Dark Matter Halos

Dark matter halos form through gravitational instability.

*Density perturbations grow linearly until they reach a critical density, after which they turn around from the expansion of the Universe and collapse to form virialized dark matter halos.*

These halos continue to grow in mass (and size), either by accreting material from their neighborhood or by merging with other halos.

- The mass function of dark matter halos (i.e. the number density of halos as a function of halo mass), its dependence on cosmology, and its evolution with redshift.
- The mass distribution of the progenitors of individual halos.
- The merger rate of dark matter halos.
Dark Matter Halos

In the peak formalism, it is *assumed* that the material which will collapse to form non-linear objects (i.e. dark matter halos) of given mass can be identified in the *initial* density field by first smoothing it with a filter of the appropriate scale and then locating all peaks above some threshold.

The properties of peaks in a Gaussian random field can be analysed in a mathematically rigorous way, but the formalism nevertheless has significant limitations:

- it cannot be used to obtain the mass function of nonlinear objects (dark halos)
- it does not provide a model for how dark matter halos grow with time

Alternative formalism: Press-Schechter 1974
- mass function, the mass distribution of progenitors, merger rate and clustering properties
Density Peaks

In the linear regime the overdensity field grows as $\delta(x, t) \propto D(t)$, with $D(t)$ the linear growth rate:

for a given cosmology [which governs $D(t)$] we can determine the properties of the density field at any given time from those at some early time $t_i$.

An object of mass $M$ forms from an overdense region with volume $V = M/\rho$ in the initial density field.

It is tempting to assume that such a region will correspond to a peak in the density field after smoothing with window of characteristic scale $R \propto M^{1/3}$.

This suggests that a study of the density peaks of the smoothed density field can shed some light on the number density and spatial distribution of collapsed dark matter halos.
Shortcomings of the Peak Formalism

It is tempting to interpret the number density of peaks in terms of a number density of collapsed objects of mass \( M \propto \rho R^3 \). However, there is a serious problem with this identification, because a mass element which is associated with a peak of \( \delta_1(x) = \delta_3(x; R_1) \) can also be associated with a peak of \( \delta_2(x) = \delta_3(x; R_2) \), where \( R_2 > R_1 \).

Should such a mass element be considered part of an object of mass \( M_1 \) or of mass \( M_2 \)?

If \( \delta_2 < \delta_1 \) the mass element can be considered part of both. The overdensity will first reach (at time \( t_1 \)) the critical value for collapse on scale \( R_1 \), and then (at time \( t_2 > t_1 \)) on scale \( R_2 \). \( M_1 \) is the mass of a collapsed object at \( t_1 \) which merges to form a bigger object of mass \( M_2 \) at \( t_2 \).

In the opposite case, where \( \delta_2 > \delta_1 \), the mass element apparently can never be part of a collapsed object of mass \( M_1 \) but rather must be incorporated directly into a larger system of mass \( M_2 \). Such peaks of the field \( \delta_1 \) should therefore be excluded when calculating the number density of \( M_1 \) objects.

This difficulty is known as the “cloud-in-cloud” problem.
Press-Schechter Formalism

Consider the overdensity field $\delta(x,t)$, which in the linear regime evolves as $\delta(x,t) = \delta_0(x)D(t)$, where $\delta_0(x)$ is the overdensity field linearly extrapolated to the present time and $D(t)$ is the linear growth rate normalized to unity at the present.

According to the spherical collapse model described, regions with $\delta(x,t) > \delta_c \simeq 1.69$, or equivalently, with:

$$\delta_0(x) > \delta_c/D(t) \equiv \delta_c(t),$$

will have collapsed to form virialized objects.

In order to assign masses to these collapsed regions, Press & Schechter (1974) considered the smoothed density field

$$\delta_s(x; R) \equiv \int \delta_0(x') W(x + x'; R) \, d^3 x'$$

where $W(x; R)$ is a window function of characteristic radius $R$ corresponding to a mass $M = \gamma_f \rho R^3$, with $\gamma_f$ a filter-dependent shape parameter (i.e. $\gamma_f = 4\pi/3$ for a spherical top-hat filter).

The ansatz of the Press-Schechter (PS) formalism is that the probability that $\delta_s > \delta_c(t)$ is the same as the fraction of mass elements that at time $t$ are contained in halos with mass greater than $M$.

The probability that an overdense region of radius $R$ has collapsed gives the fraction of mass in objects above that mass.
Merger Trees

The PS mass function can be used to describe the statistical properties of the merger histories of dark matter halos.

If we start from an early time, the mass which ends up in a halo at the present time $t_0$ (the trunk of the tree) is distributed over many small branches; pairs of small branches merge into bigger ones as time goes on.

Fig. 7.4. Illustration of a merger tree, depicting the growth of a dark matter halo as the result of a series of mergers. Time increases from top to bottom, and the width of the tree branches represent the masses of the individual halos. A horizontal slice through the tree, such as that at $t = t_f$ gives the distribution of the masses of the progenitor halos at a given time. [Adapted from Lacey & Cole (1993)]
**Merger Trees**

In order to construct a merger tree of a halo, it is most convenient to start from the trunk and work upwards.

Consider a halo with mass $M$ at a time $t$ (which can be the present time, or any other time). At a slightly earlier time $t - \Delta t$, the progenitor distribution, in a statistical sense, is given by Eq. (7.81).

The construction of a set of progenitor masses for a given parent halo mass needs to obey two requirements:

1) the *number* distribution of progenitor masses needs to follow Eq. (7.81).

2) mass needs to be conserved, so that in each individual realization the sum of the progenitor masses is equal to the mass of the parent halo.

In principle, this requirement for mass conservation implies that the probability for the mass of the $n^{th}$ progenitor needs to be conditional on the masses of the $n - 1$ progenitor halos already drawn.

Unfortunately, these conditional probability functions are not derivable from the EPS formalism, so additional assumptions have to be made.
Main Progenitor

The main trunk is defined by following the branching of a merger tree back in time, and selecting at each branching point the most massive (main) progenitor.

The left panel shows examples of main-progenitor histories for a halo of $M_0 = 5 \times 10^{11} h^{-1} M_\odot$ - large amount of scatter.

The right panel shows the average main progenitor histories for five values of $M_0$, as indicated.

More halos massive assemble later, a characteristic of hierarchical structure formation.
Gravity is easy but,

The astronomical objects we observe are made of baryons and electrons, so gas-dynamical and radiative processes must also be taken into account.

Since baryons and dark matter are expected to be well mixed initially, the density perturbation fields of the baryons, $\delta_b$, and dark matter, $\delta_{dm}$, are expected to be equal in the linear regime.
8.1.1 Basic Fluid Dynamics and Radiative Processes

The gas component can normally be approximated as an ideal fluid, which means that we can **neglect heat conduction and viscous stress** in the fluid equations.

\[
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{8.1}
\]

\[
\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = - \left( \nabla \Phi + \frac{\nabla P}{\rho} \right), \tag{8.2}
\]

\[
\frac{\partial}{\partial t} \left[ \rho \left( \frac{v^2}{2} + \varepsilon \right) \right] + \nabla \cdot \left[ \rho \left( \frac{v^2}{2} + \frac{P}{\rho} + \varepsilon \right) \mathbf{v} \right] - \rho \mathbf{v} \cdot \nabla \Phi = \mathcal{H} - \mathcal{C}. \tag{8.3}
\]

Here \( \rho, \mathbf{v}, P, E \) are the density, velocity, pressure and specific internal energy of the fluid, respectively, and \( H \) and \( C \) are the heating and cooling rates per unit volume.
Continuity

- Mass in a region **increases when mass flows inward and decreases when it flows outward** (density*velocity=Flux);
- Mass in a region **increases when new mass is created and decreases when mass is destroyed** (delta density/delta t);

Apart from these two processes, there is no other way for the amount of q in a region to change.

Continuity: when a fluid is in motion, it must **move in such a way that mass is conserved**.
Euler

\[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \]  
\[ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\left( \nabla \Phi + \frac{\nabla P}{\rho} \right), \]  
\[ \frac{\partial}{\partial t} \left[ \rho \left( \frac{v^2}{2} + \varepsilon \right) \right] + \nabla \cdot \left[ \rho \left( \frac{v^2}{2} + \frac{P}{\rho} + \varepsilon \right) \mathbf{v} \right] - \rho \mathbf{v} \cdot \nabla \Phi = \mathcal{H} - \mathcal{E}. \]

The Euler's equation for steady flow of an ideal fluid along a streamline comes from Newton's second law.

- The fluid is non-viscous (i.e., the frictional losses are zero).
- The fluid is homogeneous and incompressible (i.e., mass density of the fluid is constant).
- The flow is continuous, steady and along the streamline.
- The velocity of the flow is uniform over the section.
- No energy or force (except gravity and pressure forces) is involved in the flow.
The Euler's equation for steady flow of an ideal fluid along a streamline comes from Newton's second law.

- First term, \( \frac{dV}{dt} \), called local or temporal acceleration results from velocity changes with respect to time at a given point. Local acceleration happens when the flow is unsteady.

- Second term, \( V \cdot \text{div} V \), is called convective acceleration because it is associated with spatial gradients of velocity in the flow field. Convective acceleration happens when the flow is non-uniform, that is, if the velocity changes along a streamline.
Euler

\[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (8.1) \]

Gravity

\[ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\left( \nabla \Phi + \frac{\nabla P}{\rho} \right), \quad (8.2) \]

Pressure

\[ \frac{\partial}{\partial t} \left[ \rho \left( \frac{v^2}{2} + \varepsilon \right) \right] + \nabla \cdot \left[ \rho \left( \frac{v^2}{2} + \frac{P}{\rho} + \varepsilon \right) \mathbf{v} \right] - \rho \mathbf{v} \cdot \nabla \Phi = \mathcal{H} - \mathcal{E}. \quad (8.3) \]

The gravitational potential \( \Phi \) satisfies the Poisson equation

\[ \nabla^2 \Phi = 4\pi G \rho_{\text{tot}}, \quad (8.5) \]

If you accelerate a fluid element you have an additional pressure on the right end side of the element:

\[ = P \cdot dA - (P + dP) \cdot dA \]

\[ = -dP \cdot dA \]

The gravitational flux through any closed surface is proportional to the enclosed mass.
Energy - Bernoulli’s equation

\[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \]  
\[ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = - \left( \nabla \Phi + \frac{\nabla P}{\rho} \right), \]  
\[ \frac{\partial}{\partial t} \left[ \rho \left( \frac{v^2}{2} + \mathcal{E} \right) \right] + \nabla \cdot \left[ \rho \left( \frac{v^2}{2} \frac{P}{\rho} + \mathcal{E} \right) \mathbf{v} \right] - \rho \mathbf{v} \cdot \nabla \Phi = \mathcal{H} - \mathcal{C}. \]  
\[ \frac{P}{\gamma - 1} \left( \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \ln \left( \frac{P}{\rho^\gamma} \right) = \mathcal{H} - \mathcal{C}, \]

For an ideal fluid where \( P = \text{density}(\gamma-1) \times \mathcal{E} \)
8.3 The Formation of Hot Gaseous Halos

The baryonic component is expected to cluster in the same way as collisionless dark matter particles during the early evolutionary stages, when the perturbations on galaxy scales are still in the linear regime.

However, once overdense regions become non-linear and start to collapse, the baryonic gas associated with the overdensity can be shocked and heated, while collisionless dark matter particles cannot. Consequently, in a collapsed object the distribution of the gas component may deviate from that of the dark matter - formation of gaseous halos.
8.3.1 Accretion Shocks

Shock waves can be produced:

- When an obstructive body moves through a gaseous medium with supersonic speed. This is what happens, for example, when supernova explosions drive gas clouds (shells) into the interstellar medium.

- When cold gas is accreted into a gravitational potential well, such as a dark matter halo.

Hot gas temperature \( \sim T_{\text{vir}} \)
To exemplify the essence of how the state of gas is affected by a shock, let us idealize the problem by a planar shock propagating at a constant speed $v_{sh}$ through a uniform medium (Fig. 8.3).

The motion is one-dimensional, and we assume $v_{sh}$ to be along the $-x$-axis.

For an observer moving with the shock wave, the fluid is static. The properties in the upstream (pre-shock fluid) and downstream (post-shock fluid) are therefore related by the steady-state fluid equations.

Fig. 8.3. An illustration of different regions of a planar shock. The supersonic flow in the region $x < x_1$ is shocked between $x_1$ and $x_2$ (i.e. the shock has a finite, though small, width), after which it becomes a hot subsonic flow. In between $x_2$ and $x_3$ the gas is out of thermal equilibrium resulting in net cooling ($\mathcal{L} > 0$). At $x > x_3$ the gas has cooled, reached a new thermal equilibrium, and continues to flow subsonically. The arrows indicate the direction (but not the speed) of the flow.
8.1.2, 8.1.3, 8.1.4 Cooling and Heating

In order to study the evolution of the baryonic component using the ideal-fluid approximation, one has to deal with processes that can heat or cool the baryonic gas. Basic processes are:

**Compton Cooling:** compton scattering of photon that interact with electrons in a gas

**Radiative Cooling:** two body interactions: bremsstrahlung, collisional ionization, recombination, collisional excitation

**Photoionization Heating:** atoms ionized by absorbing photons
Radiative Cooling

Fig. 8.1 shows the cooling function as function of $T$ for gases with three different metallicities + the primordial case. In all cases, the cooling rate is dominated by bremsstrahlung at the high-temperature end, where $\Lambda \propto T^{1/2}$.

Fig. 8.1. Cooling functions for primordial ($Z = 0$) gas (assuming $n_{\text{He}}/n_{\text{H}} = 1/12$), and for gases with metallicities $Z/Z_\odot = 0.01$, 0.1 and 1.0, as indicated. [Based on data published in Sutherland & Dopita (1993)]
Radiative Cooling

For a primordial gas with $T < 10^{5.5}$ K, a large fraction of the electrons are bound to their atoms, and the dominant cooling process is collisional excitation followed by radiative de-excitation; the peaks in the cooling function at $\sim 15000$ K and $\sim 10^5$ K are due to collisionally excited electronic levels of H$^0$ and He$^+$, respectively.
For an enriched gas, there is an even stronger peak at $T = 10^5 \text{K}$ due to the collisionally excited levels of ions of oxygen, carbon, nitrogen, etc. In an enriched gas, the cooling function is also enhanced at $\sim 10^6 \text{K}$ by other common elements, noticeably neon, iron and silicon.

Fig. 8.1. Cooling functions for primordial ($Z = 0$) gas (assuming $n_{\text{He}}/n_{\text{H}} = 1/12$), and for gases with metallicities $Z/Z_\odot = 0.01$, 0.1 and 1.0, as indicated. [Based on data published in Sutherland & Dopita (1993)]
Radiative Cooling Time Scales for Uniform Clouds

The cooling time for gas at temperature $T$ and density $n$ as a result of radiative cooling can be written as:

$$t_{\text{cool}} \equiv \frac{\rho c}{\mathcal{E}} = \frac{3nk_{B}T}{2n_{H}^{2} \Lambda(T)} \approx 3.3 \times 10^{9} \frac{T_{\odot}}{n_{-3} \Lambda_{-23}(T)} \text{yr},$$  \hspace{1cm} (8.94)

where $E$ is the internal energy per unit mass, $n_{H}$ is the number density of hydrogen atoms [$n_{H} = (12/27)n$ for a fully ionized primordial gas], and we have assumed an ideal gas with adiabatic index $\gamma = 5/3$. Note that we have written the cooling function as $\Lambda = 10^{-23} \Lambda_{-23}\text{ergcm}^{3}\text{s}^{-1}$.

The $10^{11} M_{\odot}$ protogalaxy forming at $z = 3$ has $n_{-3} \approx 5.5$ (Since a newly collapsed object has an overdensity $\delta \sim 200$) and $\Lambda_{-23} \approx 0.5$ assuming primordial gas, which implies a cooling time of $t_{\text{cool}} \sim 7.4 \times 10^{8}\text{yr}$. This is roughly twice as long as its free-fall time,

$$t_{\text{ff}} = \sqrt{\frac{3\pi}{32G\rho}} = \sqrt{\frac{3\pi f_{\text{gas}}}{32Gn\mu m_{p}}} \approx 2.1 \times 10^{9} f_{\text{gas}}^{1/2} n_{-3}^{1/2} \text{yr},$$  \hspace{1cm} (8.95)

or $t_{\text{ff}} \approx 3.5 \times 10^{8}\text{yr}$ for the case we are considering. Note that for a given temperature $t_{\text{cool}} \propto (1 + z)^{-3}$ and $t_{\text{ff}} \propto (1 + z)^{-3/2}$ so that cooling is more effective at higher redshifts.

Note also that, in the absence of a strong UV background, protogalaxies with $10^{4}\text{K} < T_{\text{vir}} < 10^{8}\text{K}$ (where $\Lambda_{-23} \sim 10$) have $t_{\text{cool}}$ much smaller than $t_{\text{ff}}$. Thus, the gas in small halos at high redshifts is expected to cool effectively and on short time-scales.
Radiative Cooling Time Scales for Uniform Clouds

Fig. 8.6 shows the locus of $t_{\text{cool}} = t_{\text{ff}}$ in the $n$–$T$ plane, which separates clouds that can cool effectively ($t_{\text{cool}} \ll t_{\text{ff}}$) from those that cannot. Halos with (primordial) gas masses larger than about $10^{11} \, M_\odot$ cannot cool effectively. This mass increases to $\sim 10^{12} \, M_\odot$ if the gas has Solar metallicity. This is why clusters and groups of galaxies at the present time usually contain large amounts of hot gas.
Fig. 8.7. Same as Fig. 8.1, except that now we also indicate $\Lambda_{\text{crit}}$ for halos in the redshift range $0 \leq z \leq 3$ (shaded area). Halos with $\Lambda > \Lambda_{\text{crit}}$ accrete their gas in the cold mode, while those with $\Lambda < \Lambda_{\text{crit}}$ experience hot mode accretion.
9. Star Formation in Galaxies

Galaxies are mostly observed through stars - any theory of galaxy formation has to address the question of how stars form.

- Baryonic gas in halos can cool within a time that is shorter than the age of the halo.
- Loses pressure support and flows to the center of the potential well - density increases.
- When density of cooling gas > dark matter density (in the central part of the halo) - gas becomes self-gravitating and collapses under its own gravity.

If cooling is efficient - self-gravitating gas is unstable and collapses catastrophically - formation of dense, cold gas clouds within which star formation can occur.
9. Star Formation in Galaxies

Understanding SF is challenging:

- typical mass of gas: $\sim 10^{11} \, M_\odot$, typical density of gas $\sim 10^{-24} \, \text{g cm}^{-3}$
- typical mass of a star $\sim 1 \, M_\odot$, typical density of a star $\sim 1 \, \text{g cm}^{-3}$.

Formation of an individual stars spans 11 orders of magnitude in mass and 24 orders of magnitude in density.

However, for our purpose (galaxy formation and evolution) - we are only interested in properties of SF in a volume comparable to a galaxy.
9.1 Giant Molecular Clouds: the Sites of Star Formation

Star formation is slow in present-day spiral galaxies, such as the Milky Way. The time scale for star formation, the gas consumption time scale, is defined as the ratio between the gas mass available and the rate at which gas turns into stars, \( \tau \equiv \frac{M_{\text{gas}}}{M^*} \).

Normal Spirals:

- Spiral galaxies \( \tau = (1-5) \times 10^9 \text{yr} \), \( \sim 10 \) times longer than the relevant dynamical time scales of the gas.
- For the Milky Way (we have great detail), SF is always associated with giant molecular clouds (GMCs) embedded in the interstellar medium (ISM).
- The GMCs spatially correlate with young star clusters (ages < \( 10^7 \) yr) but not with older star clusters: implies that typical lifetime of a GMC is \( \sim 10^7 \) yr, much shorter than the age of a galaxy.
- Rare to find GMCs not forming stars - star formation in a GMC starts when cloud forms.

Rate of large-scale star formation is essentially proportional to the formation rate of GMCs.
9. Star Formation in Galaxies

Star Bursts:

Star formation time scales $\sim 10^7 - 10^8$ yr, comparable to the dynamical times of the cold gas in these galaxies much shorter than those in normal disk galaxies.

Observations of CO emission associated with large amounts of molecular gas ($10^8$–$10^{10} M_\odot$) confined to small volumes with sizes <2 kpc.

Molecular gas densities are 10-100 times higher than in the inner 1 kpc region of the Milky Way.

Although starbursts have very different star-formation properties, SF also driven by the availability of dense molecular clouds.

It is now generally believed that GMCs are the principal sites for SF, and SFR is determined by its ability to form GMCs. Star formation in galaxies can be divided into: ‘microphysics’ and ‘macrophysics’.

- the formation of individual stars in the dense regions of GMCs
- the formation and structure of GMCs in galaxies
9.1.1 Structure and Dynamical State of GMCs

Molecular fraction increases with gas density: 0 in the diffuse ISM, almost 100% in dense regions, near Galactic center.

The state of a GMC can be described qualitatively with the aid of the virial theorem:

\[
\frac{1}{2} \dot{I} = 2K + W - \Sigma + B. \tag{9.1}
\]

Note that \( I \) is the moment of inertia, \( K \) is the kinetic energy of the system, and \( \Sigma \) is the work done by the external pressure. The potential energy, \( W \), is equal to the gravitational energy of the system and \( B \) is the magnetic energy.

A system whose kinetic energy is much larger (smaller) than the negative value of its potential energy tends to expand (contract). For a static system, \( d^2I/dt^2 = 0 \), and

\[
2K + W + \Sigma = 0, \tag{5.134}
\]

giving a constraint on the global properties of any system in a static state. Finally, for \( \Sigma = 0 \) we have

\[
E = -K = \frac{W}{2}, \tag{5.135}
\]

where \( E = K + W \) is the total energy.
Gravitational Mass Scale

\[
\frac{1}{2} \ddot{I} = 2K + W - \Sigma + \mathcal{B}. 
\] (9.1)

A cloud is said to be virialized if \( \ddot{I} = 0 \), expanding if \( \ddot{I} > 0 \), and contracting if \( \ddot{I} < 0 \).

Quantities on the right-hand side can be used to set relevant scales that describe the dynamic state of a cloud.

1. gravitational mass scale:

2. maximum mass for an isothermal sphere embedded in a pressurized medium in hydrostatic equilibrium (in the absence of magnetic fields and turbulence), the Bonnor-Ebert mass:

3. \( B = |W| \) defines a mass:
Clumps Properties

Properties of Clumps:

Molecular clumps themselves are also highly inhomogeneous, containing dense cores with densities $> 10^{-20}$ g cm$^{-3}$ and temperatures $\sim 10$ K.

**Low mass cores:** appear to be supported by thermal motion, rather than by turbulent motion or magnetic pressure. Mass comparable to Bonnor-Ebert indicating that these cores will contract further in the presence of effective cooling.

**Most massive cores:** have significant levels of internal turbulence and appear to be magnetically critical or supercritical. However, may also collapse once the turbulence energy has been dissipated.

*Cores in molecular clumps are the potential sites within which individual stars form. Indeed, young stars in the Milky Way are typically found to be associated with dense cores embedded in large molecular clouds.*

Turbulence is ubiquitous in the ISM over a large range of scales:

- from that comparable to the host galaxy all the way down to that of massive cores of the molecular clouds.

Based on these observations, and modelling, large-scale, supersonic turbulence is assumed to govern the macrophysics of SF in galaxies.
Formation of Molecular Clouds

In a self-gravitating turbulent medium, the collapse of gas clouds can be affected by turbulence in two different ways.

1. Turbulent motion increases the velocity dispersion of the gas - delay or suppress gravitational collapse.

2. Gas can be compressed by shocks produced by the supersonic flow - promote collapse by increasing the gas density.

Thus, turbulence may suppress gravitational collapse globally, but can promote local gravitational collapse. 

Question is whether shock velocity is comparable to turbulent velocity.
9.1.3 Mechanisms Driving Turbulence

The cold gas in a star-forming galaxy is assembled through the cooling and accretion of gas in a dark matter halo.

Accreted gas contains angular momentum - settles into a disk-like structure supported by angular momentum

The accreted gas is expected to be clumpy – especially in mergers, where large chunks of gas are being added into the galaxy – and to generate high levels of turbulence in the gas.

Cosmological simulations of galaxy disk formation show that gravitational instabilities + shear flows + tidal interactions - generate high levels of turbulence in the ISM on large scales.

The large-scale turbulence may then cascade down in scales before it is dissipated, producing a supersonic turbulent medium similar to the ISM.
9.3.1 The Kennicutt-Schmidt law
9.3.2 Local Star Formation Laws

High resolution data for $\Sigma^*\,\,\,\,\Sigma_{\text{gas}}$ and $\Omega$ as function of galactocentric radius - although there is a roughly linear trend between SFE and orbital frequency among galaxies no strong relation exists within galaxies. **Orbital time not important.**

The $\Sigma^*\,\,\,\Sigma_{\text{gas}}$ relation reveals a pronounced break at low surface densities of the gas.
9.4 The Initial Mass Function

\[ \phi(m) \propto m^{-2.35} \]
9.4 Observationally Deriving the IMF

Mass is generally not observable - estimated from luminosity.

IMF derived in 4 steps:

1) measure the apparent magnitudes and distances of all stars in a particular volume down to some magnitude limit (only possible in the Milky way).

2) Observed apparent magnitudes are converted into absolute magnitudes, \( M \), using their distances, and LF is derived, \( \Phi(M) \), defined as the number density of stars as a function of their absolute magnitude.

3) LF of stars is transformed into a MF, using M/L from stellar evolution models.

4) The IMF is obtained from the MF, correcting for stars that died (depends on the star formation history)
IMF Based on the Core Mass Function

Based on the mass distribution of molecular cloud cores expected in a turbulent medium.

Since the strength of compression in a shock must be related to the pre-shock spatial scale, the mass function of dense cores in the GMCs generated by supersonic turbulence with a power-law spectrum, $|\Delta v|(l) \propto l^q$ (which is the rms velocity on scale $l$), should have a mass function

$$\frac{dN}{d\ln m} \propto m^{-3/(3-2q)}.$$  \hspace{1cm} (9.44)

For the expected turbulence $q = 0.5$ and $N = -1.5$.

Since only high-density cores collapse, the IMF should be related to the gravitationally-bound fraction of the mass function. For a given gas density $n$, the thermal Jeans mass is:

$$m_J = m_{J,0} \left( \frac{n}{n_0} \right)^{-1/2}, \text{ where } m_{J,0} \approx 1.2 \times \left( \frac{T}{10K} \right)^{3/2} \left( \frac{n_0}{1000 \text{ cm}^{-3}} \right)^{1/2} M_\odot \hspace{1cm} (9.45)$$

The distribution function of local Jeans mass is $P(y)dy = 2y^2 p[x(y)]dy$, where $y \equiv m/m_{J,0} = x^{-1/2}$. For cores with a given mass $m$, only the fraction with local Jeans mass $m_J < m$, $f(m) = m/m_{J,0} P(y)dy$, can collapse. Hence, the mass function of the cores that can collapse is:

$$\frac{dN}{d\ln m} \propto f(m)m^{-3/(3-2q)}.$$  \hspace{1cm} (9.46)

If $p(x)$ is log-normal, then $f(m) \sim 1$ for $m \gg m_{J,0}$, and $f(m) \sim 0$ for $m \ll m_{J,0}$. Therefore, the resulting core mass function is a power law. Such a core mass function has the same characteristics as the observed IMF of stars!
10. Stellar Populations and Chemical Evolution

How individual stars evolve with time?

Most stars, during most of their evolution, can be considered as:
- spherically symmetric
- with constant mass
- and in hydrostatic equilibrium.

Under these conditions, the evolution of a star is almost completely determined by its mass and chemical composition through a set of ordinary differential equations that describe the structure of the star.
Spherically Symetric?

(i) stars are in hydrostatic state
(ii) stars are spherically symmetric.

Structure of a star described by a set of equations in which all quantities depend only on the distance to the center.

Of course stars are not completely spherical; as a gaseous body a star can be flattened by rotation.

Consider an element of mass $m$ near the equator of a star.

The centrifugal force on the mass element due to rotation is $m\omega^2 R$, where $\omega$ is the angular velocity of the rotation and $R$ is the equatorial radius of the star.

This force will not cause significant departure from spherical symmetry if it is much smaller than the gravitational force, $GMm/R^2$,

if $\omega^2 R^3 / GM \ll 1$.

For the Sun, $\omega \approx 2.5 \times 10^{-6}\text{s}^{-1}$, $\omega^2 R^3 / GM \approx 2 \times 10^{-16}$, and so flattening due to rotation is completely negligible.

In general, normal stars can be considered to be spherical to very high accuracy.
To check hydrostatic equilibrium, consider the acceleration of a mass element at a radius $r$ in a spherical star:

$$
\ddot{r} = -\frac{GM}{r^2} - \frac{1}{\rho(r)} \frac{\partial P}{\partial r}.
$$

(10.1)

partial derivative of $P$ is used because the pressure is a function of both radius $r$ and time $t$.

Two time scales based on the ratio between the radius of the star, $R$, and the acceleration $|\ddot{r}|$.

1) gravitational dynamical time scale, $t_{\text{dyn}} = (G\rho)^{-1/2}$, which corresponds to setting the pressure gradient to zero.

2) hydrodynamical time scale, $t_{\text{hydro}} = R/c_s$ (with $c_s$ the sound speed of the stellar gas), which corresponds to setting the gravitational term to zero.

For the Sun with $\rho \sim 1$ g cm$^{-3}$, $R \sim 7 \times 10^{10}$ cm and $T \sim 7000$K, these two time scales are $t_{\text{dyn}} \sim 50$min and $t_{\text{hydro}} \sim 1$day, respectively.

These two time scales characterize how fast the gravity and sound waves can adjust a perturbed stellar structure to a new equilibrium configuration - should be compared to any perturbator.
Hydrostatic Equilibrium?

**Thermal time scale of a star**, ratio between
- thermal energy of a star, $E_{th}$
- energy loss rate at its surface, $L$: $t_{th} = E_{th}/L$.

Gives the rate at which a star changes its structure by radiating its thermal energy.
For the Sun, $t_{th} \sim 10^7$ yr. Since this is much longer than both $t_{dyn}$ and $t_{hydro}$, the structure of a normal star can adjust quickly to a new configuration as it radiates.

Main energy source in a normal star is nuclear reactions and, for a star in equilibrium, the energy generation rate is equal to its luminosity.

**Nuclear energy generation time-scale**, ratio between
- energy resource in a star
- current luminosity, $L$: $t_{nuc} \sim \eta Mc^2/L$, $\eta \sim 10^{-3}$ is the efficiency of nuclear reactions converting rest-mass to radiation

For the Sun, this time scale is too long, about $10^{10}$ yr.

Thus, for a normal star like the Sun, $t_{nuc} \gg t_{th} \gg t_{dyn}$. Implies that the evolution of a star is determined by nuclear reactions, with thermal and mechanical equilibrium achieved during the evolution.
Basic Equations of Stellar Structure

Eqs. (10.2), (10.3), (10.7), and (10.9), are the four structure equations for the five quantities $\rho$, $M$, $P$, $T$, $L$. Together with the equation of state, they provide a complete description of a radiative star with given composition $\{X_i\}$, provided that $P$, $\kappa$ and $\varepsilon$ are known functions of $(\rho, T, \{X_i\})$.

\[
\frac{dP(r)}{dr} = -\frac{GM(r)\rho(r)}{r^2}, \quad (10.2)
\]

\[
\frac{dM(r)}{dr} = 4\pi r^2 \rho(r). \quad (10.3)
\]

\[
P = P(\rho, T, \{X_i\}), \quad (10.4)
\]

\[
\frac{dT}{dr} = -\frac{3\kappa L\rho}{16\pi a_c r^2 T^3}, \quad (10.7)
\]

\[
\frac{dL}{dr} = 4\pi r^2 \rho \varepsilon. \quad (10.9)
\]
Stellar Nucleosynthesis and Energy Production

Our Sun is currently emitting with a luminosity $L_\odot \approx 3.9 \times 10^{33}$ ergs$^{-1}$.

Based on radioactive dating of rocks on Earth, it can be inferred that the Sun has been shining with such a luminosity for at least about $5 \times 10^9$ yr.

The total energy that has been released by the Sun is therefore about $6 \times 10^{50}$ erg, or about a fraction of $3 \times 10^{-4}$ of its present total mass.

An obvious question is, what is the basic energy source for the Sun?

- The gravitational energy, $GM_\odot^2/R_\odot$, can sustain the Sun at its present luminosity for a period of $GM_\odot^2/R_\odot L_\odot \sim 3 \times 10^7$ yr, much too short compared to the age of $5 \times 10^9$ yr. (Before dating of rocks on Earth that was believed to be the age of the Sun.)

- The total thermal energy, which is about half of the gravitational energy, according to the virial theorem, is also far too small to supply the energy.

- Chemical reactions associated with the burning of material can transform a fraction of $\sim 10^{-10}$ of the rest mass into heat. Much smaller than $3 \times 10^{-4}$ (also not enough)

- **The only energy source that can convert rest mass into energy with an efficiency larger than $3 \times 10^{-4}$ is nuclear reactions.**

Rest mass of He is $3.97m_p$. Mass converted into energy is therefore $4m_p - 3.97m_p = 0.03m_p$, or about $7 \times 10^{-3}$ of the original mass, $4m_p$.

Large enough to power the Sun for $8 \times 10^{10}$ yr. In fact structure and luminosity of the Sun change drastically once it uses 13% of its total hydrogen. So fusion of H to He can sustain the Sun for about $9 \times 10^9$ yr (twice its age)
Binding energy per nucleon becomes larger for heavier elements. Reversed for elements heavier than $^{56}$Fe.

Balance between the net attraction of the strong nuclear force and the repulsion of the positively charged protons.

For small nuclei, the short-range nuclear force wins over the long range electric force, causing the binding energy per nucleon to increase when a nucleon (proton or neutron) is added to the nucleus.

However, since the nuclear density is nearly constant, heavier nuclei are also physically bigger. The increase in the binding energy per nucleon with the number of nucleons saturates with iron-56; further addition of nucleons actually reduces the binding energy per nucleon. This is the reason why very heavy nuclei, such as uranium, are unstable.
Stellar Nucleosynthesis and Energy Production

If it is more advantageous for nucleons to be synthesized into heavy elements why is most of the cosmic material still in the form of light elements such as H and He?

**Nuclear fusion reactions can occur only under certain conditions.** Since atomic nuclei are positively charged, the two (or more) nuclei in a fusion reaction tend to repel each other due to the Coulomb force.

In order for the components to come close enough so that the attractive nuclear force becomes dominant, they have to move rapidly towards each other to overcome the Coulomb barrier.

Nucleosynthesis is efficient only where **gas density and temperature are both high** so that many high-velocity collisions among nuclei can occur. Stellar interiors are such places.
pp chain and CNO cycle

There are two important channels for converting hydrogen into helium. The first is through the pp chain and the second is through the CNO cycle. The pp chain starts with two hydrogen nuclei (i.e. two protons) to form a deuteron (D) which then captures another proton to form a $^3$He:

$$p + p \rightarrow D + \bar{e} + \nu_e, \quad D + p \rightarrow ^3\text{He} + \gamma. \quad (10.45)$$

The $^3$He so created is converted into $^4$He either through the collision of two $^3$He,

$$^3\text{He} + ^3\text{He} \rightarrow ^4\text{He} + 2p, \quad (10.46)$$

or, in the presence of $^4$He, through

$$^3\text{He} + ^4\text{He} \rightarrow ^7\text{Be} + \gamma, \quad ^7\text{Be} + e \rightarrow ^7\text{Li} + \nu_e, \quad ^7\text{Li} + p \rightarrow 2^4\text{He}. \quad (10.47)$$

The CNO-cycle involves the following set of reactions:

$$^{12}\text{C} + p \rightarrow ^{13}\text{N} + \gamma, \quad ^{13}\text{N} \rightarrow ^{13}\text{C} + \bar{e} + \nu_e, \quad ^{13}\text{C} + p \rightarrow ^{14}\text{N} + \gamma,$$

$$^{14}\text{N} + p \rightarrow ^{15}\text{O} + \gamma, \quad ^{15}\text{O} \rightarrow ^{15}\text{N} + \bar{e} + \nu_e, \quad ^{15}\text{N} + p \rightarrow ^{12}\text{C} + ^4\text{He}. \quad (10.48)$$

The net effect of the CNO-cycle is to convert four protons into a helium nucleus, with C, N, and O nuclei acting only as catalyst. Note that the completion of the CNO-cycle requires the pre-existence of carbon.
pp chain and CNO cycle

The rate with which the pp-chain and the CNO-cycle convert H into He is proportional to the square of the density (because the reactions involved are two-body in nature) and increases rapidly with temperature.

The CNO-cycle is more sensitive to temperature because the elements involved are heavier.

Rate of energy production (from each reaction) = reaction rate x energy released from each conversion (function of T)

The specific energy production rate, $\varepsilon$, is in general a smooth function of $T$, and can be represented as a power law of $T$:

$$\varepsilon = \varepsilon_0 \rho T^n.$$  \hspace{1cm} (10.49)

In the temperature range where the pp-chain is important,

$$\varepsilon \propto X_H^2 \rho T^4 \quad \text{(pp-chain)},$$  \hspace{1cm} (10.50)

while in the temperature range where the CNO-cycle dominates,

$$\varepsilon \propto X_H X_C \rho T^{17} \quad \text{(CNO-cycle)}.$$  \hspace{1cm} (10.51)

Since thermonuclear reactions need high temperature to proceed, there must be systems in which the temperature is never high enough to ignite H-burning. Such stars are called black dwarfs, because they only radiate very little.

The minimum temperature corresponds to a minimum mass, $M_{\text{min}} \approx 0.08M_\odot$.

Might still burn deuterium (which has a lower ignition temperature) - brown dwarf.
Helium Burning - triple-alpha

For systems with masses larger than $M_{\odot}$, the conversion of H to He first occurs in the central region where the temperature is the highest.

1) There will then be a time when most hydrogen in the central region is exhausted.

2) Star contracts with no thermonuclear reaction to provide pressure support and thermal energy flows out.

3) According to the virial theorem, such contraction leads to increase in the T in the centre of the star.

4) Contraction continues until the next important nuclear reaction takes place to generate sufficient thermal energy to halt the contraction.

As can be inferred from Eq. (10.44), when the temperature reaches $\sim 10^8$ K, fusion of helium begins to synthesize the next stable nuclei, $^{12}$C, through the following reactions:

$$^4\text{He} + ^4\text{He} \rightarrow ^8\text{Be} + \gamma, \quad ^8\text{Be} + ^4\text{He} \rightarrow ^{12}\text{C} + \gamma.$$ \hspace{1cm} (10.52)

Called the triple-\(\alpha\) process because three \(\alpha\) particles (\(^4\text{He}\)) are involved.

Since \(^8\text{Be}\) is unstable and can decay back to two \(^4\text{He}\) in a short time (after the end of first reaction), only small amounts of \(^{12}\text{C}\) would be produced if the second reaction occurred with a ‘normal’ rate. This is known as the ‘bottleneck’ for the synthesis of heavy elements.

A breakthrough of this ‘bottleneck’-problem occurred when it was realized that the second reaction can proceed through a fast (resonant) channel with the production of an excited state \(^{12}\text{C}^*\), which then de-excites to the ground level \(^{12}\text{C}\) (Hoyle, 1954). Once \(^{12}\text{C}\) is produced, the synthesis of heavy elements can proceed further.
Heavier Elements

Two groups of elements are particularly important.

One group is called the \( \alpha \)-elements, which includes \( ^{16}\text{O} \), \( ^{20}\text{Ne} \), \( ^{24}\text{Mg} \), \( ^{28}\text{Si} \), \( ^{32}\text{S} \), \( ^{36}\text{A} \), \( ^{40}\text{Ca} \) (all the elements can be formed by adding \( \alpha \)-particles to \( ^{12}\text{C} \)).

The synthesis of these elements is either through the capture of \( \alpha \)-particles in reactions such as:

\[
^{12}\text{C} + ^{4}\text{He} \rightarrow ^{16}\text{O} + \gamma \quad \text{and} \quad ^{20}\text{Ne} + ^{4}\text{He} \rightarrow ^{24}\text{Mg} + \gamma, \quad (10.53)
\]

or through the burning of C and O in reactions like

\[
^{12}\text{C} + ^{12}\text{C} \rightarrow ^{20}\text{Ne} + ^{4}\text{He} \quad \text{and} \quad ^{16}\text{O} + ^{16}\text{O} \rightarrow ^{28}\text{Si} + ^{4}\text{He}. \quad (10.54)
\]

Can be synthesized in a star that starts out with pure hydrogen and helium - abundances are independent of the initial metallicity.

The other group, called the iron-peak elements, includes all elements with atomic numbers in the range \( 40 < A < 65 \), i.e. Sc, Ti, V, Cr, Mn, Fe, Co, Ni and Cu.

High nuclear charges - form late in a star, when its core becomes extremely hot.

Elements that can be synthesized directly in stars of zero initial metallicity are referred to as primary elements, while elements that can form only in the presence of primary elements are called secondary elements.
Fig. 10.3. Post-main-sequence evolutionary tracks of stars with Solar metallicity. To avoid confusion, tracks for stars with masses smaller than $2M_\odot$ are terminated at their He-flash. The six points on each track mark the positions reached by the star after the times from the zero-age main sequence listed in Table 10.1. The shaded area is the instability strip within which various variable stars are located. [Based on data published in Girardi et al. (2000)]
Low Mass Stars

Even after hydrogen has been exhausted in the core of a star, energy transport will continue in its interior.

1) Since the core is still too cold to burn He, there is no thermal energy to counterbalance the energy leak, and so the core contracts.

2) As it contracts, the core is heated up (according to the virial theorem), and so is the gas layer above the He core.

3) Since hydrogen has a lower ignition temperature than helium, the hydrogen in the shell just above the He core will burn first as the temperature increases, while the He core itself still remains dormant.

The H shell-burning is quite effective in generating energy. But as long as the star is not convective, the amount of energy that can leak from the star is limited by photon diffusion and is roughly a constant (generates extra heat causing the star to expand). Luminosity almost constant and a slight increase in T.

4) As the temperature in the outer envelope drops (due to expansion), the temperature gradient between the outer and inner regions increases, and the star eventually becomes fully convective.

When this happens the effective temperature cannot decrease any further and the star start to ascend almost vertically in the $T_{\text{eff}} - L$ plane to the red giant branch (RGB) - Luminosity increases because energy can leak.

At this stage, the envelope of the star has been distended so much that substantial mass-loss (stellar wind driven by the radiation pressure) may occur.

5) The He core becomes hotter and hotter as it contracts and as more He ash is added to it from the H-burning shell. When the core temperature reaches $\sim 10^8$K, helium begins to burn into carbon through the triple-$\alpha$ process; s

6) The He-burning heats the core and causes it to expand, lowering the gravity on the H-burning shell and thereby reducing the strength of shell H-burning. Consequently, the star reaches the tip of the red-giant branch and then settles onto the horizontal branch (HB) where the star is mainly powered by He burning in its core. The He core behaves like a He main-sequence star.
Low Mass Stars

Fig. 10.4. An illustration of the complete evolutionary track of low-mass stars.
Low Mass Stars

7) The core He burning will come to a halt once most of the helium nuclei in the core are synthesized into carbon and oxygen. At this stage, the star contains a C/O core, an inner He-burning shell and an outer H-burning shell.

8) As the core temperature is still too low for carbon to burn, the C core contracts. The He-burning shell outside the C core and the H-burning shell outside the He-burning shell also contract in this phase because of the loss of pressure support. Such contraction enhances the double-shell burning, causing the star to ascend the Asymptotic Giant Branch (AGB) in a way similar to what happened during the RGB phase.

During the AGB phase, the envelope of the star can be greatly distended, and a large amount of mass loss can occur. When the star reaches the top of the AGB, where it becomes a red supergiant, it can lose all of its blanketing hydrogen layer.

9) The star then makes a sudden move leftwards in the H-R diagram as it peels away its relatively cold mantel. At this stage, the intense ionizing radiation from the star may cause the ejected H envelope to shine as a fluorescent planetary nebula.

10) The He burning will gradually die off, leaving behind a C/O core with a mass in the range $0.55 - 0.6 \, M_\odot$ (independent of the initial mass). Because of the small mass, further nuclear burning cannot be ignited in the core. The star will then contract until it is supported by the degeneracy pressure of electrons, becoming a white dwarf.

Valid for all stars with initial masses $M_\star < 8 \, M_\odot$. 
Massive Stars

For stars with initial masses $M_s > 8 M_\odot$, core-H exhaustion, He-core contraction, shell-H burning, and core-He ignition proceed in much the same way as for low-mass stars.

1) However, because of the high mass, core-ignition and shell-burning can proceed successively to heavier elements. During this process, a star in general moves to a higher $T_{\text{eff}}$ after a core ignition, and to a lower $T_{\text{eff}}$ during core contraction. Since a massive star can maintain radiative equilibrium during the evolution, its luminosity changes only little. As a result, the evolutionary tracks of such stars appear more-or-less horizontal in the H-R diagram.

2) At some stage, all the material in the core may be converted into iron-peak elements that cannot be ignited anymore. The burning in the core will then cease, while Si, O, Ne, C, He, and H continue to burn in shells at successively larger radii. When this happens, the core contracts until it becomes supported by the degeneracy pressure of electrons.

3) Since the shell-burning outside the core keeps dropping ash onto the core, the degenerate core will grow until it reaches Chandrasekhar’s mass limit, $M \approx 1.4 M_\odot$, beyond which the iron-core can no longer be supported by the degeneracy pressure of electrons and must collapse further. This collapse can heat the core to such a high temperature ($>10^{9}$ K) that iron nuclei can be photo-disintegrated into $\alpha$-particles ($^4$He nuclei).

4) Since thermal energy is used in the photo-disintegration, the core will collapse further and be heated to an even higher temperature. The $\alpha$-particles produced will then also be photo-disintegrated into protons and neutrons, draining more thermal energy from the core.

5) The core then undergoes a phase of catastrophic contraction, and its density can become so high that almost all electrons are squeezed into protons to form neutrons. If the mass is smaller than $\sim 3 M_\odot$, the core will be a neutron star supported by the degeneracy pressure of neutrons; otherwise the core will collapse further to form a black hole. The collapse of the iron-core is so violent that a supernova explosion may take place.
10. Star formation driven feedback

Stars can output energy in three different channels:

- radiation
- neutrino emission
- mass flow

The energy output in radiation can be obtained by integrating the luminosity of a star over its lifetime. Typically, a star of Solar mass emits about $10^{35.5} \text{erg}$ in radiation, while the corresponding numbers for $10 \, M_\odot$ and $100 \, M_\odot$ stars are $\sim 10^{52.5} \text{erg}$ and $\sim 10^{53.7} \text{erg}$, respectively.

A star also emits $\sim 10^{56} \text{neutrinos}$ when it becomes a white dwarf or a red giant, and the total energy involved is of the order $10^{53} \text{erg}$. 
10.5.1 Mass-Loaded Kinetic Energy - Stellar Winds

Some stars can lose substantial amounts of mass via stellar winds during their late evolutionary stage. Mass loss can be calculated from stellar evolution theory:

Table 10.4. *The ejected masses (in $M_\odot$) from winds and Type II SNe for stars with Solar metallicity. [Based on data published in Marigo et al. (1996) and Portinari et al. (1998)]*

<table>
<thead>
<tr>
<th>$M_s$</th>
<th>$M_{ej}^{\text{wind}}$</th>
<th>$M_{ej}^{\text{total}}$</th>
<th>He</th>
<th>C</th>
<th>N</th>
<th>O</th>
<th>Ne</th>
<th>Mg</th>
<th>Fe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>0.43</td>
<td>0.43</td>
<td>0.13</td>
<td>0.001</td>
<td>0.001</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.49</td>
<td>0.91</td>
<td>0.91</td>
<td>0.27</td>
<td>0.003</td>
<td>0.002</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td>1.40</td>
<td>1.40</td>
<td>0.43</td>
<td>0.011</td>
<td>0.004</td>
<td>0.014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.00</td>
<td>2.34</td>
<td>2.34</td>
<td>0.73</td>
<td>0.014</td>
<td>0.007</td>
<td>0.022</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.00</td>
<td>3.14</td>
<td>3.14</td>
<td>0.96</td>
<td>0.014</td>
<td>0.010</td>
<td>0.029</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.00</td>
<td>4.70</td>
<td>4.70</td>
<td>1.58</td>
<td>0.017</td>
<td>0.019</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.00</td>
<td>5.70</td>
<td>5.70</td>
<td>2.15</td>
<td>0.019</td>
<td>0.027</td>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.00</td>
<td>0.32</td>
<td>7.69</td>
<td>2.99</td>
<td>0.074</td>
<td>0.040</td>
<td>0.178</td>
<td>0.021</td>
<td>0.007</td>
<td>0.057</td>
</tr>
<tr>
<td>12.0</td>
<td>0.53</td>
<td>10.56</td>
<td>3.95</td>
<td>0.170</td>
<td>0.055</td>
<td>0.712</td>
<td>0.125</td>
<td>0.028</td>
<td>0.121</td>
</tr>
<tr>
<td>15.0</td>
<td>0.78</td>
<td>13.13</td>
<td>4.70</td>
<td>0.268</td>
<td>0.067</td>
<td>1.27</td>
<td>0.177</td>
<td>0.036</td>
<td>0.098</td>
</tr>
<tr>
<td>20.0</td>
<td>1.90</td>
<td>17.89</td>
<td>6.01</td>
<td>0.281</td>
<td>0.089</td>
<td>2.76</td>
<td>0.217</td>
<td>0.069</td>
<td>0.186</td>
</tr>
<tr>
<td>30.0</td>
<td>17.11</td>
<td>22.82</td>
<td>8.28</td>
<td>0.310</td>
<td>0.128</td>
<td>3.08</td>
<td>0.504</td>
<td>0.124</td>
<td>0.022</td>
</tr>
<tr>
<td>40.0</td>
<td>34.00</td>
<td>37.94</td>
<td>19.05</td>
<td>3.03</td>
<td>0.211</td>
<td>2.06</td>
<td>0.417</td>
<td>0.039</td>
<td>0.133</td>
</tr>
<tr>
<td>60.0</td>
<td>52.65</td>
<td>57.91</td>
<td>32.14</td>
<td>4.93</td>
<td>0.343</td>
<td>2.66</td>
<td>0.474</td>
<td>0.056</td>
<td>0.234</td>
</tr>
<tr>
<td>100.0</td>
<td>91.89</td>
<td>97.88</td>
<td>55.48</td>
<td>11.40</td>
<td>0.682</td>
<td>4.28</td>
<td>1.07</td>
<td>0.130</td>
<td>0.251</td>
</tr>
</tbody>
</table>
For supernova explosions, the typical mass-loaded kinetic energy can be estimated by modeling in detail supernovae observed in the local Universe:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_{ej}/M_\odot$</td>
<td>15</td>
<td>17</td>
<td>2.2</td>
<td>3.26</td>
<td>1.1</td>
<td>0.87</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>$E_{51}$</td>
<td>1.7</td>
<td>1.7</td>
<td>1.0</td>
<td>1.7</td>
<td>1.48</td>
<td>1.18</td>
<td>1.0</td>
<td>0.69</td>
</tr>
</tbody>
</table>

The mass-loaded kinetic energies are quite similar for Type Ia and Type II SNe.
8.6.1 Blast Waves - Supernova

Let's consider supernova remnants, which are produced by supernova explosions at the late evolutionary stages of relatively massive stars.

The typical energy output of such an explosion is about $10^{51}$ erg, released in seconds. Almost all this energy is initially in the form of kinetic energy of the ejecta.

Blast wave consists of four well-defined stages:

1) first the ejected mass exceeds the swept-up ambient mass, and to lowest order the ejecta undergo free expansion.

2) once the mass of the swept-up material becomes comparable to the mass of the ejecta, the blast-wave enters the adiabatic (or Sedov) phase, in which the evolution is self-similar as long as the ambient medium is homogeneous.

3) Eventually the radiative losses from the interior of the blast-wave become significant, and the supernova remnant enters the third, radiative stage of its evolution.

4) Finally, once the interior pressure becomes comparable to that of the ambient medium, the supernova remnant merges with the ISM.
If the energy input from star formation is equal to the binding energy of the cold gas, the star formation rate, $\dot{M}_\star$, is given by $E_0\dot{M}_\star = (\dot{M}_g - \dot{M}_\star)V_c^2/2$, where $E_0$ measures the energy feedback per unit mass of formed stars, and the right-hand side is a crude estimate of the binding energy of the cold gas. Solving for $\dot{M}_\star$ we obtain

$$\dot{M}_\star = \frac{\dot{M}_g}{1 + (V_0/V_c)^2}, \quad (8.214)$$

where $V_0^2 = 2E_0$.

Based on the discussion presented above, we have $V_0 \sim V_{\text{crit}}$. In this simple model, the star formation efficiency in halos with $V_c \ll V_{\text{crit}}$ is reduced by a factor proportional to $V_c^2$.

It is necessary to suppress the efficiency of star formation in low-mass halos in order to explain the observed galaxy luminosity function at the faint end in the CDM scenario of galaxy formation.

**Star formation feedback through supernova explosions provides an appealing mechanism.**
14.1 AGN, Black Holes and feedback

Object defined as AGN if one or more of the following properties are observed:

1) a **compact nuclear region much brighter** than a region of the same size in a normal galaxy;

2) **non-stellar (non-thermal) continuum emission**;

3) **strong emission lines**;

4) **variability** in continuum emission and/or in emission lines on relatively short time scales.

AGN population is classified into subgroups according to their observational properties. Some elements are mainly historical and do not necessarily differentiate the underlying physical processes.

Table 14.1 gives number densities.

AGN population only makes up a relatively minor fraction of all galaxies observed at the present time. However, since they are believed to have short lifetimes it is possible that many galaxies had an AGN phase. This is supported by the fact that virtually all spheroids seem to harbor a SMBH.

<table>
<thead>
<tr>
<th>Type of object</th>
<th>Number density [Mpc$^{-3}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field galaxies</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td>Luminous spirals</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>Seyfert galaxies</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>Radio galaxies</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>QSOs</td>
<td>$10^{-7}$</td>
</tr>
<tr>
<td>Quasars</td>
<td>$10^{-9}$</td>
</tr>
</tbody>
</table>
Seyfert Galaxies

Named after their discoverer Carl Seyfert (1943), are:

1) active galaxies
2) with spiral-like morphologies
3) with bright star-like nuclei.
4) spectra have non-thermal continua and contain strong and broad emission lines of high excitation (Figure).

radiation from the nuclei can vary by a factor of more than two in less than a year - sizes of these nuclei are smaller than one light year across.

Fig. 14.1. A composite spectrum of QSOs revealing the typical non-thermal continuum and various emission lines. Lines in brackets are forbidden lines, those in semi-brackets are semi-forbidden lines, and lines without brackets are permitted lines. Note that the permitted lines are much broader than the forbidden lines, as is typical for QSOs and Seyfert 1 galaxies. [Courtesy of C. Foltz and P. Hewett, based on an extension of the data published in Francis et al. (1991)]
Seyfert Galaxies

Subdivided into two categories, Seyfert 1 and Seyfert 2, according to the widths of their emission lines.

**Seyfert 1 galaxies:**
- the permitted lines, mainly from hydrogen, are very broad with full widths at half maxima (FWHM) corresponding to velocities in the range $1000 - 5000 \text{ km s}^{-1}$.
- the forbidden lines, such as [OIII], are much narrower, with FWHM corresponding to velocities of a few hundred km s$^{-1}$.

**In Seyfert 2 galaxies:**
- both permitted and forbidden lines have **narrow velocity widths**, typically of a few hundred km s$^{-1}$.

**In general, Seyfert 1 galaxies have stronger non-stellar continua and stronger hard X-ray emission than Seyfert 2 galaxies.**
Radio Galaxies

Radio galaxies have strong radio emission.

Normal spirals have weak radio emission (due to supernova remnants), with power $P_{1.4\text{GHz}} < 2 \times 10^{23} \text{W Hz}^{-1}$ at $\sim 1.4$ GHz, so radio galaxies are commonly defined as galaxies with $P_{1.4\text{GHz}}$ larger than this.

Initially identified in radio surveys, being later identified with optical sources, which has revealed that almost all host galaxies of radio AGN are ellipticals.

The spectra of the optical sources associated with radio galaxies typically reveal strong emission lines.

Similar to Seyfert galaxies, one distinguishes radio galaxies between broad-line radio galaxies (BLRGs) and narrow-line radio galaxies (NLRGs).

In principle BLRGs and NLRGs can be considered as radio-loud Seyfert 1 and Seyfert 2 galaxies, respectively, albeit with a different morphology of the host galaxy.
Radio Galaxies

Radio galaxies are also distinguished based on their radio morphology.

**Powerful radio galaxies:**

have a double-lobed structure extending to several hundred kiloparsecs or even megaparsecs from the central nucleus that coincides with the center of the host galaxy.

jet-like structures are observed to stretch from the compact core towards the lobes, suggesting that these jets are responsible for transporting energy from the core out into the radio lobes.

The jets are rarely symmetric: often only one jet is observed, and in sources with two jets, one is usually significantly brighter than the other.

![Radio galaxy image](image-url)
Radio Galaxies

Radio galaxies are also distinguished based on their radio morphology.

**Weak radio galaxies:**
Radio emission has a more irregular and more concentrated structure.

Based on radio morphology Fanaroff & Riley (1974) divided radio galaxies into two subgroups:

FR I: the distance between the two most intense spots on either side of the nucleus (called ‘hot spots’) is smaller than half the overall source size

FR II: this distance is larger.

It turns out that class FR II are powerful radio sources, with $P_{1.4\text{GHz}} > 4 \times 10^{25} \text{WHz}^{-1}$, while class FRI typically have weaker radio power.
Quasars and QSOs

The name quasar (from Quasi-Stellar Radio Source) was originally used for the optical identifications of compact radio sources with Seyfert-like spectra.

Radio characteristics are similar to powerful radio sources, but:

- optical images are unresolved (at θ~1′)
- luminous ($M_b<-21.5 + 5 \log h$) nuclei that are unusually blue and often variable.

Radio observations of high spatial resolution have shown that flat-spectrum quasars often contain very compact nuclei (~10^{-3} arcsec).

Some of these have elongated structure (jets) consisting of two or more closely separated components whose relative motion to each other can be detected on time scales of a few years.
14.2 The Supermassive Black–Hole Paradigm

An AGN is extraordinary in that it emits a large amount of energy from a very small region. An obvious question is how this energy is generated.

The small size of the emission region and the large amount of energy output suggest that the central engine must be compact and have relatively large mass.

It is now generally believed that this central engine is a supermassive black hole (SMBH), an idea originally proposed by Salpeter (1964), Zel’dovich & Novikov (1964) and Lynden-Bell (1969).

In addition to the central engine, the black hole (BH), the standard AGN model also assumes the existence of other components, such as broad and narrow line regions (BLR and NLR), an accretion disk, and jets.

Fig. 14.3. Different components of an AGN in the standard paradigm.
14.2.1 The Central Engine

In the SMBH paradigm, an AGN is assumed to be powered by a SMBH accreting gas, and the energy source is the gravitational potential of the central black hole.

To comprehend the energy scale involved with such accretion, consider a simple spherical model in which a central source with luminosity $L$ is surrounded by gas with density distribution $\rho(r)$. The flux at a radius $r$ from the source is $L/(4\pi r^2)$ and the radiation pressure is:

$$P_{\text{rad}}(r) = \frac{L}{4\pi r^2 c}. \tag{14.1}$$

If the gas is ionized, the pressure force on a unit volume of gas due to the scattering of photons by electrons is

$$\mathbf{F}_{\text{rad}} = \sigma_T P_{\text{rad}}(r)n_e(r)\hat{\mathbf{r}}, \tag{14.2}$$

where $\sigma_T$ is the Thompson scattering cross section, and $n_e(r)$ is the electron density at radius $r$. In order for the gas not to be dispersed quickly, this pressure force must be smaller than the gravitational force on the gas, i.e.,

$$|\mathbf{F}_{\text{rad}}| \leq F_{\text{grav}} = \frac{GM_{\text{BH}}\rho(r)}{r^2}. \tag{14.3}$$
14.2.1 The Central Engine

This defines a maximum luminosity (the Eddington luminosity) for a given black hole $M_{\text{BH}}$:

$$|F_{\text{rad}}| \leq F_{\text{grav}} = \frac{G M_{\text{BH}} \rho(r)}{r^2}. \quad (14.3)$$

with $m_p$ the proton mass. This Eddington luminosity is the largest possible luminosity that can be achieved by spherical accretion (if larger gas is dispersed too quickly and accretion stops).

Can be inverted to give a minimum central mass required to achieve a given luminosity,

$$L_{\text{Edd}} \equiv \frac{4\pi G c m_p}{\sigma_T} M_{\text{BH}} \approx 1.28 \times 10^{46} M_8 \text{erg s}^{-1} \quad (M_8 \equiv M_{\text{BH}}/10^8 M_\odot), \quad (14.4)$$

with $m_p$ the proton mass. This Eddington luminosity is the largest possible luminosity that can be achieved by spherical accretion (if larger gas is dispersed too quickly and accretion stops).

A bright quasar with a luminosity $L \sim 10^{46} \text{ erg s}^{-1}$ may be powered by a black hole with a mass $M_{\text{BH}} \sim 10^8 M_\odot$. 

$$M_{\text{Edd}} = 8 \times 10^7 L_{46} M_\odot, \quad [L_{46} \equiv L/(10^{46} \text{ erg s}^{-1})]. \quad (14.5)$$
The synchrotron emission from a relativistic electron is concentrated within a small beam, with an opening angle $\Delta \theta \sim \gamma^{-1}$, aligned with the instantaneous velocity of the electron. Thus, as an electron is spiraling in the magnetic field, the radiation signals received by an observer are short pulses.
Despite the many uncertainties, the formation of a SMBH generically requires a considerable amount of gas to be funneled into the center of a dark matter halo on a short time scale.

Even with $\varepsilon_r \sim 0.1$ and $L \sim L_{\text{Edd}}$, the timescale $t_{\text{BH}}$ is shorter than the free-fall timescale of a dark matter halo at $z < 5$.

Once a SMBH formed in the center of a galaxy, whether or not it can be observed as a bright AGN again depends on whether or not it is fed with gas at a sufficiently high rate.

The typical rate required to power a black hole at the Eddington luminosity is that given by Eq. (14.9), or $\sim 2 \, M_\odot \, \text{yr}^{-1}$ for a $10^8 M_\odot$ black hole.

\[
\dot{M}_{\text{Edd}} = \frac{L_{\text{Edd}}}{\varepsilon_r c^2} \approx 2.2 M_8 \left(\frac{\varepsilon_r}{0.1}\right)^{-1} M_\odot \, \text{yr}^{-1},
\]  

(14.9)

Note that the fueling of a bright AGN leads to further growth of the black hole mass, and the time scale of the growth is $M_{\text{BH}}/\dot{M}_{\text{BH}} \sim \varepsilon_t t_E (L/L_{\text{Edd}})^{-1}$.

This time scale is comparable to that for the growth of a SMBH accreting at the Eddington limit, indicating that the growth of a SMBH and the fueling of a bright AGN go hand in hand.

Hence, the key step towards understanding the formation of an AGN is therefore to identify the mechanism(s) that can effectively bring gas into the center of the host galaxy.
AGN fueling

One process to remove angular momentum from gas is gravitational interaction with other galaxies. Tidal forces during galaxy encounters can cause otherwise stable disks to develop bars, and gas in such barred disks, subjected to strong gravitational torques, can lose angular momentum very effectively and flow into the central region.

Numerical simulations show that the merger of a pair of galaxy-sized gaseous disks can produce a central cloud with mass $\sim 10^9 M_\odot$ and size $\sim 100$ pc. The detailed structure of such clouds is not yet resolved by current simulations, and it is unclear whether the gas can indeed be funneled directly into the central parsec region.

Significant gas inflow may also be generated in minor mergers in which a bar instability is induced in a disk by the perturbation from a smaller satellite galaxy. Although less effective, more common.

In some cases, a disk may become unstable spontaneously and form a bar-like structure. This may occur as a result of the interaction between the disk and other components of the galaxy, or when the disk becomes globally unstable during the formation process. It is therefore also possible that some AGN are fueled without involving strong interaction with other galaxies.
12.3 Dynamical Friction

An alternative but equivalent way to think about dynamical friction, is that the moving subject mass perturbs the
distribution of field particles causing a trailing enhancement (or “wake”) in their density. The gravitational force of this
wake on the subject mass $M_s$ then slows it down (see Fig. 12.3).

![Diagram of dynamical friction](image)

Fig. 12.3. As a massive object $M$ moves through a sea of particles, the particles passing by are accelerated
towards the object. As a result, the particle number density behind the object is higher that that in front of
it, and the net effect is a drag force (dynamical friction) on the object.
12.3 Dynamical Friction

The time it takes for a satellite to merge into a central galaxy due to the dynamical friction force can be calculated from the Chandrasekhar’s formula:

\[
F_{df} = -\frac{4\pi G^2 m_{\text{sat}}^2 \ln(\Lambda) \rho B(x)}{v_{\text{rel}}^2},
\]

where \( m_{\text{sat}} \) represents the satellite mass, \( \ln(\Lambda) \) is the Coulomb logarithm (assumed to be \( \ln (1 + M/m_{\text{sat}}) \)), \( \rho \) is the local density and \( v_{\text{rel}} \) represents the relative velocity of the satellite.

\[
t_{df} \approx 1.17 \frac{V_{\text{vir}} r_{\text{sat}}^2}{G m_{\text{sat}} \ln(\Lambda)},
\]
12.2 Tidal Stripping

We now examine how tidal forces impact a collisionless system in the more general case. As we will see, even in a static configuration tidal forces can strip material from the outer parts of a collisionless system, which is generally known as tidal stripping.

Assuming a slowly varying system (a satellite in a circular orbit) with a spherically symmetric mass distribution, material outside the tidal radius will be stripped from the satellite. This radius can be identified as the distance from the satellite centre at which the radial forces acting on it cancel.

These forces are the gravitational binding force of the satellite, the tidal force from the central halo and the centrifugal force and following King (1962) the disruption radius $R_t$ will be given by:

$$R_t \approx \left( \frac{G m_{\text{sat}}}{\omega_{\text{sat}}^2 - \frac{d^2 \phi}{dr^2}} \right)^{1/3},$$  \hspace{1cm} (5.1)

where $m_{\text{sat}}$ is the mass of the satellite, $\omega$ is its orbital angular velocity and $\phi$ represents the potential of the central object.
12.2 Tidal Stripping

From the assumption that satellites follow circular orbits we get $\omega_{\text{sat}}^2 = \frac{GM}{r^3}$ and

\[
R_t \approx \left( \frac{\sigma_{\text{sat}}^2 r^2}{\frac{GM}{2r} + \sigma_{\text{halo}}^2} \right)^{1/2}.
\]  

(5.6)

The final expression for the disruption radius will therefore be:

\[
R_t \approx \frac{1}{\sqrt{2} \sigma_{\text{halo}}} \sigma_{\text{sat}} r
\]

(5.7)

The material outside this radius is assumed to disrupted and becomes part of the intra-cluster medium component.
12.4 Galaxy Merging

the structure of the remnant of a merger between two galaxies depends primarily on four properties:

• The progenitor mass ratio \( q \equiv \frac{M_1}{M_2} \), where \( M_1 \geq M_2 \). If \( q < \sim 4 \) (\( q > \sim 4 \)) one speaks of a major (minor) merger. In major mergers violent relaxation plays an important role during the relaxation of the merger remnant, and the remnant typically has little resemblance to its progenitors. In minor mergers, on the other hand, phase mixing and/or Landau damping dominate, and the merger is less destructive. Consequently, the remnant of a minor merger often resembles its most massive progenitor.

• The morphologies of the progenitors (disks or spheroids).

Galactic disks are fragile, and are therefore relatively easy to destroy, especially when \( q \) is small. Disks that accrete small satellites (i.e., in the minor merger regime with \( q > \sim 10 \)) typically survive the merger event but can undergo considerable thickening. Mergers that involve one or more disk galaxies tend to create tidal tails, which are absent in mergers between two spheroids.
12.4 Galaxy Merging

• The gas mass fractions of the progenitors.

Unlike stars and dark matter particles, gas responds to pressure forces as well as gravity and can lose energy through radiative cooling. Moreover, gas flows develop shocks whereas streams of stars can freely interpenetrate. Consequently, mergers between gas-rich progenitors (often called ‘wet’ mergers) can have a very different outcome from mergers between gas-poor progenitors (‘dry’ mergers).

• The orbital properties.

The orbital energy and angular momentum not only determine the probability for a merger to occur, but also have an impact on the merger outcome. For example, the relative orientation of the orbital spin with respect to the intrinsic spins of the progenitors (prograde or retrograde) is an important factor determining the prominence of tidal tails.
12.4.3 The Connection between Mergers, Starbursts and AGN

An important aspect of (wet) mergers, and of interactions in general, is that they may be responsible for triggering (nuclear) starbursts and AGN activity.

Numerical simulations of mergers and encounters between gas-rich disks show that the tidal perturbations can cause the disks to become globally unstable and to develop pronounced bars. Since the gas and stars do not have the same response to the tidal force, the gaseous and stellar bars generally have different phases.

This phase difference gives rise to torques that can effectively remove angular momentum from the gas. As a result, the gas flows towards the central region and eventually forms a dense gas concentration at the center of the merger remnant.
Denser environments host larger fractions of galaxies morphologically classified as early-types. In addition, galaxies in denser environments are on average redder, less gas-rich, and have lower specific star formation rates.

This strong environment dependence is often interpreted as indicating that galaxies undergo transformations (e.g. late-type $\rightarrow$ early type, star forming $\rightarrow$ passive) once they enter or become part of a denser environment.

Clusters of galaxies are the largest virialized structures in the Universe, with masses of about $10^{14} - 10^{15} \, M_\odot$, and velocity dispersions of about 1000 km s$^{-1}$, and the environments with the highest number densities of galaxies.

Hence, galaxy interactions are frequent, making clusters the ideal environments to look for possible transformation processes.

Roughly speaking, cluster galaxies can be effected by the cluster environment in three different ways:

(i) tidal interactions with other cluster members and with the cluster potential,

(ii) dynamical friction, which causes the galaxy to slowly make its way to the cluster center, and

(iii) interactions with the hot, X-ray emitting intra-cluster medium (ICM) that is known to permeate clusters.
12.5.1 Galaxy Harassment

In a cluster of galaxies, the typical velocity of a galaxy is of the order of the velocity dispersion of the cluster, which is much larger than the internal velocity dispersion of the galaxy.

Encounters are all high-speed in nature - colliding galaxies are impulsively heated (i.e. its internal energy is increased).

As a result, the perturbed galaxy becomes less bound, and more vulnerable to disruptions by further encounters and by tidal interactions with the global cluster potential - harassment.

Early studies of this process focused on how it modifies the structure of elliptical galaxies in clusters and whether it can account for the extended outer envelopes of central cluster galaxies. A particularly interesting result by Aguilar & White (1986) was that the de Vaucouleurs surface brightness profile appears invariant to harrassment, i.e. rapid encounters between galaxies with this structure can cause substantial mass loss but the profile of the final object is still well fit by the \( r^{-1/4} \)-law; there is no tendency for harrassment to produce a tidal cut-off in the density profiles of cluster galaxies.
12.5.2 Galactic Cannibalism

Galaxies in clusters are unlikely to merge since their encounter speed is typically much larger than their internal velocity dispersion.

One important exception: because of dynamical friction, galaxies lose energy and momentum which causes them to ‘sink’ towards the center of the potential well. If the dynamical friction time is sufficiently short, the galaxy will ultimately reach the cluster center, where it will merge with the central galaxy already residing there. **Hence, a central cluster galaxy may accrete satellite galaxies, a process called galactic cannibalism.**

**Roughly speaking, a satellite galaxy will be cannibalized by its central galaxy when dynamical friction can bring it to the center of the potential well within a Hubble time** \(1/H(z)\). From Eq. (12.45) we see that the critical radius for this to occur to a satellite galaxy of mass \(M_S\) in an isothermal host halo of mass \(M_h\) is

\[
    r_{\text{crit}} \sim 0.1r_h \left( \frac{\ln \Lambda}{10} \right)^{1/2} \left( \frac{M_S}{10^{-4}M_h} \right)^{1/2},
\]

where we have used \(V_v/r_h = 10H(z)\). Thus, although it is preferentially the massive satellites that will be cannibalized, in the central regions of a cluster even low mass galaxies can be acccreted by the central galaxy within a Hubble time.
12.5.2 Galactic Cannibalism

Galactic cannibalism has two important effects:

it causes the central galaxy to increase in mass, and

it causes a depletion of massive satellite galaxies, for which the dynamical friction time is the shortest.

Consequently, cannibalism causes an increase of the magnitude difference, $\Delta M_{12}$, between the brightest and second brightest member of a cluster.

If galactic cannibalism is the main mechanism regulating the luminosity of the central galaxy, the magnitude gap $\Delta M_{12}$ can thus be used as a measure for the dynamical age of the cluster: older systems will have a larger magnitude gap.
12.5.3 Ram-Pressure Stripping

When a galaxy moves through the intracluster medium (ICM), its gas component experiences a ram pressure, just like one feels wind drag when cycling. As first discussed by Gunn & Gott (1972), if the ram pressure is sufficiently strong, it may strip the gas initially associated with the galaxy.

Consider a disk galaxy of radius $R_d$ moving through an ICM of density $\rho_{\text{ICM}}$ with velocity $V$. For simplicity, assume that the velocity vector of the disk galaxy is pointing perpendicular to the disk, so that it experiences a face-on wind. The amount of ICM material swept per unit time by the disk is $\pi R_d^2 \rho_{\text{ICM}} V$.

If we assume that the wind is stopped by the interstellar medium (ISM) of the disk, then the momentum transferred to the disk per unit time is $\pi R_d^2 \rho_{\text{ICM}} V^2$, corresponding to a ram pressure $P_{\text{ram}} = \rho_{\text{ICM}} V^2$.

**If this pressure exceeds the force per unit area that binds the ISM to the disk, the gas will be stripped.**

To estimate the binding force, assume that the mean surface density of interstellar gas is $\Sigma_{\text{ISM}}$ and that the mean mass density (usually dominated by stars) of the disk is $\Sigma_{\ast}$. The gravitational field of the disk is approximately $2\pi G \Sigma$ in the disk and so the gravitational force per unit area on the interstellar gas is $2\pi G \Sigma \Sigma_{\text{ISM}}$. Hence, we expect that stripping occurs if

$$\rho_{\text{ICM}} > \frac{2\pi G \Sigma_{\ast} \Sigma_{\text{ISM}}}{V^2}.$$  \hfill (12.78)
12.5.4 Strangulation

The ram-pressure stripping discussed above may strip a galaxy of its entire cold gas reservoir, causing an abrupt quenching of its star formation.

However, if only the outer parts of a galaxy’s gas disk are stripped, star formation may continue until all fuel is exhausted. In normal field spirals, the gas consumption timescale is typically only a few Gyrs. Their lifetimes can be significantly extended only if they continue to accrete fresh gas.

Galaxies are believed to be surrounded not only by halos of dark matter, but also by substantial amounts of hot/warm gas. This consists:

in part of gas that is just falling onto the system for the first time,

in part of gas which has already fallen in and been shocked to high temperature, but has not yet cooled, and

in part of gas that has been reheated and expelled from the galaxy by feedback processes.

The infall and cooling of this material can replenish the ISM as it is consumed by star formation, allowing the galaxy to continue forming stars over long timescales. Hence, this extended gas component can be regarded as a reservoir of fuel for future star formation.

Since this gas reservoir is only relatively loosely bound to the galaxy, it is easily removed, either by tides or by ram-pressure when the galaxy falls into a larger system such as a cluster. Thereafter the galaxy loses its supply of new gas and gradually exhausts the remaining fuel in its disk. This process is called strangulation, and produces a gradual decline of a galaxy’s star formation rate on the gas consumption timescale.