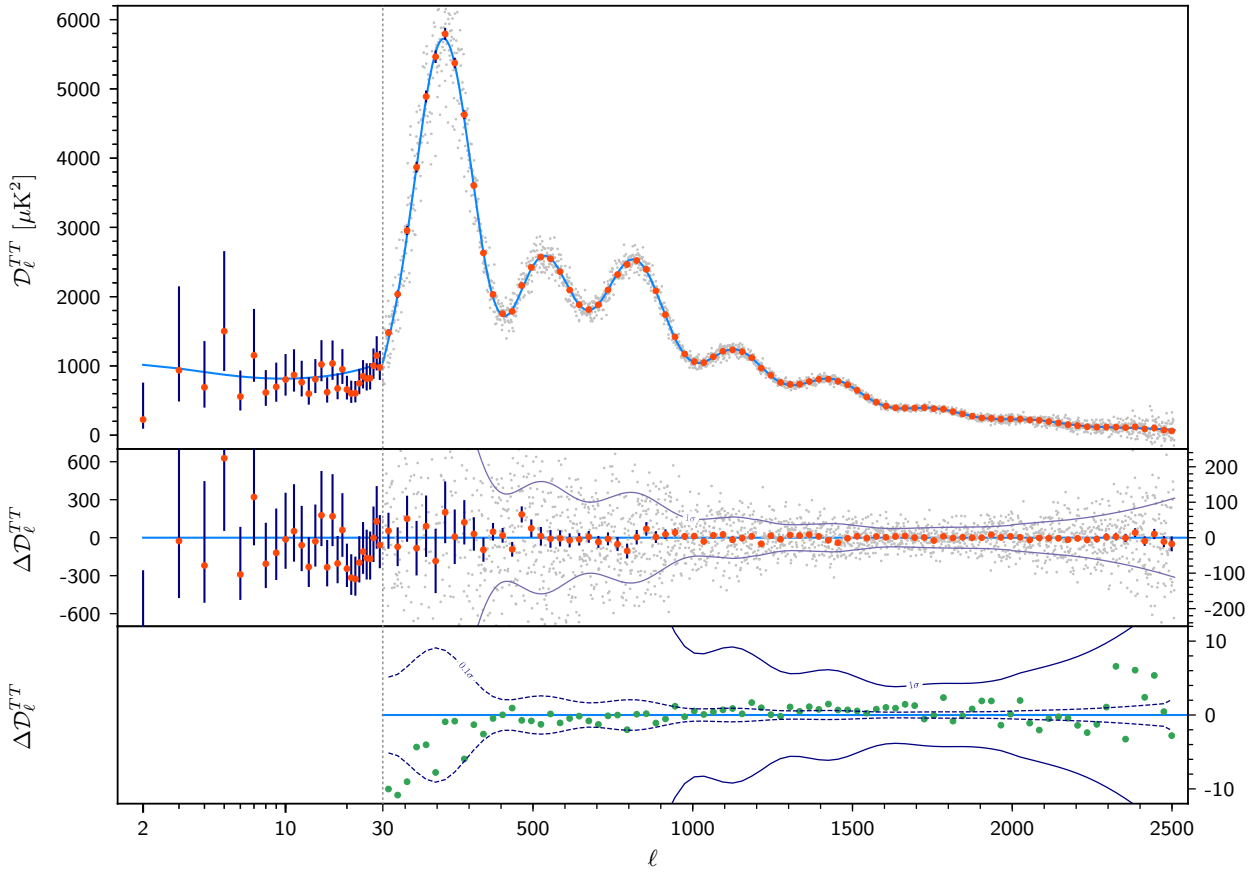
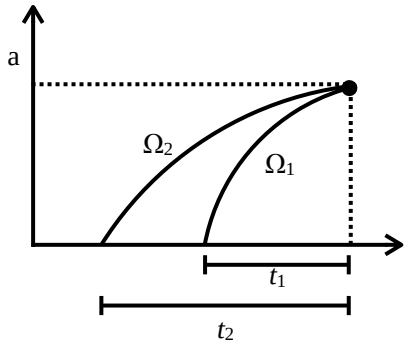


Figure 1: [Fig. 57 of Planck Collaboration V. 2020, A&A, 641, A5] Planck 2018 temperature power spectrum. At multipoles $\ell \geq 30$ we show the frequency-coadded temperature spectrum computed from the `Plik` cross-half-mission likelihood, with foreground and other nuisance parameters fixed to a best fit assuming the base-ΛCDM cosmology. In the multipole range $2 \leq \ell \leq 29$, we plot the power-spectrum estimates from the `Commander` component-separation algorithm, computed over 86% of the sky (see Sect. 2.1.1). The base-ΛCDM theoretical spectrum best fit to the likelihoods is plotted in light blue in the *upper panel*. Residuals with respect to this model are shown in the middle panel. The vertical scale changes at $\ell = 30$, where the horizontal axis switches from logarithmic to linear. The error bars show $\pm 1\sigma$ diagonal uncertainties, including cosmic variance (approximated as Gaussian) and not including uncertainties in the foreground model at $\ell \geq 30$. The 1σ region in the middle panel corresponds to the errors of the unbinned data points (which are in grey). *Bottom panel*: difference between the 2015 and 2018 coadded high-multipole spectra (green points). The 1σ region corresponds to the binned data errors. The vertical scale differs from the one of the middle panel. The trend seen for $\ell < 300$ corresponds to the change in the dust correction model described in Sect. 3.3.2.

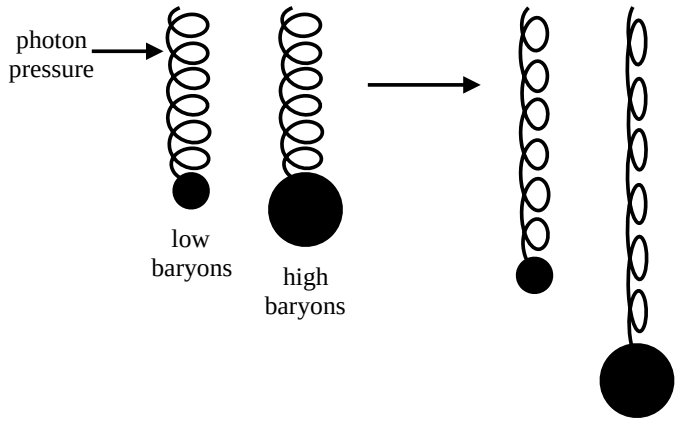


$\Omega_1 > \Omega_2$ leads to $t_1 < t_2$, so there's not as much time to form structures and we get a smaller peak!

Note: based on the first peak, we get Ω_m and Ω_k, so also Ω_Λ (assuming flat universe)!

- Derive Ω_b from the second peak:
  Ω_b is degenerate with Ω_m, so we need the second peak. This can also be derived from Big Bang nucleosynthesis.

  The idea is that a higher baryon mass is like adding mass to a spring ("baryon loading"). More mass causes a deeper fall:



  With more loading, the mass falls deeper but rebounds to the same position. Thus, odd peaks are associated with compression—i.e., how deep the baryons fall into the well. Those peaks get enhanced with more baryons, so the second peak is compressed compared to the first peak. We can therefore constrain Ω_b with the ratio of the two peaks.

2 Thermal history of the Universe

The main idea is that the Universe was very hot in the beginning since $T \propto (1+z)$. For a particle with mass m_x and temperature such that $kT \gtrsim m_x c^2$, we have creation and annihilation reactions. Once T falls low enough, we get *freeze-out* and the reactions stop, freezing the abundance of those particles. (Note: $1\,\mathrm{eV} = 1.1605 \times 10^4\, k_B\mathrm{K} \Rightarrow 1\,\mathrm{eV} \leftrightarrow 10^4\,\mathrm{K}$.)

We first discuss in some more depth the freeze-out of dark matter, which happens around $10 - 100\,\mathrm{GeV}$. We then briefly discuss the remaining thermal history for temperatures below $\sim 16\,\mathrm{GeV}$ (standard model physics). Big Bang nucleosynthesis will be discussed in the next chapter.

2.A Thermal history of dark matter

Where does dark matter come from? At early times ($T > 10^{12}\,\mathrm{K} \cong 100\,\mathrm{GeV}$), we have $kT \geq m_x c^2$ for leading dak matter candidates. For non-relativistic particles in equilibrium:

$$n_{\mathrm{eq}} = g \left(\frac{mkT}{2\pi\hbar^2} \right)^{3/2} e^{-\frac{mc^2}{kT}} \,. \tag{466}$$

We have equilibrium between creation and annihilation:

$$x + \bar{x} \leftrightharpoons 2\gamma \,. \tag{467}$$

For creation rate ψ and annihilation rate $n^2\langle\sigma v\rangle$, where $\langle\sigma v\rangle$ is the velocity averaged cross section for annihilation, we have:

$$= n_{\mathrm{eq}}^2 \langle\sigma v\rangle \,. \tag{468}$$

At late times, kT falls below mc^2, so $e^{-mc^2/kT} \to 0$. If annihilation continues to happen, no particles will be left since new particles cannot be created at low temperatures, which would leave no *relic abundance*. However, the annihilation rate $n^2\langle\sigma v\rangle$ also goes down since $n \propto a^{-3}$. If there's no creation or annihilation:

$$\begin{aligned} \text{comoving: } \frac{\mathrm{d}n_c}{\mathrm{d}t} &= 0 \quad \left(n_c = n \left(\frac{a}{a_0} \right)^{-3} \right) \\ \frac{\mathrm{d}n}{\mathrm{d}t} &= -3\frac{\dot{a}}{a} n \\ \Rightarrow \frac{\mathrm{d}n}{\mathrm{d}t} &+ 3Hn = 0 \end{aligned} \tag{469}$$

Thus, annihilation will stop and then there will be a relic abundance. The abundance equations with reactions is:

$$\frac{\mathrm{d}n_c}{\mathrm{d}t} = -\langle\sigma v\rangle \left(n_c^2 - n_{c,\mathrm{eq}}^2 \right) \tag{470}$$

so the reactions drive n_c towards the equilibrium value.

There are two competing timescales: the expansion of the Universe and the mean interaction timescale. We can rewrite the above equation:

$$\begin{aligned} \frac{\mathrm{d}n_c}{\mathrm{d}a} \underbrace{\frac{\mathrm{d}a}{\mathrm{d}t}}_{\dot{a}} &= -\langle\sigma v\rangle n_{c,\mathrm{eq}}^2 \left[\left(\frac{n_c}{n_{c,\mathrm{eq}}} \right)^2 - 1 \right] \\ &\left(H = \frac{\dot{a}}{a} \right) \\ \frac{a}{n_{c,\mathrm{eq}}} \frac{\mathrm{d}n_c}{\mathrm{d}a} &= -\frac{\langle\sigma v\rangle n_{c,\mathrm{eq}}}{H} \left[\left(\frac{n_c}{n_{c,\mathrm{eq}}} \right)^2 - 1 \right] \,. \end{aligned} \tag{471}$$

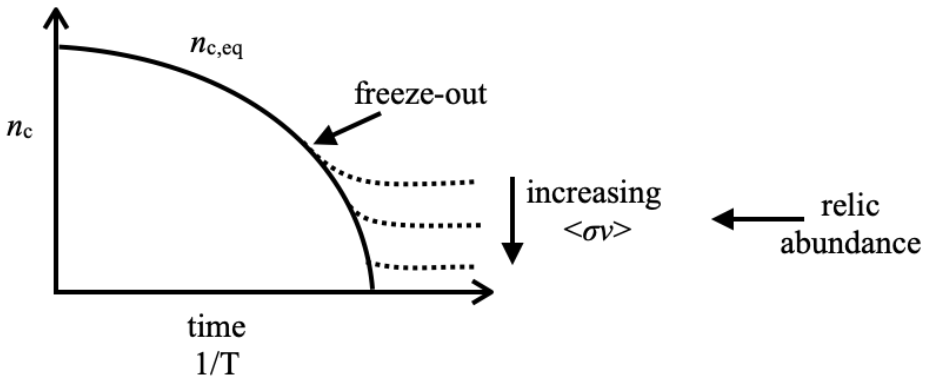
There are two timescales in this equation: $\tau = 1/H$ and $\tau_{\rm coll} = 1/(n_{\rm eq}\langle\sigma v\rangle)$. Then

$$\frac{a}{n_{c,\rm eq}}\frac{{\rm d}n_c}{{\rm d}a} = -\frac{\tau_H}{\tau_{\rm coll}}\left[\left(\frac{n_c}{n_{c,\rm eq}}\right)^2 - 1\right] . \tag{472}$$

We have two regimes that give us two solutions:

- At early times, $\tau_{\rm coll} \ll \tau_H \Rightarrow n_c \approx n_{c,\rm eq}$.
- At late times, $\tau_{\rm coll} \gg \tau_h \Rightarrow n_c \approx$ constant $\approx n_{c,\rm eq}(z_{\rm freeze})$.

At redshift $z_{\rm freeze}$, we have $\tau_{\rm coll} \sim \tau_H$, so particles freeze out of equilibrium and the comoving number density stays fixed.



From observed relic abundances, we get m and σ at the electroweak scale, which is predicted for WIMPs! This is known as the WIMP miracle. So far, however, nothing has been detected.

Hot and cold dark matter:
Are particles moving relativistically (hot dark matter) or non-relativistically (cold dark matter) at freeze-out? Particles become non-relativistic when:

$$3kT(t_{\rm nr}) \approx mc^2 . \tag{473}$$

There are two cases:

- $t_{\rm nr} > t_{\rm freeze} \Rightarrow$ hot relic and hot dark matter
- $t_{\rm nr} < t_{\rm freeze} \Rightarrow$ cold relic and cold dark matter

Hot dark matter:
Consider an analogy to the Jeans length, the free-streaming length:

$$\lambda = v\sqrt{\frac{\pi}{G\rho}} . \tag{474}$$

This wipes out structures on small scales! Since $v \approx c$, scales below the horizon are suppressed. But particles slow down due to expansion:

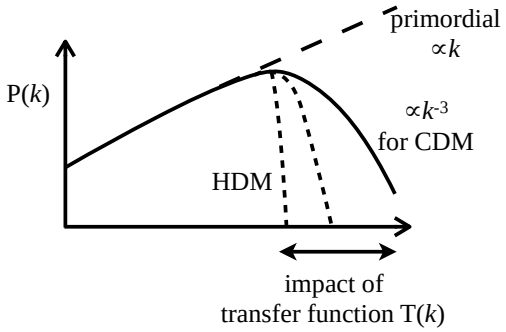
$$v_{\text{pec}} = v - Hd \Rightarrow v \propto a^{-1} \tag{475}$$

and the particle becomes non-relativistic once $mc^2 \sim 3kT$ (relic time and temperature), so:

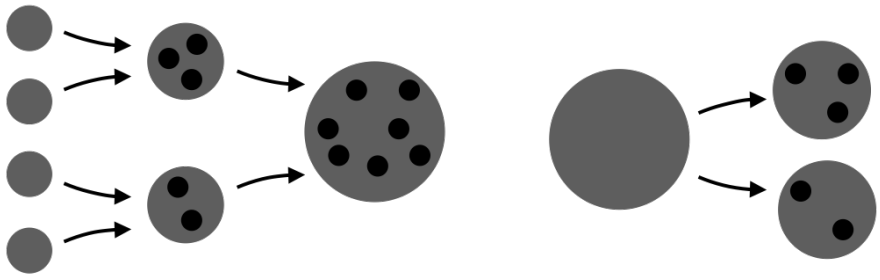
$$\begin{aligned} t &\sim 2 \times 10^{12} s \left(\frac{mc^2}{2\,\text{keV}} \right)^{-2} \\ l_h &\sim ct_h \sim 60\,\text{Mpc} \left(\frac{mc^2}{3\,\text{keV}} \right)^{-1} \end{aligned} \tag{476}$$

so hot dark matter erases all structures below l_h due to free-streaming.

Cold dark matter:
Cold dark matter is already non-relativistic at freeze-out, so structures can grow. It still has some free-streaming scale, but it is much smaller.



Hot dark matter would not be captured by small potential wells, so it needs large potential wells to form structures. This leads to top-down structure formation, where large large structures form first and fragment into smaller structures. Cold dark matter can form small halos that merge into larger ones in bottom-up formation.



Bottom-up theory: smaller structures merge to form larger structures with subhalos

Top-down theory: larger structures fragment into smaller structures

2.B Thermal history of the Universe and other particles

- $T \sim 10^{19}$ GeV, $t \sim 10^{-43}$ s:
  quantum gravity regime
- $T \sim 10^{16}$ GeV, $t \sim 10^{-38}$ s:
  GUT phase transition: strong and electroweak interactions are indistinguishable at earlier times

- $T \sim 10^{12}\,\text{GeV}, t \sim 10^{-30}\,\text{s}$:
  Peccei-Quinn phase transition, if PQ mechanism is the correct explanation for the strong CP problem
- $T \sim 10\text{s} - 100\text{s}\,\text{GeV}, t \sim 10^{-8}\,\text{s}$:
  WIMPs freeze out
- $T \sim 100 - 300\text{s}\,\text{MeV}, t \sim 10^{-5}\,\text{s}$:
  quark-hadron phase transition: quarks and gluons first become bound into neurons and protons
- $T \sim 0.1\,\text{MeV} - 10\,\text{MeV}, t \sim \text{seconds} - \text{minutes}$:
  Big Bang nucleosynthesis (BBN): neutrons and protons first combine to form D, ^{4}He, ^{3}He, and ^{7}Li nuclei
- $T \sim$ keV, $t \sim 1$ day:
  photons fall out of equilibrium, and the number density of photons is conserved
- $T \sim 3\,\text{eV}, t \sim 10^{4-5}$ yrs:
  matter-radiation equality: energy density is dominated by photons at earlier times
- $T \sim \text{eV}, t \sim 400,000$ yrs: electrons and protons combine to form hydrogen
- $T \sim 10^{-3}\,\text{eV}, t \sim 10^9$ yrs:
  first stars and galaxies form
- $T \sim 10^{-4}\,\text{eV}, t \sim 10^{10}$ yrs:
  today

3 Big Bang nucleosynthesis

Once protons and neutrons become available, they can fuse into elements. This allows detailed predictions about the abundance of the first stars.

Proton/neutron reactions:
After n, p production from the gluon-gluon plasma:

$$\begin{aligned} n + \nu_e &\leftrightharpoons p + e^- \\ n + e^+ &\leftrightharpoons p + \bar{\nu}_e \end{aligned} \tag{477}$$

with weak interactions mediated by neutrinos and

$$n_{\text{eq}} = g \left(\frac{mkT}{2\pi\hbar} \right)^{3/2} e^{-\frac{mc^2}{kT}} . \tag{478}$$

Protons and neutrons have $g_n = g_p = 2$, so

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p} \right)^{3/2} e^{-\frac{(m_m - m_p)c^2}{kT}} . \tag{479}$$

So before freeze-out of the above reactions, we have:

$$\frac{n_n}{n_p} \approx e^{-\frac{1.29\,\mathrm{MeV}}{kT}} \,. \tag{480}$$

Weak interactions stop once neutrinos freeze out, $T \sim 0.8\,\mathrm{MeV}, t \sim 1\,\mathrm{s}$:

$$\frac{n_n}{n_p} \approx e^{-1.29/0.8} \approx 0.2 \,, \tag{481}$$

a 5-to-1 ratio. Now nucleosynthesis starts.

Deuteron fusion:
We have the strong reaction:

$$p + n \leftrightharpoons D + \gamma \tag{482}$$

with the binding energy of deuteron approximately 2.22 MeV. At the time of neutrino freeze-out, the temperature is already smaller than 2.22 MeV, but there are so many more photons than baryons. The high energy tail of photons is still sufficient to destroy deuteron, so we need $k_B T \ll 2.22\,\mathrm{MeV}$ to efficiently form deuteron! (The deuteron fusion bottleneck means that 4_{He} fusion afterwards is quick.)

The time delay needed for the temperature to drop below 2.22 MeV causes neutrons to decay through β-decay before they can fuse to deuteron. Without fusion to deuteron, all neutrons would be gone!

Note: If we assume the deuteron fusion is instantaneous, what helium/baryon mass fraction (Y) would we get?
All neutrons would fuse into 4_{He}:

- think of a group of 12 nucleons: $10p + 2n$ (5:1 ratio, see above)
- all neutrons fuse into 4_{He}, so we get one 4_{He} atom and eight free protons
- $$\Rightarrow Y = \frac{4}{4+8} = \frac{4}{12} \approx 0.33 \tag{483}$$
  but we observe 0.24, which is lower due to β-decay.

We now do the precise calculation:

- $p + n \leftrightharpoons D + \gamma$ never freezes out. It stops once all neutrons are used up. We can use the Saha equation to find the abundance:
  $$\frac{n_D}{n_p n_n} = \frac{g_D}{g_p g_n} \left(\frac{m_D}{m_p m_n}\right)^{3/2} \left(\frac{kT}{2\pi\hbar^2}\right)^{-3/2} e^{\frac{2.22\,\mathrm{MeV}}{kT}} \tag{484}$$
  where $g_p = g_n = 2$ (2 spin configurations) and $g_D = 3$ (3 spin configurations: $\uparrow\uparrow, \downarrow\downarrow, \uparrow\downarrow$). We also know $m_p = m_n = m_D/2$, so:
  $$\frac{n_D}{n_p n_n} = 6 \left(\frac{m_n kT}{\pi\hbar^2}\right)^{-3/2} e^{\frac{2.22\,\mathrm{MeV}}{kT}} \tag{485}$$

- Protons always outnumber neutrons, so define the time of deuteron fusion is the time when half the neutrons have fused into deuteron: $n_D = n_n$. Then from the Saha equation:

$$\frac{n_D}{n_n} = 1 = 6n_p \left(\frac{m_n kT}{\pi\hbar^2}\right)^{-3/2} e^{\frac{2.22\,\text{MeV}}{kT}} \tag{486}$$

We now want to know when this happens. We can relate n_p to the temperature to calculate the corresponding temperature and time. We can relate n_p to n_b, which we can relate to n_γ through the fixed baryon-to-photon ratio η.

$$n_b = n_p + n_n = \frac{6}{5} n_p \tag{487}$$

since $n_n = 0.2 n_p = \frac{1}{5} n_p$. Then, for a black-body,

$$\frac{n_p}{n_b} = \frac{5}{6} \Rightarrow n_p = \frac{5}{6}\eta n_\gamma = \frac{5}{6}\eta \left(0.24 \left(\frac{kT}{\hbar c}\right)^3\right) \tag{488}$$

so we get:

$$\begin{aligned}
1 &= 6\frac{5}{6}\eta \left(\frac{kT}{\hbar c}\right)^3 \cdot 0.24 \left(\frac{m_n kT}{\pi\hbar^2}\right)^{-3/2} e^{\frac{2.22\,\text{MeV}}{kT}} \\
&= \frac{0.24 \cdot 5}{1.25}\eta \left(\frac{(kT)^2}{)\hbar c)^2}\frac{\pi\hbar^2}{m_n kT}\right)^{3/2} e^{\frac{2.22\,\text{MeV}}{kT}} \\
&= \eta \underbrace{\pi^{3/2} \cdot 1.25}\left(\frac{kT}{m_n c^2}\right)^{3/2} e^{\frac{2.22\,\text{MeV}}{kT}} \\
&\quad 5.5 \cdot 1.25 = 6.9 \\
&\approx 6.9\eta \left(\frac{kT}{m_n c^2}\right)^{3/2} e^{\frac{2.22\,\text{MeV}}{kT}} \\
&\quad (\text{fiducial } \eta \sim 5 \times 10^{-10}) \\
&\approx 3.4 \times 10^{-9} \left(\frac{kT}{m_n c^2}\right)^{3/2} e^{\frac{2.22\,\text{MeV}}{kT}}
\end{aligned} \tag{489}$$

So we get

$$T \approx 8 \times 10^8\,\text{K}, t \approx 200\,\text{s} \tag{490}$$

for the time of deuteron fusion!

- We lose neutrons through β-decay with a half-life $t_{1/2} = 890\,\text{s}$. After 200 s,

$$\frac{n_n}{n_p} \approx 0.15 < 0.2 \tag{491}$$

so we get a helium-to-baryon ratio:

$$Y = \frac{4n_{\text{He}}}{4n_{\text{He}} + n_{\text{H}}} = \frac{2n_n}{2n_n + (n_p - n_n)} = \frac{2n_n}{n_p + n_n} \tag{492}$$

since $n_{\mathrm{He}} = n_n/2$ (every 4_{He} nucleus has $2n$) and $n_{\mathrm{H}} = n_p - n_n$ (since 4_{He} nucleus has equal number of protons and neutrons), leaving us with

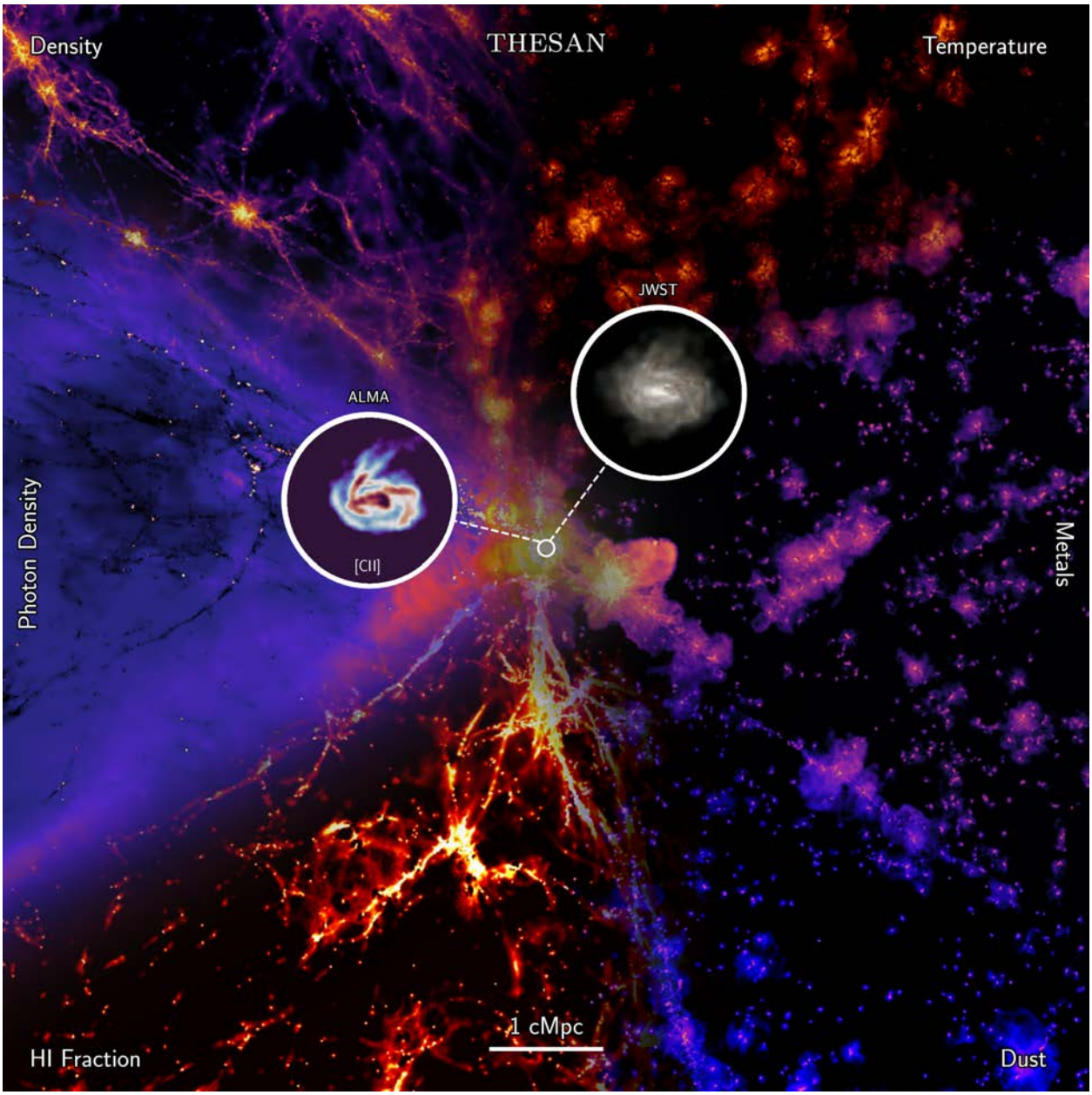
$$Y = \frac{2(n_n/n_p)}{1 + (n_n/n_p)} \approx 0.25 \tag{493}$$

Notes:

- Large Ω_b leads to larger η, so deuteron can form earlier and there is less neutron decay. This leads to a larger n_n/n_p, so Y increases with Ω_b.
- Measurements of 4_{He} and D allow us to determine η and Ω_b. Deuteron abundance is a sensitive measure for Ω_b, and can be found with the Lyman-α forest relating line strength of H and D.

Part IV

Selected Topics



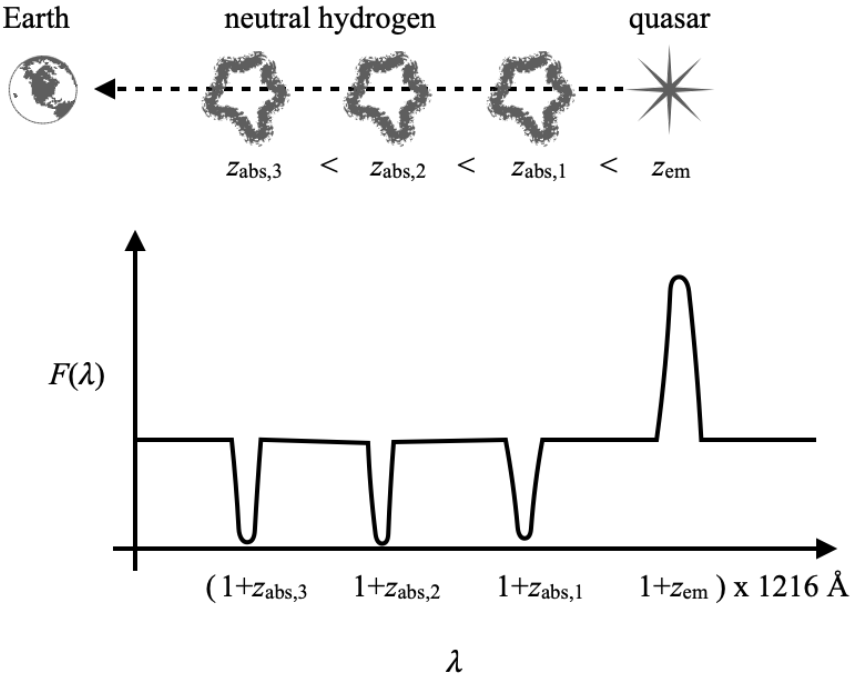
Courtesy of THESAN Collaboration. Used with permission.

So far, we have gone through the basic concepts of the early universe, galaxies, and structure formation. We have built the basis to discuss some more advanced topics.

1 The Lyman-α forest

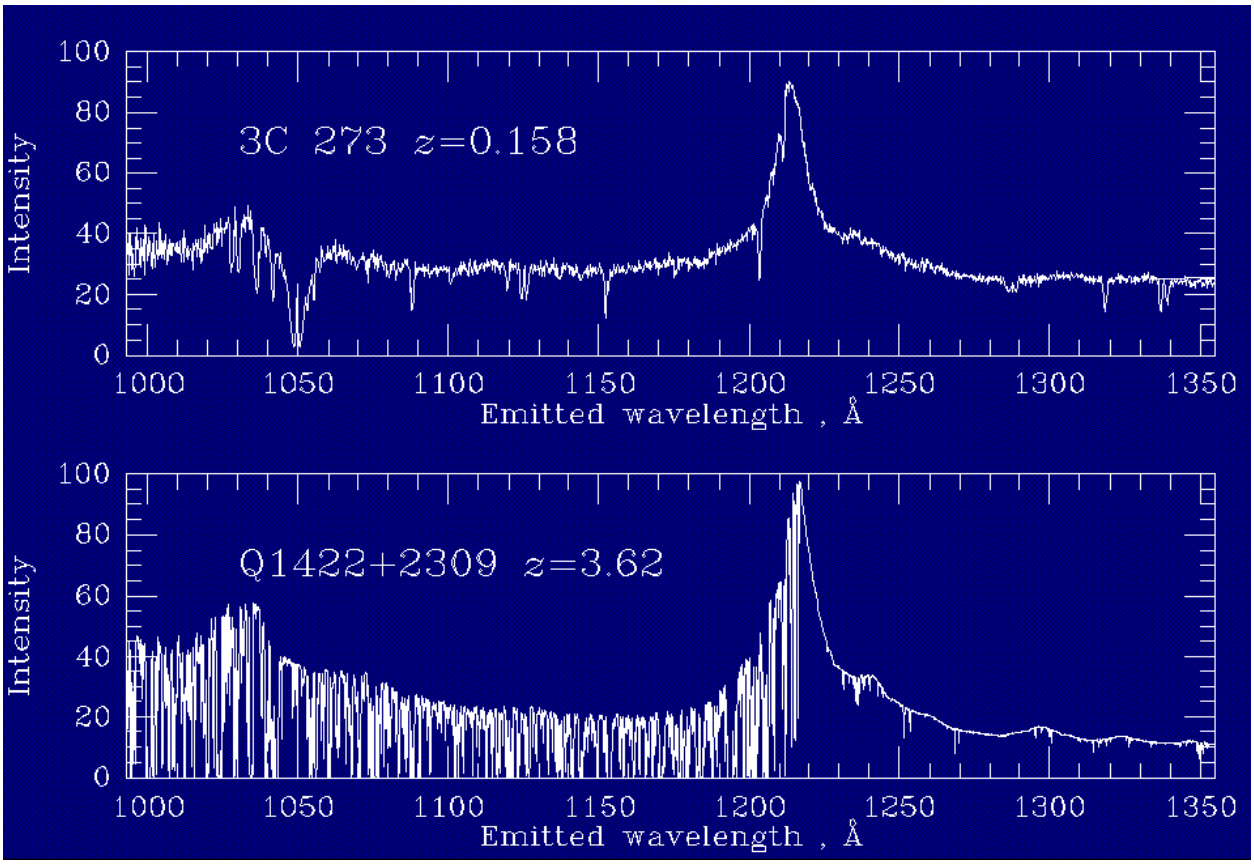
1.A Basics

The Lyman-α forest is the absorption spectrum of quasars. Quasars are very bright from accretion onto supermassive black holes so can be observed out to very high redshifts. They can therefore probe gas between the quasar and us along the line of sight through their absorption lines.



The quasar emits at 1216 Å from the hydrogen $n = 2$ to $n = 1$ transition. This emission line is redshifted as it travels through space. The light also passes through neutral hydrogen clouds, which absorb at 1216 Å, and these absorption lines are also redshifted as the light continues to travel through space. By the time the light reaches Earth, there is a series of absorption lines redshifted from 1216 Å, so the absorption lines are observed at different wavelengths.

By observing quasar spectra passing through the intergalactic medium (IGM), we can use the Lyman-α forest to probe the density, ionization, temperature, chemistry, and structure of the IGM. Below are a few examples of quasar spectra. At higher redshift, there are more opportunities for the light to pass through neutral hydrogen clouds.

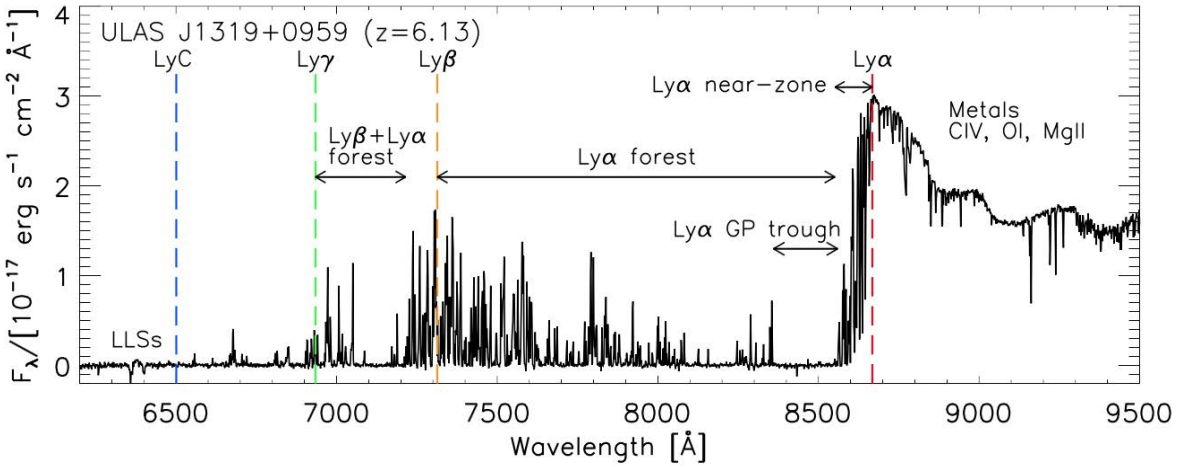


Notes:

- Identify lines in general through doublets: μ_{0II} ($\lambda = 2795$Å, 2802Å...)
- Lyman-α forest is only visible if it extends to shorter wavelengths than the observed Lyα emission line at $(1 + z_{\rm em})1216$ Å. Photons emitted with $1216\text{Å}(1 + z_{\rm em} < \lambda) < 1216\text{Å}$ will have at some point along the line of sight the right rest frame wavelength (1216Å) to be absorbed.
- There are three cases for HI along the line of sight
  - column density $N_{\rm H} \lesssim 10^{17}$ cm^{-2} gives narrow lines, i.e. the forest
  - column density $N_{\rm H} \gtrsim 10^{17}$ cm^{-2} are Lyman-limit systems, i.e. photons with $\lambda \lesssim 912\text{Å} = 13.6\text{eV}$ in the rest frame are completely absorbed as the light moves through the cloud
  - column density $N_{\rm H} \gtrsim 10^{32}$ cm^{-2} are damped Ly-α systems, i.e. absorption lines become very broad.

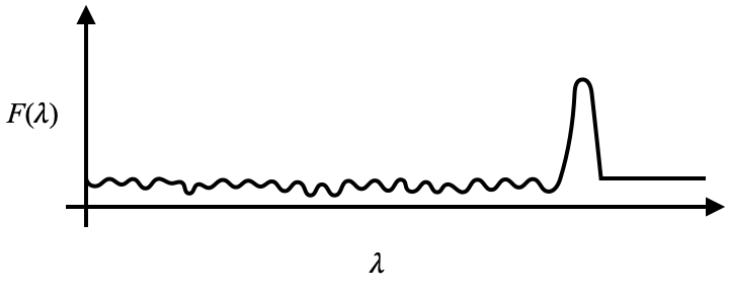
The Gunn-Peterson Test:
We can use quasar absorption spectra as a hint about reionization and determining how baryons are distributed in the universe and in which state. The key idea is that neutral hydrogen along the line of sight leads to absorption, so if there is a significant amount of

neutral hydrogen then a quasar spectrum should be totally absorbed. When the redshift is high enough so that the hydrogen is almost all neutral, the high-wavelength region of the spectrum will be almost totally absorbed. This is called the *Gunn-Peterson trough*.

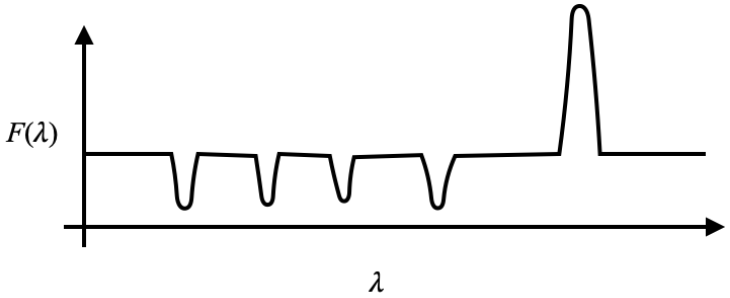


What is observed?

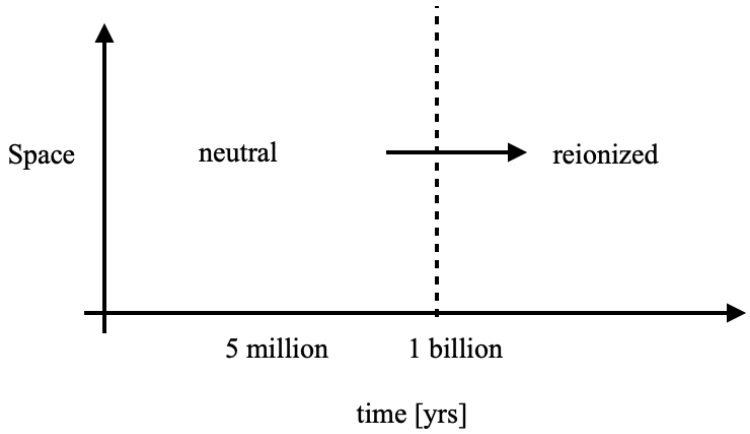
For $z \gtrsim 6$ we see suppression due to lots of neutral hydrogen along the line of sight.



For $z \lesssim 6$, we see little suppression due to little neutral hydrogen along the line of sight.



This implies that hydrogen above $z \approx 6$ is mostly neutral and mostly ionized below $z \approx 6$.

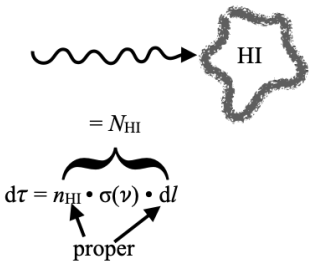


1.B A quantitative approach to Lyman-α

We now look in more detail at the absorption process:

$$F(\lambda_{\rm obs}) = F(\lambda_{\rm em}(1+z))e^{-\tau} \tag{494}$$

where τ is the optical depth, which we need to calculate.



Photons en route through a neutral HI cloud hit atoms in the ground state which then transition to an excited state ($n = 1 \rightarrow 2,\ n = 1 \rightarrow 2, ...n = 1 \rightarrow$ ionized). Each transition has a frequency dependent cross-section. The atoms then settle back to the ground state and a photon is emitted in some other direction within the solid angle 4π.

From atomic physics, we know that

$$\sigma(\nu) = \frac{\pi e^2}{m_{\rm e}c} f\phi(\nu - \nu_0) \tag{495}$$

where f is the oscillator strength (i.e. the probability for absorption) and ϕ is the Voigt profile with $\int \phi {\rm d}\nu = 1$. For the Lyman-α transition, $\sigma(\nu) = 10^{-2}\ {\rm cm}^2\phi(\nu - \nu_0)$ with units ${\rm cm}^2{\rm Hz}^{-1}$.

The proper length ${\rm d}l$ can be related to redshift (taking only the magnitude and ignoring signs):

$${\rm d}l = c{\rm d}t = c\frac{{\rm d}a}{\dot{a}} = \frac{c}{a}\frac{{\rm d}a}{\dot{a}/a} = \frac{c}{a}\frac{{\rm d}a}{H} = (1+z)c\frac{{\rm d}a}{H} \tag{496}$$

and since $a = 1/(1+z)$ then $\frac{{\rm d}a}{{\rm d}z} = 1/(1+z)^2$ so

$${\rm d}l = \frac{1}{1+z}c\frac{{\rm d}z}{H(z)} = \frac{c{\rm d}z}{(1+z)H(z)} \tag{497}$$

giving

$${\rm d}l = \frac{c{\rm d}z}{(1+z)H(z)} = c\left(H_0(1+z)\sqrt{\Omega_{M,0}(1+z)^3 + \Omega_\Lambda}\right)^{-1} {\rm d}z\ . \tag{498}$$

For the matter dominated regime ($z \gtrsim 1$), we get

$$\mathrm{d}l = \frac{c\mathrm{d}z}{H_0\sqrt{\Omega_{M,0}}(1+z)^{\frac{5}{2}}} \ . \tag{499}$$

Then the optical depth is

$$\tau_\nu = \sigma_0 \frac{c}{H_0\Omega_{M,0}^{\frac{1}{2}}} \int_0^z \mathrm{d}z' \frac{n_{\mathrm{HI}}(z')}{(1+z')^{\frac{5}{2}}} \phi(\nu(1+z') - \nu_0) \tag{500}$$

and we can assume that ϕ, which has a Gaussian shape, is very narrow so is approximately a delta function

$$\begin{aligned} &\approx \sigma_0 \frac{c}{H_0\Omega_{M,0}^{\frac{1}{2}}} \int_0^z \mathrm{d}z' \frac{n_{\mathrm{HI}}(z')}{(1+z')^{\frac{5}{2}}} \delta(\nu(1+z') - \nu_0) \\ &= \sigma_0 \frac{n_{\mathrm{HI}}(z)}{H_0\Omega_{M,0}^{\frac{1}{2}}} \frac{c}{\nu_0} \frac{1}{(1+z)^{\frac{3}{2}}} \end{aligned} \tag{501}$$

and using $\lambda_0 = \frac{c}{\nu_0}$, we get

$$\tau_\nu(z) = \sigma_0 \frac{n_{\mathrm{HI}}(z)\lambda_0}{H_0\Omega_{M,0}^{\frac{1}{2}}(1+z)^{\frac{3}{2}}} \tag{502}$$

where $\sigma_0 = 10^{-2}\,\mathrm{cm}^2$ and z is the redshift when $\lambda(1+z) = \lambda_0$, i.e. when absorption happens. This gives the optical depth for one frequency, so is necessary to evaluate at many frequencies to get the optical depth for different parts of a spectrum.
As an example, we can evaluate $\tau_\nu(z=3)$ assuming hydrogen is uniformly distributed and neutral. Using

$$\begin{aligned} H_0 &= 70\,\mathrm{km/s/Mpc} = 2.3 \times 10^{-18}\,\mathrm{s}^{-1} \\ n_{\mathrm{HI}} &= \frac{\rho_{\mathrm{crit}}\Omega_{b,0}}{m_{\mathrm{H}}}(1+z)^3 \sim 10^{-5}\,\mathrm{cm}^{-3}\left(\frac{1+z}{4}\right)^3 \approx 10^{-5}\,\mathrm{cm}^{-3} \qquad \text{at } z=3 \\ \lambda_0 &= 1216\,\text{Å} \\ \sigma_0 &= 10^{-2}\,\mathrm{cm}^2 \end{aligned} \tag{503}$$

we get $\tau_{\mathrm{Ly}\alpha} \sim 10^5$, but we observe that $\tau \sim 1$.

There are a couple possible solutions to this discrepancy. It could be that gas isn't in intergalactic space. However, we know that this is not the case since we have observed it. The other option is that the gas isn't neutral. To bring $\tau_{\mathrm{Ly}\alpha}(z=3)$ down to ~ 1, we need to have the neutral hydrogen fraction $X_{\mathrm{HI}} \equiv \frac{n_{\mathrm{HI}}}{n_{\mathrm{H}}} \sim 10^{-5}$.

Ionization:
How does the hydrogen get ionized?

Ionization can occur in hot temperatures. We can estimate the temperature from absorption line widths:

$$\begin{aligned} &\Delta v \sim 20\,\mathrm{km/s} \\ &\frac{1}{2}mv^2 \sim kT \\ \Rightarrow &20\,\mathrm{km/s} \sim \sqrt{\frac{2kT}{m}} \\ \Rightarrow &T \sim 30,000\,\mathrm{K} \sim 3eV \end{aligned} \tag{504}$$

which is not enough to ionize hydrogen.

Another option is photoionization. Integrated light from galaxies and quasars emits $\Gamma \sim 10^{-12}$ ionizing photons per second. The ionization rate is then Γn_{HI}, and the ionization timescale is

$$\frac{n_{\mathrm{HI}}}{n_{\mathrm{HI}}\Gamma} \sim 10^{12}\,\mathrm{s} \sim 30,000\,\mathrm{yr} \tag{505}$$

We can compare this with the recombination rate $Rn_{\mathrm{HII}}n_{\mathrm{e}}$ where $R = 4.3 \times 10^{-13} \left(\frac{T}{10^4\,\mathrm{K}}\right)^{-0.7}$ which gives the recombination timescale

$$\frac{n_{\mathrm{HII}}}{Rn_{\mathrm{e}}n_{\mathrm{HII}}} \sim 2 \times 10^{17}\,\mathrm{s} \sim 3 \times 10^6\,\mathrm{yr} \tag{506}$$

so recombination is much slower than ionization and photoionization is plausible.

To establish the predicted ionization fraction, we find equilibrium by setting the ionization and recombination timescales equal:

$$Rn_{\mathrm{HII}}n_{\mathrm{e}} = \Gamma n_{\mathrm{HI}} \tag{507}$$

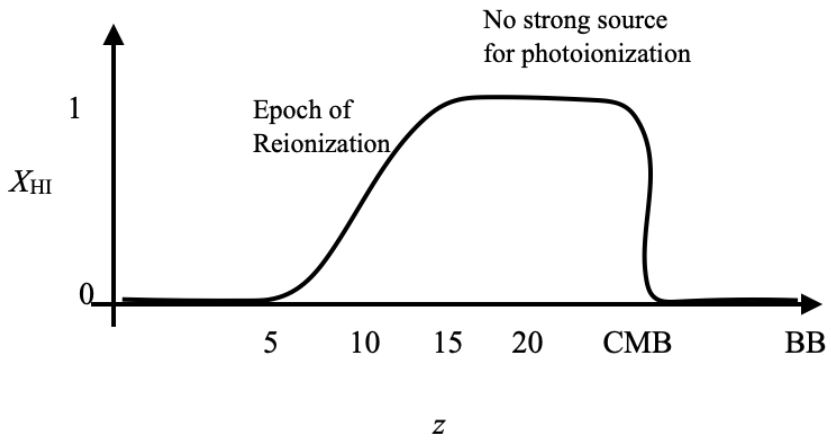
and assume that $n_{\mathrm{HII}} \sim n_{\mathrm{e}} \sim n_{\mathrm{H}}$ and $n_{\mathrm{HI}} \ll n_{\mathrm{HII}}$ so

$$Rn_{\mathrm{H}}^2 = X_{\mathrm{HI}}n_{\mathrm{H}}\Gamma \tag{508}$$

which gives a neutral fraction of

$$X_{\mathrm{HI}} \approx \frac{Rn_{\mathrm{H}}}{\Gamma} \sim 5 \times 10^{-6}\,. \tag{509}$$

So at the present day, the ionization fraction is very small. However, it took a while between recombination and the present day for ionizing sources to form and begin emitting radiation to ionize the neutral gas. Once they formed, the gas was ionized over a period of time. Before recombination, the gas was mostly ionized in the hot universe. Exactly how and when reionization occurred is an active area of research that is being probed by both telescopes like HERA and simulations like THESAN.


No strong source
for photoionization
1
Epoch of
Reionization
X_{HI}
0
5
10
15
20
CMB
BB
z

8.902 Astrophysics II

Fall 2023