The Physics of Galaxy Clusters

Prof. Dr. Christoph Pfrommer
Leibniz-Institute for Astrophysics Potsdam (AIP)
University of Potsdam

Contents

Chapter 1

Overview and Background

1.1 Why are Galaxy Clusters Interesting?

1.1.1 General Syllabus

Galaxy clusters are fascinating objects as they lie at the *cross-roads of astrophysics and cosmology,* which makes them unique tools for answering questions that reach into both areas. Let me explain to you why this is by lying out the general syllabus of the four chapters.

1. **Overview and background.** We will become familiar with the various appearances of clusters in a number of different observational windows; each of which allows us to probe physics that is either specific to a waveband or probes a common feature of a given cluster. We will encounter a vast range of (length and time) scales as well as physical processes. To master this problem, we need to introduce the powerful technique of *order of magnitude estimates,* a very useful tool for contemporary research in astrophysics that we will frequently use in the course of these lectures.
2. **Evolution of the dark component.** Most of the matter in the universe is in form of dark matter that interacts primarily gravitationally with baryonic matter (that is described by the standard model of particle physics). Galaxy clusters are the largest gravitationally collapsed objects. Hence, they represent a fair sample of the universe and are also dominated by dark matter. We will first learn how (the dark component of) a cluster forms and grows. This knowledge is the basis for using clusters as cosmological tools.
3. **Evolution of the baryonic component.** We will then encounter the rich and interesting astrophysics that governs the assembly and evolution of baryons in clusters. This chapter starts with basic thermodynamics and conservation laws and ends with plasma and high-energy astrophysics.
4. **Cluster astrophysics and cosmology.** We will see how we can take advantage of these physical processes to observe clusters across wavelengths and deepen our understanding of the underlying fundamental astro- and plasma physics, and introduce the concept of *cluster cosmology*.

I will briefly introduce and overview the main motivation to study clusters s and present the concepts that we will develop in these lectures. I hope that those concepts will be of good use for you in your further career as (astro-)physicist (even if you won't be studying galaxy clusters).

1.1.2 Clusters as Tools for Cosmology

- Galaxy clusters are the largest and most massive gravitationally bound structures known in the Universe. As such, they represent the latest stage of the structure formation, presently assembling through mergers of smaller groups of galaxies and gas accretion. Hence they provide us with the opportunity to study an "ecosystem" – a volume that is a high-density microcosm of the rest of the Universe.

- At the same time, clusters are extremely rare events, forming at sites of constructive interference of long waves in the primordial density fluctuations. Hence, they are very sensitive tracers of the growth of structure in the universe and the cosmological parameters governing it, which puts them into focus of constraining the properties of Dark Energy or to test whether our understanding of gravity is complete.

- What are the most basic questions one could ask about clusters? And what are the concepts that we will develop to answer those?

  1. **When and where do clusters form?** We will learn how structures grow from tiny perturbations to non-linear structures and how we describe these by appropriate statistics, in particular correlation functions and power spectra.
  2. **How do clusters form?** We will develop a simplified model of the spherical collapse of a perturbation into a (dark matter) halo that defines all characteristic halo parameters.
  3. **How many clusters are there?** We will study the statistics of collapsed halos giving rise to the Press-Schechter mass function.
  4. **What is the structure of a cluster?** We will analyze halo density profiles and the concept of virial masses.

- These concepts are presented in Chapter 2 and enable us to build the dark matter backbone of clusters – by understanding the structure of the gravitational potential of an individual cluster as well as understanding the distribution of the cluster population as a whole. The next chapter asks what happens if we fill baryons into clusters and addresses the beautiful physics associated with this. We will come back to the exciting topic of *cluster cosmology* at the very end of the lecture in Section 4.6 and discuss state-of-the-art methods and current research to improve our knowledge on cosmology and the nature of dark matter using observations of galaxy clusters.

1.1.3 Clusters as Laboratories for Astrophysics

- Galaxy clusters are excellent laboratories for studying the rich astrophysics of dark matter and baryons. In particular, they allow us to study plasma and high-energy astrophysics under conditions that are unique and not reproducible anywhere else, especially not in Earth-bound laboratories.

- In Chapter 3, we will "assemble" clusters by starting simple and consecutively adding more complicated physics. First (in Section 3.1), we consider only non-radiative physics: what is a stable thermodynamic configuration of the gas in a stratified atmosphere and how do perturbations propagate? Clusters are dynamically evolving systems that are shaken by merging groups and gas accretions, which has two consequences: 1. perturbations of an otherwise stable atmosphere can induce vortical motions that feed turbulence and 2. shock waves can be excited that irreversibly change the thermodynamical cluster state, building a new equilibrium configuration that we will characterize. We finally develop powerful cluster scaling relations that link cluster observables to fundamental cluster properties such as its mass. Its evolution differs for different cosmologies and as such, allows to solve for cosmological parameters. However, the scaling relations are intertwined with cluster physics which causes significant modifications of the scaling laws. While this enables us to infer details about complex baryonic processes in clusters, it also complicates the inference of cosmological parameters.

- In a second step (Section 3.2), we explore radiative gas physics, namely radiative cooling, star formation, and energy feedback by exploding stars in galaxies (supernovae) and accreting supermassive black holes that are thought to exist at the center of every galaxy. Detailed physical processes close to the Schwarzschild horizon are able to launch relativistic outflows that carry enormous momentum and energy to macroscopic scales in clusters, thereby modifying its thermodynamical structure in an important way. How this exactly works is currently under intense investigation. We will learn about the strength and weaknesses of various promising suggestions, some of which include transport processes of gas (turbulence, conduction).

- Last but certainly not least (Section 3.3), we will discover the physics of non-thermal processes such as magnetic fields and relativistic particle populations in galaxy clusters. We discuss proposals for the origin and transport of cluster magnetic fields and how magnetic fields modify hydrodynamic turbulence. Moreover, I present basic concepts how particles get accelerated to relativistic energies and how they interfere with the thermal plasma of a cluster, an exciting cutting-edge topic in cluster research. Those can be directly observed in form of giant radio relics and halos that constitute a puzzling glow of the outer fringes as well as entire galaxy clusters and enable us to watch powerful shocks and plasma physics at work.

- We will connect these more basic astrophysical concepts and processes to exciting current-day research on the intracluster medium in Chapter 4. We will illuminate the interesting cluster astrophysics that can be pursued with high-angular resolution observations of the intracluster medium in the X-ray band in Section 4.3. This includes the topic of hydrostatic equilibrium masses and biases, which cluster populations exist and how these evolve with redshift, and what we can learn about plasma physics by studying intracluster medium turbulence, merger shocks, and cold fronts. In Section 4.4, I will detail the physics of the famous Sunyaev-Zel'dovich effect, discuss why this effect is uniquely suited to characterize the intracluster medium by means of scaling relations and power spectra and finally touch upon a new window into cluster astrophysics by means of millimeter wavelengths observations with the Sunyaev-Zel'dovich effect. Finally, I will elucidate the non-thermal physics probed by the radio window to galaxy clusters in Section 4.5, which enables us to study particle acceleration at shocks and magnetized turbulence as well as magnetism in galaxy clusters.

1.1.4 Clusters as Laboratories for Galaxy Evolution

- Observing galaxies in the optical wavelength regime and realizing that they like to cluster together was the first window to galaxy clusters (not surprisingly as the name suggests). Even today, there are many interesting questions about galaxy formation that take advantage of the increased density of a cluster environment, which accelerates the formation time of galaxies and enables us look at ancient relics of galaxy formation.

- However, a cluster does not simply represent a museum that conserves galaxy properties and supports a passive aging of the stellar populations within them. Instead, the high galaxy density in a cluster environment transforms galaxy populations via different evolutionary processes, including tidal stripping and shocks, dynamical friction, and ram pressure stripping, which we will all study in detail in Section 4.1.

- In that Section, we will learn how the virial theorem can be used to weight a galaxy cluster. It is interesting to compare the masses obtained from this method to another approach that assumes the equation of hydrostatic equilibrium of the cluster gas.

- Finally, galaxy clusters literally act as magnifying glasses for very distant galaxies that happen to be in projection behind a massive cluster. The processes of gravitational lensing not only magnifies the galaxies' surface brightness but also increases their solid angle on the sky. We will review the theory of gravitational lensing in Section 4.2, derive the lens equation, and discuss the lensing potential. After X-ray based hydrostatic equilibrium masses and applying the virial theorem to cluster galaxies, this represents a third, independent method for weighting clusters.

1.1.5 General Remarks

This is a course in the Masters program. The lectures aim at students who

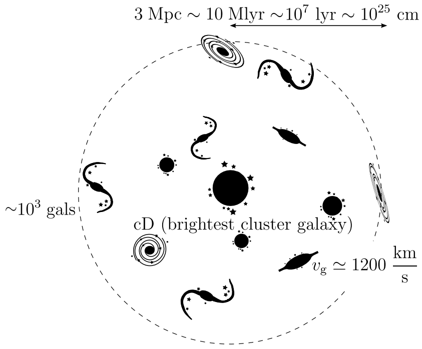
- wish to extend and deepen their understanding of theoretical physics;
- are interested in astronomy and astrophysics; or
- (intend to) carry out a masters thesis or Ph.D. dissertation on an astronomical or astrophysical subject.

I assume basic knowledge of *Cosmology*. While this is not absolutely necessary to follow most of the lectures, I recommend working through the *Cosmology* lecture notes by Prof. Bartelmann. You can download a revised version freely from my home page if you want to refresh your memory.

I use Gaussian cgs units throughout these lecture notes, denote $\boldsymbol{ab}$ as the dyadic product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, and define $\boldsymbol{:}$ as the double-dot product of two rank-2 tensors.

1.2 What Characterizes a Galaxy Cluster?

A galaxy cluster looks different, depending on how you look at it. Using observations at various wavelengths, we get a wealth of diverse insights into physics. This section is meant to provide a general overview of the various appearances of clusters in a number of different observational windows. Rather than lying out the most complete and accurate description, I try to convene the basic concepts by using the powerful technique of *order of magnitude estimates*, and leave the detailed discussion of the physics to later chapters.



Schematic drawing of galaxies in a galaxy cluster.

1.2.1 Optical Window

- In a rich galaxy cluster, there are $\sim 10^3$ galaxies that have to good approximation a Gaussian velocity distribution with a dispersion $\sigma_v = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} \approx 1200\,\mathrm{km\ s^{-1}}$. The typical radius of such a cluster is
$$r_\mathrm{cl} \approx 3\ \mathrm{Mpc} \approx 10^7\ \mathrm{lyr} \approx 10^{25}\ \mathrm{cm}. \tag{1.1}$$
This defines a dynamical cluster timescale, $t \approx r_\mathrm{cl}/\sigma_v \approx 2$ Gyr.

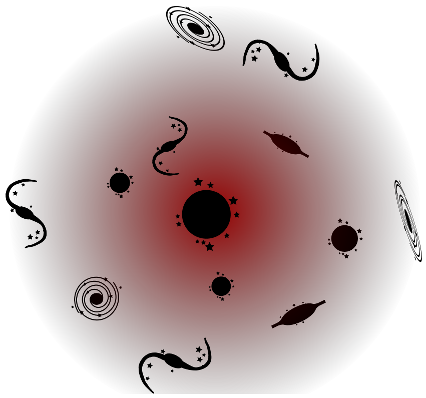
- Assuming that the cluster is a closed system in dynamical equilibrium, the virial theorem relates the kinetic energy, E_kin, of a galaxy of mass M_gal to its potential energy, E_pot,
$$2E_\mathrm{kin} + E_\mathrm{pot} = 0, \tag{1.2}$$
$$M_\mathrm{gal}\sigma_v^2 - \frac{G M_\mathrm{cl} M_\mathrm{gal}}{r_\mathrm{cl}} = 0, \tag{1.3}$$
where G is Newtons gravitational constant. Solving for the gravitating mass of a cluster, M_cl, we get
$$\begin{aligned} M_\mathrm{cl} = \frac{r_\mathrm{cl}\sigma_v^2}{G} &\approx \frac{10^{25}\,\mathrm{cm}\,1.4\times10^{16}\,\mathrm{cm^2 s^{-2}}}{7\times10^{-8}\,\mathrm{erg\ cm\ g^{-2}}} \\ &\approx 2\times10^{48}\,\mathrm{g} \approx 10^{15}\,\mathrm{M_\odot} \end{aligned} \tag{1.4}$$
Note that M_cl sources the high velocity dispersion of galaxies. A typical mass range for clusters is $(10^{14} \ldots 10^{15})\,\mathrm{M_\odot}$.

- However, by adding up the all the luminous stellar mass within the galaxies, we only get
$$M_* \approx \frac{1}{50} M_\mathrm{cl}. \tag{1.5}$$
This discrepancy of the gravitating and luminous mass in galaxy clusters was already noted by Fritz Zwicky in the 1930s and led him to postulate the existence of dark matter more than 80 years ago! To be precise, back then the "dark matter" could have been baryonic in form of compact objects (such as planets) or in form of diffuse gas.

1.2.2 X-ray Regime

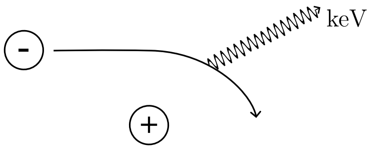
- With the onset of X-ray astronomy in the 1970s, it was discovered, that galaxy clusters are among the brightest X-ray emitting sources. Improved angular resolution demonstrated that the X-rays were not emitted by individual point sources but instead the entire galaxy cluster is glowing in X-rays, filling in the volume in between the galaxies. What emission process could produce X-rays? There are three possibilities:

  1. bremsstrahlung emission of hot thermal electrons,
  2. line emission from recombination of atoms, or
  3. inverse Compton emission: if relativistic electrons interact with low-energy photons (such as those from the cosmic microwave background), they can cool by upscattering these photons into the X-ray regime. This would typically produce power-law spectra that are imprinted by the power-law spectra of the underlying non-thermal relativistic electrons.



The hot and dilute intracluster medium fills the space in between the galaxies and emits X-rays.

- The observed X-ray spectrum instead shows a flat spectrum with an exponential decline that is characteristic of thermal bremsstrahlung emission. Additionally, there are lines imprinted on the spectrum. The bremsstrahlung emissivity scales as $j_X \propto n_e n_i \sqrt{T_e}$, where T_e, n_e, and n_i are the electron temperature, density and the ion density, respectively. The physical principle of bremsstrahlung emission is shown below:

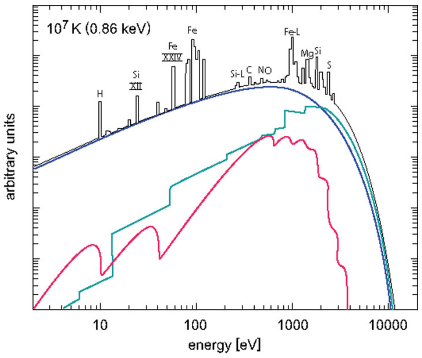


The amount of X-rays and the location of the exponential break (as well as the location of the individual lines) enable to characterize the properties of the emitting gas,

$$n \approx (10^{-4} \dots 10^{-3})\,\mathrm{cm}^{-3}, \tag{1.6}$$

$$T \approx (10^{7} \dots 10^{8})\,\mathrm{K}, \tag{1.7}$$

i.e., a hot, dilute, and thermal gas (as inferred from the exponential shape of the bremsstrahlung spectrum).



X-ray spectrum for a plasma with temperature 10^7 K with solar abundance. The different radiative processes contributing to the emission are bremsstrahlung continuum (blue), recombination lines (green) and 2-photon radiation (red). Credits: Boehringer and Werner 2009.

- We usually talk about the temperature of a gas in terms of particle energies,

$$k_B T \approx (1 \dots 10)\,\mathrm{keV} = (10^{3} \dots 10^{4})\,\mathrm{eV}, \tag{1.8}$$

where k_B is the Boltzmann factor. At these temperatures, most of the elements are fully ionized, except for highly-ionized iron, e.g., hydrogen-like iron which is an iron nucleus with one bound electron, Fe XXVI. The transition energy of such highly ionized iron is

$$\begin{aligned}\text{Fe XXV}: &\quad h\nu \approx Z(Z-1)\,\mathrm{Ry} = 26\times25\times13.6\,\mathrm{eV} \approx 8.8\,\mathrm{keV}\\ \text{Fe XXVI}: &\quad h\nu \approx Z^2\,\mathrm{Ry} = 26^2\times13.6\,\mathrm{eV} \approx 9\,\mathrm{keV}, \qquad (1.9)\end{aligned}$$

i.e., the higher the temperature, the higher the ionization state.

- Assuming that this hot gas with energy E_{th} is in hydrostatic equilibrium with the cluster potential, we have

$$E_{\mathrm{th}} = E_{\mathrm{pot}}, \qquad (1.10)$$

$$\frac{3}{2}k_{\mathrm{B}}T = \mu m_{\mathrm{p}}\frac{GM_{\mathrm{cl}}}{r_{\mathrm{cl}}}, \qquad (1.11)$$

where m_{p} is the proton mass and the mean molecular weight of primordial gas is given by $\mu = 4/(5X+3) \approx 0.588$ for a primordial hydrogen fraction $X = 0.24$ (see Appendix A.1). Solving for the gravitating mass of a cluster with $k_{\mathrm{B}}T = 6$ keV yields

$$\begin{aligned}M_{\mathrm{cl}} &= \frac{3}{2}\frac{k_{\mathrm{B}}Tr_{\mathrm{cl}}}{\mu m_{\mathrm{p}}G} \approx \frac{1.5\times10^{-8}\,\mathrm{erg}\,10^{25}\,\mathrm{cm}}{0.6\times1.7\times10^{-24}\,\mathrm{g}\,7\times10^{-8}\,\mathrm{erg\,cm\,g^{-2}}}\\ &\approx 2\times10^{48}\,\mathrm{g} \approx 10^{15}\,\mathrm{M}_{\odot}. \qquad (1.12)\end{aligned}$$

- Resolved X-ray imaging of a galaxy cluster produces an X-ray surface brightness map. Deprojection enables us to back out the mass density profile. Integrating that over the cluster volume yields the total gas mass,

$$M_{\mathrm{gas}} \approx \frac{1}{7}M_{\mathrm{cl}}. \qquad (1.13)$$

Hence, we found some of the matter that was "dark" in the optical by looking at a different waveband. The rest cannot be directly seen in any other waveband (at least no significant amounts). It only can be indirectly inferred through its gravitational interaction. We call this "dark matter", which reflects our ignorance of the composition. It dominates the total cluster mass and is mostly responsible for the gravitational cluster potential.

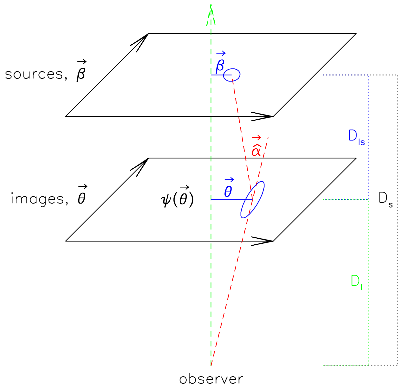
- We can now summarize an inventory of cluster mass

$$\begin{aligned}M_{*} &\approx \ \ 2\% : &&\text{stars in galaxies,}\\ M_{\mathrm{gas}} &\approx 13\% : &&\text{hot gas } (1-10\,\mathrm{keV}), \qquad (1.14)\\ M_{\mathrm{dm}} &\approx 85\% : &&\text{dark matter.}\end{aligned}$$

The value of the baryon fraction in a cluster of $f_{\mathrm{b,clus}} \approx 0.15$ is somewhat smaller than the cosmic mean of $f_{\mathrm{b,clus}} \approx 0.166$. This points to interesting physics, including non-gravitational energy input from supernovae and super-massive black holes. Interestingly, $f_{\mathrm{b,clus}}$ declines toward less massive clusters in which those feedback processes have a comparably larger impact because of the shallower cluster potential.

1.2.3 Gravitational Lensing

- Galaxy clusters or galaxies act as gravitational lenses for galaxies behind them. The processes of gravitational lensing not only magnifies the surface brightness of the source galaxies but also increases their solid angle on the sky. According to general relativity, light travels on geodesics (straightest possible lines) through curved space time. Mass acts as a source of gravity, curving space time at the location of a lensing galaxy cluster or galaxy and causing the light rays to be deflected by the gravitational potential of the lensing object. This causes a single galaxy to be mapped onto multiple images (or even a so-called Einstein ring, provided that we have a very symmetric configuration and a point-like source). We define the angular diameter distance to the light-deflecting cluster or galaxy, D_{l}, the distance to the source galaxy, D_{s}, and the angular diameter distance between deflector and source, D_{ls}. The drawing on the right explains the geometry of a lensing system.



Geometry of a gravitational lensing system.

- Later on in the lectures, we will derive the Einstein radius θ_{E}. We state the result and insert some values to get an idea about the involved angular scales.

$$\begin{aligned}\theta_{\mathrm{E}} &= \left[\frac{4GM(\theta_{\mathrm{E}})}{c^2}\frac{D_{\mathrm{ls}}}{D_{\mathrm{l}}D_{\mathrm{s}}}\right]^{1/2} && (1.15)\\ &\approx 3'' \left(\frac{M}{10^{12}\,M_{\odot}}\right)^{1/2}\left(\frac{D}{1\,\mathrm{Gpc}}\right)^{-1/2} \quad \text{(galaxy lensing)} && (1.16)\\ &\approx 30'' \left(\frac{M}{10^{14}\,M_{\odot}}\right)^{1/2}\left(\frac{D}{1\,\mathrm{Gpc}}\right)^{-1/2} \quad \text{(cluster lensing).} && (1.17)\end{aligned}$$

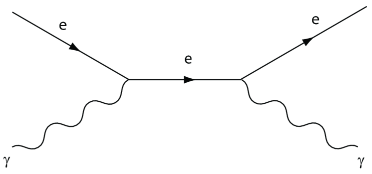
  Here, $D = D_{\mathrm{l}}D_{\mathrm{s}}/D_{\mathrm{ls}}$ is the lensing efficiency distance. In the case of galaxy lensing, the approximation of a point lens is justified whereas for cluster lensing, the size of the lens is much larger than the size of the source. This makes a detailed mass modeling necessary since only a fraction of the cluster mass is contained within the Einstein radius and will contribute to the lensing potential. The inferred values of $\theta_{\mathrm{E}} \approx 30''$ correspond to angular scales of observed giant (tangential) arcs.

- We distinguish two types of lensing:

1. **Strong lensing** is sensitive to the projected mass within θ_E and leads to radial arcs that are clearly visible in optical images. In this regime, a source can be imaged onto multiple different images.
2. **Weak lensing** causes weaker distortions of a galaxy image in the tangential direction than cannot be detected on a individual basis because the effect is very small. We need to assume that orientation of the of neighboring galaxies is random and average over an aperture to detect a weak shear signal that is induced by the gravitational tidal field of the cluster lens.

1.2.4 Sunyaev-Zel'dovich Effect

- The universe is filled with 2.75 K photons of the cosmic microwave background (CMB), which is radiation left behind from the early universe when hydrogen recombined at a redshift of $z \approx 1100$. If such a "cold" photon passes through a galaxy cluster that is filled with "hot" electrons there is the chance that this photon experiences Compton scattering off an electron:

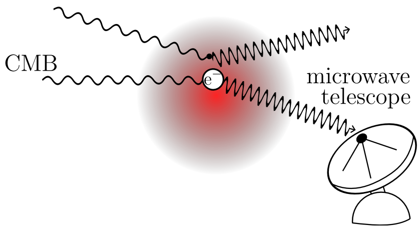
  

  This elastic scattering event conserves the number of CMB photons. However, during this reaction there is a mean energy transfer from the "hot" electron to the "cold" photon (which is the reason why this is called *inverse* Compton scattering). This causes a unique distortion of the CMB spectrum, a decrement in thermodynamic temperature at frequencies below $\nu_0 \approx 220$ GHz, and an excess above. As a result, galaxy clusters appear as holes in the CMB sky at $\nu < \nu_0$ and as extended sources above. This Sunyaev-Zel'dovich (SZ) effect provides a complementary method for detecting and characterizing galaxy clusters.

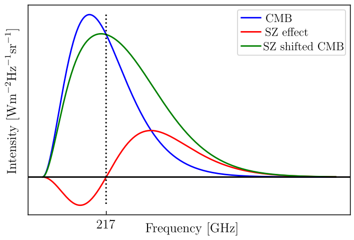
- How many CMB photons experience inverse Compton scattering on passing through a cluster? To answer this, we compute the optical depth,

$$\tau = \int_0^L n_\mathrm{e}\sigma_\mathrm{T}\mathrm{d}l \approx n_\mathrm{e}\sigma_\mathrm{T}L, \qquad (1.18)$$

  where L is the effective path length through the hot intracluster



Visualizing the Sunyaev-Zel'dovich (SZ) effect in a galaxy cluster. The underlying physics are inverse Compton scatterings of CMB photons on hot thermal electrons of the intracluster medium.



Spectral distortion due to the SZ effect. The black-body spectrum of the CMB (blue) is shifted as a result of the energy transfer from hot electrons to cold CMB photons via inverse Compton scattering (green). The spectral distortion due to the SZ effect (red) shows a reduction of flux in the CMB spectrum below $\nu < \nu_0 = 217$ GHz and excess above. Here we exaggerate the SZ amplitude and adopt $y = 6 \times 10^{-2}$, following equation (4.184).

medium (ICM) and σ_{T} is the Thomson cross section,

$$\sigma_{\mathrm{T}} \approx 2\pi r_0^2 = 2\pi \left(\frac{e^2}{m_{\mathrm{e}}c^2}\right)^2 \approx 6 \left[\frac{\left(4.8 \times 10^{-10}\right)^2}{10^{-27}\,10^{21}}\right]^2 \mathrm{cm}^2$$
$$\approx 6 \left(3 \times 10^{-13}\right)^2 \mathrm{cm}^2 \approx 6 \times 10^{-25}\,\mathrm{cm}^2, \qquad (1.19)$$

where r_0 is the classical electron radius. Hence, we obtain an optical depth

$$\tau = n_{\mathrm{e}}\sigma_{\mathrm{T}}L \approx 10^{-4}\,\mathrm{cm}^{-3}\, 6 \times 10^{-25}\,\mathrm{cm}^2\, 10^{25}\,\mathrm{cm} \approx 6 \times 10^{-4} \ll 1. \qquad (1.20)$$

This means that on average only one photon in 2000 experiences a scattering event.

- What is the amplitude of the SZ effect? To answer this, we integrate the typical energy gain experienced by a photon in a Compton interaction ($k_{\mathrm{B}}T_{\mathrm{e}}/m_{\mathrm{e}}c^2$) times the differential scattering probability of a photon ($\mathrm{d}\tau = n_{\mathrm{e}}\sigma_{\mathrm{T}}\mathrm{d}l$) over the photon path length, D. This is the exact definition of the Compton-y parameter,

$$y = \int_0^D \frac{k_{\mathrm{B}}T_{\mathrm{e}}}{m_{\mathrm{e}}c^2} n_{\mathrm{e}}\sigma_{\mathrm{T}}\mathrm{d}l \approx 10^{-2} \times 6 \times 10^{-4} = 6 \times 10^{-6}. \qquad (1.21)$$

Here, we adopted a line-of-sight averaged temperature of our massive ($10^{15}\,\mathrm{M}_{\odot}$) cluster with $k_{\mathrm{B}}T_{\mathrm{e}} \approx 6$ keV. As we can see, the SZ signal is proportional to the integrated electron pressure ($P_{\mathrm{e}} = k_{\mathrm{B}}T_{\mathrm{e}}n_{\mathrm{e}}$), so the hot gas of the galaxy clusters dominates the effect. This implies that only the path length L through the cluster contributes significantly to the integral that formally extends over the light travel distance, D, from us to the release of the CMB photons. The resulting small value for y implies a small change in intensity that is challenging to detect.

- The thermal SZ effect directly observes the solid-angle integrated Compton-y parameter over the cluster face

$$Y = \int_{\Omega} \mathrm{d}\Omega\, y = \frac{1}{D_{\mathrm{ang}}^2} \int_A \mathrm{d}^2 r\, y, \qquad (1.22)$$

where D_{ang} denotes the angular diameter distance and $D_{\mathrm{ang}}^2\Omega = A$. The quantity Y is a measure of the cluster's global gas heat-energy content, a volume-average of the thermal gas pressure,

$$Y_{\mathrm{sph}} = Y D_{\mathrm{ang}}^2 = \frac{\sigma_{\mathrm{T}}}{m_{\mathrm{e}}c^2} \int_0^{r_{\mathrm{cl}}} P_{\mathrm{e}}\mathrm{d}V \propto E_{\mathrm{th}}(< r_{\mathrm{cl}}), \qquad (1.23)$$

and this is related to gravitational energy through the virial relation.

- The SZ surface brightness is independent of the redshift a specific cluster is at. This can be easily understood by the following consideration: CMB photon are continuously redshifted from the surface-of-last-scattering to us as the Universe expands. Irrespective of when exactly the inverse-Compton scattering event happened, that energized the photon by a fixed amount of energy, the CMB photons still experience continuous redshifting. This independence on redshift makes the SZ effect a prime candidate for cluster cosmology since it allows for an almost flat selection of cluster at a given mass with redshift. Note that this is quite different for X-ray selected clusters that suffer from the strong flux dimming as the square of the distance.

1.2.5 Synthesis of Observational Windows

Different cluster observables have different strengths and weaknesses. Hence the various windows to clusters are complementary: combining observations of different wavelengths is essential to learn more about clusters. Here is a short (and possibly incomplete) summary about the most important (dis-)advantages of different cluster observables:

- **X-ray window.**
  - It is well-suited for observing cluster centers since $j_\mathrm{X} \propto n^2$ which emphasizes dense cluster gas.
  - It allows for high-resolution (arcsec) observations (because of the use of Bragg reflection for focusing X-rays).
  - It is difficult to observe X-ray clusters at large distances due to flux dimming, $F_\mathrm{X} \propto L_\mathrm{X}/D^2$.
  - If the X-rays are emitted by an inhomogeneous medium, data analyses need to be aware of the bias of the inferred density. In this case, $\bar{n}^2 = \langle n \rangle^2$ is biased high by the clumping factor $C = \langle n^2 \rangle / \langle n \rangle^2$ which is in general not known.
- **Sunyaev-Zel'dovich effect.**
  - It is an excellent tool for studying cluster outskirts as $y \propto P_\mathrm{e} = n_\mathrm{e} k_\mathrm{B} T_\mathrm{e}$ and declines less steeply with radius in comparison to the X-ray emission.
  - It is well adapted to detect and observe clusters at large distances because y is independent of redshift (since CMB photons experience continuous redshifting from the surface-of-last-scattering to us as the Universe expands).
  - The comparably small signal-to-noise ratio makes it difficult to detect small clusters and groups.

- Current resolutions (typically arcmin) preclude the study of detailed cluster physics.

- **Optical window.**
  - Galaxies are collisionless tracers of the gravitational potential and its dynamical state. However, the method needs many galaxies to sample the velocity distribution well enough (which is only possible for large clusters).
  - Gravitational lensing is an invaluable tool for directly probing the total cluster mass which is dominated by dark matter. However, projection of structures along the line-of-sight needs to be accounted for carefully.

1.2.6 Relation to the Average Universe

- How does a galaxy cluster relate to the average Universe around us? The critical density of the universe today is

$$\rho_{\mathrm{cr},0} = \frac{3H_0^2}{8\pi G} \approx 10^{-29}\ \mathrm{g\ cm^{-3}}, \tag{1.24}$$

  where $H_0 \approx 70\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$ is the Hubble constant that defines the recession velocity of local galaxies, which are at distance d, from us according to Hubble's law, $v = H_0 d$. Using the critical density, we can define the density parameters of total matter, Ω_m, and baryons, Ω_b,

$$\Omega_{\mathrm{m},0} \equiv \frac{\bar{\rho}_{\mathrm{m},0}}{\rho_{\mathrm{cr},0}} \approx 0.27, \quad \text{and} \quad \Omega_{\mathrm{b},0} \equiv \frac{\bar{\rho}_{\mathrm{b},0}}{\rho_{\mathrm{cr},0}} \approx 0.045. \tag{1.25}$$

- Hence the mean matter density of the Universe is

$$\bar{\rho}_{\mathrm{m},0} \approx 4 \times 10^{10}\,\mathrm{M_\odot\ Mpc^{-3}} \approx 10^{9}\,\mathrm{M_\odot\ Mlyr^{-3}}. \tag{1.26}$$

  Compare this to typical cluster masses $M_\mathrm{cl} \sim 10^{15}\,\mathrm{M_\odot}$. In order to form clusters, you need large chunks of volume that contain $10^{15}\,\mathrm{M_\odot}$. As we will learn during the lectures, only less than 1% of cosmic matter forms an aggregation that makes a large clusters. Hence clusters are extremely rare!

- **Collapse of a cluster.** Typically, we find $\bar{\rho}_\mathrm{cl} \sim 10^3 \bar{\rho}_{\mathrm{m},0}$, hence cluster collapse roughly by a factor of 10 in radius. Let's check whether this is consistent with what we have already learned. The mean baryon density of the universe at the present time is

$$\bar{n}_{\mathrm{b},0} = \frac{\rho_\mathrm{cr}\Omega_\mathrm{b}}{\mu m_\mathrm{p}} \approx \frac{10^{-29}\,\mathrm{g\,cm^{-3}}\,0.045}{0.6 \times 1.7 \times 10^{-24}\,\mathrm{g}} \approx 4 \times 10^{-7}\ \mathrm{cm^{-3}}. \tag{1.27}$$

Upon collapsing this chunk of baryons by a radial factor of 10, we obtain a mean cluster density of

$$\bar{n}_{\rm cl} \approx 4 \times 10^{-4}\ \mathrm{cm}^{-3}, \qquad (1.28)$$

which is perfectly consistent with the densities inferred by X-ray observations of clusters.

Chapter 2

Evolution of the Dark Component

2.1 The Growth of Perturbations

2.1.1 Newtonian Equations

- There are pronounced structures in the universe on scales from galaxies to galaxy clusters and cosmic large-scale filaments. While filaments and the voids they surround can reach sizes of $\sim 50\,\mathrm{Mpc}$, they are still small compared to the Hubble radius. In this chapter, we only describe the concepts of the basic theory for structure growth in the expanding universe for the matter-dominated epoch, i.e., we only consider the conservation laws for non-relativistic fluids. A detailed derivation and more complete discussion can be found in Section 2.1 of the *Cosmology* lecture notes by Prof. Bartelmann.

- Strictly, this theory should be worked out in the framework of general relativity, which is a complicated exercise. With the inhomogeneities being "small", i.e. much smaller than the typical scale of the universe, we can neglect effects of curvature and the finite speed of information propagation and work within the framework of Newtonian dynamics. We will see that structure grows from small-amplitude seed fluctuations through gravitational instability and determine the rate of growth.

- We describe inhomogeneities in a cosmic fluid which contains at least radiation, dark matter, and baryonic matter and which moves according to Newtonian gravity.

- We begin with the continuity equation, which formulates mass conservation,
$$\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}_r \cdot (\rho\boldsymbol{v}) = 0\ , \tag{2.1}$$
where $\rho(t,\boldsymbol{r})$ and $\boldsymbol{v}(t,\boldsymbol{r})$ are the density and velocity of the cosmic fluid at position $\boldsymbol{r}$ and time t.

- The second equation is Euler's equation which formulates the conservation of momentum,
$$\frac{\partial\boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\boldsymbol{\nabla}_r)\boldsymbol{v} = -\frac{\boldsymbol{\nabla}_r P}{\rho} - \boldsymbol{\nabla}_r\Phi\ . \tag{2.2}$$
The terms on the right-hand side represent the pressure-gradient and gravitational forces.

- The Newtonian gravitational potential Φ satisfies the Poisson equation
$$\nabla_r^2\Phi = 4\pi G\rho\ . \tag{2.3}$$

2.1.2 Density Perturbations

- The next steps consist in decomposing density and velocity fields into their homogeneous background values $\bar{\rho}$ and $\bar{\boldsymbol{v}}$ and small perturbations $\delta\rho$ and $\delta\boldsymbol{v}$,

$$\rho(t,\boldsymbol{r}) = \bar{\rho}(t) + \delta\rho(t,\boldsymbol{r}) \ , \quad \boldsymbol{v}(t,\boldsymbol{r}) = \bar{\boldsymbol{v}}(t) + \delta\boldsymbol{v}(t,\boldsymbol{r}) \ . \tag{2.4}$$

- The evolution of the homogeneous background quantities are governed by the expansion of the universe. Physical coordinates, $\boldsymbol{r}$, are related to comoving coordinates, $\boldsymbol{x}$, via the equation $\boldsymbol{r} = a\boldsymbol{x}$. Here, $a(t)$ is the cosmic scale factor whose dynamics is governed by Friedmann's equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3} \ , \tag{2.5}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \ . \tag{2.6}$$

Here, K is a constant parameterising the curvature of spatial hypersurfaces and Λ is the cosmological constant. The scale factor is uniquely determined once its value at a fixed time t is chosen. We set $a = 1$ today.

- The *critical* density at scale factor a and today are defined as

$$\rho_{\rm cr}(a) \equiv \frac{3H^2(a)}{8\pi G} \ , \quad \rho_{\rm cr0} \equiv \frac{3H_0^2}{8\pi G} \ , \quad \text{and} \tag{2.7}$$

$$H^2(a) \equiv \left(\frac{\dot{a}}{a}\right)^2 = H_0^2\left[\Omega_{\rm r0}a^{-4} + \Omega_{\rm m0}a^{-3} + \Omega_{\Lambda 0} + \Omega_{\rm K}a^{-2}\right] \tag{2.8}$$

is the Hubble function that derives from Friedmann's equation (2.5) and describes the expansion rate of the universe. Quantities at the present time are denoted with a subscript 0. The density parameters of radiation, matter, the cosmological constant, and curvature are defined by

$$\Omega_{\rm r0} = \frac{\rho_{r0}}{\rho_{\rm cr0}}, \quad \Omega_{\rm m0} = \frac{\rho_{m0}}{\rho_{\rm cr0}}, \quad \Omega_{\Lambda 0} = \frac{\Lambda c^2}{3H_0^2}, \ \text{and} \ \Omega_K = \frac{-Kc^2}{H_0^2}. \tag{2.9}$$

- After this short cosmological detour, we return to the derivation of the density perturbations. To this end we transform the governing equations (2.1) through (2.3) from physical coordinates, $\boldsymbol{r}$, to comoving coordinates, $\boldsymbol{x}$, which are related by $\boldsymbol{r} = a\boldsymbol{x}$. To understand the concept of comoving coordinates, imagine a set of particles that are slightly displaced from a uniform grid. In the Eulerian point of view, the expanding universe causes the grid points to move apart from each other homogeneously and gravitational attraction increases the degree of irregularity. In the comoving frame, the large-scale homogeneous expansion is divided out. This leaves only the dynamics of the gravitational attraction to increase the irregularity. We are hence seeking an equation for the time evolution of the density perturbations in this comoving frame, $\delta\rho(t, \boldsymbol{x})$.

- We define the density contrast,

$$\delta \equiv \frac{\delta\rho}{\bar{\rho}} \,, \tag{2.10}$$

and adopt an equation of state linking the pressure fluctuation to the density fluctuation,

$$\delta p = \delta p(\delta) \equiv c_\mathrm{s}^2 \delta\rho \tag{2.11}$$

with the sound speed c_s.

- We obtain an expression for the velocity,

$$\boldsymbol{v} = \dot{\boldsymbol{r}} = \dot{a}\boldsymbol{x} + a\dot{\boldsymbol{x}} = H\boldsymbol{r} + a\dot{\boldsymbol{x}} = \bar{\boldsymbol{v}} + \delta\boldsymbol{v}, \tag{2.12}$$

where $\bar{\boldsymbol{v}} = H\boldsymbol{r}$ is the Hubble velocity and $\delta\boldsymbol{v} = a\dot{\boldsymbol{x}}$ is the peculiar velocity that deviates from the Hubble flow. Equations (2.1) through (2.3) can be combined to yield a single equation for the density contrast

$$\ddot{\delta} + 2H\dot{\delta} = \left(4\pi G\bar{\rho}\delta + \frac{c_\mathrm{s}^2 \nabla_{\boldsymbol{x}}^2 \delta}{a^2}\right) \,. \tag{2.13}$$

- We can decompose δ into plane waves,

$$\delta(\boldsymbol{x}, t) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \hat{\delta}(\boldsymbol{k}, t) \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \,, \tag{2.14}$$

introducing the time-dependent Fourier amplitudes $\hat{\delta}(\boldsymbol{k}, t)$ and decoupling the time evolution from the spatial dependence. Inserted into Eq. (2.13), this yields

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = \hat{\delta}\left(4\pi G\bar{\rho} - \frac{c_\mathrm{s}^2 k^2}{a^2}\right) \,. \tag{2.15}$$

- On a static background, $H = 0$, and Eq. (2.15) becomes the oscillator equation

$$\ddot{\hat{\delta}} + \omega_0^2 \hat{\delta} = 0 \,, \quad \omega_0 \equiv \sqrt{\frac{c_\mathrm{s}^2 k^2}{a^2} - 4\pi G\bar{\rho}} \,. \tag{2.16}$$

The oscillation frequency is real for sufficiently large comoving wave numbers k,

$$k \geq k_\mathrm{J} \equiv \frac{2\sqrt{\pi G\bar{\rho}}\, a}{c_\mathrm{s}} \,. \tag{2.17}$$

k_{J} defines the comoving Jeans length

$$\lambda_{\mathrm{J}} \equiv \frac{2\pi}{k_{\mathrm{J}}} = \frac{c_{\mathrm{s}}}{a}\sqrt{\frac{\pi}{G\bar{\rho}}} \,. \tag{2.18}$$

Perturbations smaller than the Jeans length oscillate. Others grow or decay. The Jeans length can be heuristically derived by balancing the sound crossing time, $t_{\mathrm{s}} = a\lambda_{\mathrm{J}}/c_{\mathrm{s}} = 2\pi a/(k_{\mathrm{J}}c_{\mathrm{s}})$, with the gravitational free-fall time, $t_{\mathrm{ff}} = \sqrt{\pi/(G\bar{\rho})}$, which yields the desired result of Eq. (2.17).

- We now study the behavior of perturbations on scales much larger than the Jeans length, or in pressure-less fluids. If $\Omega = 1$, we get $\rho_{\mathrm{cr}} = \bar{\rho} = 3H^2/(8\pi G)$ and the perturbation equation reads

  $$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = \frac{3}{2}H^2\hat{\delta} \,. \tag{2.19}$$

  Note that this is only valid for the matter-dominated epoch because we only considered the conservation laws for non-relativistic fluids. In this case, the Hubble rate is given by (see (1.54) of the *Cosmology* lecture notes by Prof. Bartelmann)

  $$a \propto t^{2/3} \quad \Rightarrow \quad \frac{\dot{a}}{a} = H(t) = \frac{2}{3t} \,. \tag{2.20}$$

- The *ansatz* $\hat{\delta}(\boldsymbol{k}, t) \propto t^n$ yields

  $$n^2 + \frac{n}{3} - \frac{2}{3} = 0 \tag{2.21}$$

  hence $n = -1, 2/3$, which translates to

  $$\hat{\delta} \propto \begin{cases} a \,, \\ a^{-3/2} \,. \end{cases} \tag{2.22}$$

  Decaying modes are irrelevant for cosmic structure growth, so $\delta \propto a$ during the matter-dominated era. The phases of the waves determine whether a given cosmological patch develops into an underdense region (i.e., a void) or a galaxy cluster. Constructive interference of the growing modes causes the development of overdensities, which then collapse into galaxies (in the case of small-scale modes) or clusters (for large-scale modes). Destructive interference leads to the growth of voids.

- The sound speed defines the Jeans length, below which perturbations cannot grow, but oscillate. For dark matter consisting of weakly interacting massive particles, for instance, the concept of a sound speed makes no sense because the dark matter behaves like an ensemble of collision-less particles. In that case, one can show that the comoving Jeans length (2.18) is replaced by

$$\lambda_{\mathrm{J}} = \frac{\left\langle v^{-2} \right\rangle^{-1/2}}{a} \sqrt{\frac{\pi}{G\bar{\rho}}} \,, \tag{2.23}$$

where v is the velocity dispersion of the particles. Perturbations in collision-less matter smaller than the Jeans length are thus prevented from growing because their gravity is insufficient for keeping their particles bound.

2.2 Statistics and Non-linear Evolution

2.2.1 Power Spectra

- We have seen before Eq. (2.14) that it is convenient to decompose the density contrast δ into plane waves. We introduce the Fourier transform $\hat{\delta}$ of the density contrast δ as

$$\delta(\boldsymbol{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\,\hat{\delta}(\boldsymbol{k})\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\ , \quad \hat{\delta}(\boldsymbol{k}) = \int \mathrm{d}^3x\,\delta(\boldsymbol{x})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\ . \tag{2.24}$$

- The density contrast is a random field, which must be isotropic and homogeneous in order to comply with the fundamental cosmological assumptions. This means that the statistical properties of δ, e.g. its mean or variance, do not change under rotations and translations.

- By definition, the mean of the density contrast vanishes,

$$\langle\delta\rangle = \left\langle\frac{\rho-\rho_0}{\rho_0}\right\rangle = \frac{\langle\rho\rangle}{\rho_0} - 1 = 0\ . \tag{2.25}$$

  The variance of δ *in Fourier space* defines the power spectrum $P(k)$,

$$\langle\hat{\delta}(\boldsymbol{k})\hat{\delta}^*(\boldsymbol{k}')\rangle \equiv (2\pi)^3P(k)\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}')\ , \tag{2.26}$$

  where δ_{D} is Dirac's delta distribution, which ensures that modes of different wave vector $\boldsymbol{k}$ are uncorrelated in Fourier space in order to ensure homogeneity. The power spectrum cannot depend on the direction of $\boldsymbol{k}$ because of isotropy. The brackets $\langle\ldots\rangle$ represent ensemble averages, i.e., that we average over an infinite number of ensembles of universes, where each ensemble is an independent realization of a random field that has been drawn according to a specified statistics (typically, we assume that δ is a Gaussian random field).

- The correlation function of δ in real space is defined as

$$\xi(y) \equiv \langle\delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{y})\rangle\ , \tag{2.27}$$

  where the average extends over all positions $\boldsymbol{x}$ and orientations of $\boldsymbol{y}$. The correlation function measures the coherence of the density contrast between all points on the sky separated by a distance $|\boldsymbol{y}|$. Again, ξ cannot depend on the direction of $\boldsymbol{y}$ because of isotropy.

- Inserting the Fourier integrals for $\delta(\boldsymbol{x})$ in (2.27), we find

$$\begin{aligned}\xi(y) &= \left\langle\int\frac{\mathrm{d}^3k}{(2\pi)^3}\int\frac{\mathrm{d}^3k'}{(2\pi)^3}\hat\delta(\boldsymbol{k})\hat\delta(\boldsymbol{k}')\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\mathrm{e}^{-\mathrm{i}\boldsymbol{k}'\cdot(\boldsymbol{x}+\boldsymbol{y})}\right\rangle \\ &= \int\frac{\mathrm{d}^3k}{(2\pi)^3}\int\frac{\mathrm{d}^3k'}{(2\pi)^3}\langle\hat\delta(\boldsymbol{k})\hat\delta^*(\boldsymbol{k}')\rangle\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\mathrm{e}^{+\mathrm{i}\boldsymbol{k}'\cdot(\boldsymbol{x}+\boldsymbol{y})} \\ &= \int\frac{\mathrm{d}^3k}{(2\pi)^3}P(k)\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{y}}\ , \qquad (2.28)\end{aligned}$$

which states that the correlation function is the Fourier transform of the power spectrum (and vice versa). Hence, both statistical measures carry an equivalent amount of information. Simplifying furthermore, we obtain

$$\begin{aligned}\xi(y) &= 2\pi\int\frac{k^2\mathrm{d}k}{(2\pi)^3}P(k)\int_0^\pi\sin\theta\mathrm{d}\theta\mathrm{e}^{-\mathrm{i}ky\cos\theta} \\ &= 4\pi\int\frac{k^2\mathrm{d}k}{(2\pi)^3}P(k)\frac{\sin ky}{ky}\ , \qquad (2.29)\end{aligned}$$

where θ was the angle between vectors $\boldsymbol{k}$ and $\boldsymbol{y}$. Obviously, the variance of δ is the correlation function at $y = 0$,

$$\sigma^2 = 4\pi\int\frac{k^2\mathrm{d}k}{(2\pi)^3}P(k)\ . \qquad (2.30)$$

- The variance in real space depends on the scale which we are considering. Let us introduce

$$\bar\delta(\boldsymbol{x}) := \int\mathrm{d}^3y\delta(\boldsymbol{x})W_R(|\boldsymbol{x}-\boldsymbol{y}|)\ , \qquad (2.31)$$

i.e. the density contrast field averaged on the scale R with a *window function* W_R. The idea of the window function is that it approaches a finite constant well within R, and drops to zero outside R.

- The Fourier convolution theorem says $\widehat{f*g} = \hat f\hat g$, i.e. the Fourier transform of a convolution is the product of the Fourier transforms of the convolved functions. Applying this to (2.31) yields $\hat{\bar\delta} = \hat\delta\hat W_R$. thus, the power spectrum of the density contrast filtered on the scale R is $\bar P(k) = P(k)\hat W_R^2(k)$. Using (2.30), the variance of the filtered density-contrast field is

$$\sigma_R^2 = 4\pi\int\frac{k^2\mathrm{d}k}{(2\pi)^3}P(k)\hat W_R^2(k)\ . \qquad (2.32)$$

The variance on a scale of $8\,h^{-1}$ Mpc, σ_8, is often used for characterizing the amplitude of the power spectrum.

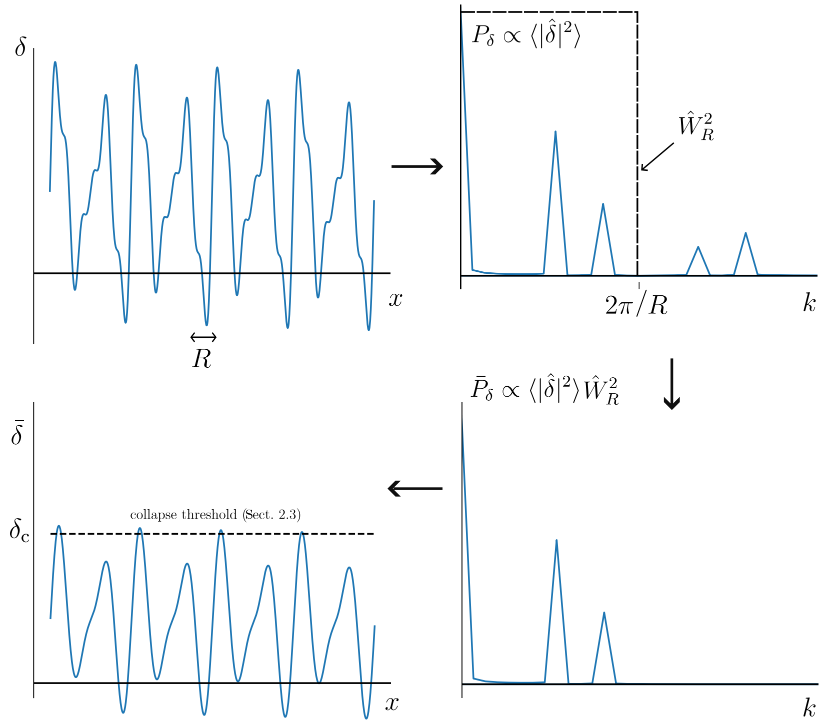


Figure 2.1: A particular realization of the unfiltered density contrast $\delta(x)$ in real space x (top left panel) is contrasted to its power spectrum in Fourier space, $P_\delta(k)$ (top right panel), where $k = 2\pi/x$ is the wave number. Also shown is the Fourier transform of the window function $\hat{W}_R$, which is applied to $P_\delta(k)$. Effectively, this cuts off contributions from waves with higher wavenumbers than $2\pi/R$ (bottom right panel). Transforming back to real space yields the filtered density contrast field $\bar{\delta}$ (bottom left panel). Es a result, all contributions from high spatial frequencies are removed. Filtered density contrasts with an amplitude higher than a critical threshold δ_c will collapse and eventually form a halo.

- In practice, we live in one Universe and and cannot average over many ensemble of universes. Hence, one adopts the ergodic hypothesis that assumes that an ensemble average equals a spatial average provided the averaging volume is large enough so that it contains many realizations of the density contrast at a given scale and can be considered to be a random field with the conditions specified above. In this case, we measure the power spectrum of a survey or a simulation where the data is discretized as follows:

  $$P(k) = \frac{1}{N}\sum_{i=1}^{N} \left|\hat{\delta}(\boldsymbol{k}_i)\right|^2, \tag{2.33}$$

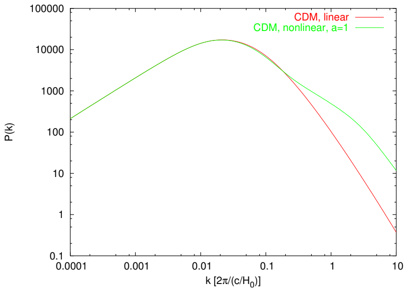
  where we average the absolute square of all (complex) Fourier components with $|\boldsymbol{k}_i| = k$.

2.2.2 Hierarchical Structure Formation

- A detailed study of how modes of different wave numbers grow in the radiation- and matter-dominated era allows to infer the shape of the power spectrum. In the linear regime of small-amplitude perturbations this can be worked out analytically (see e.g., Section 2.2.2 of the *Cosmology* lecture notes by Prof. Bartelmann). For a cold dark matter (CDM) cosmology, the resulting power spectrum reads

$$P(k) \propto \begin{cases} k & (k < k_0) \\ k^{-3} & (k \gg k_0). \end{cases} \tag{2.34}$$

Here, $k_0 = 2\pi a_{\mathrm{eq}}/\lambda_0$ is the comoving wave number of the particle horizon at matter-radiation equality.



Linear and non-linear CDM power spectra.

- We use this result to understand when galaxy clusters form in comparison to elliptical or dwarf galaxies. To this end, we define the *non-linear mass* M_*, as the mass contained in a sphere of radius R_* on which the variance becomes unity:

$$\sigma_*^2 = \int_0^{k_*} \frac{\mathrm{d}^3k}{(2\pi)^3} P(k) \overset{!}{=} 1, \tag{2.35}$$

where $k_* = 2\pi/R_*$.

- We assume that the power spectrum can be approximated locally by a power law of the form $P(k) = Ak^n$,

$$\begin{aligned} \sigma_*^2 &= 4\pi A \int_0^{k_*} \frac{k^{2+n}\mathrm{d}k}{(2\pi)^3} = \frac{4\pi A}{(2\pi)^3}\frac{k_*^{n+3}}{n+3} \overset{!}{=} 1, &(2.36)\\ \sigma^2 &= \sigma_*^2\left(\frac{k}{k_*}\right)^{n+3} = \left(\frac{k}{k_*}\right)^{n+3}. &(2.37) \end{aligned}$$

- Masses and length scales are related by background density, $M = 4\pi\bar{\rho}R^3/3 \propto k^{-3}$. For a fixed volume, density fluctuations are thus related to fluctuations in the mass of this volume, $\delta\rho \propto \delta M$. Normalized by the average density, we obtain $\delta \propto \delta M/M$. Using the definition of the variance in Eq. (2.30) and the power spectrum in Eq. (2.26), we get

$$\sigma^2 = \left\langle \left(\frac{\delta M}{M}\right)^2 \right\rangle = \left(\frac{M}{M_*}\right)^{-1-n/3}. \tag{2.38}$$

The variance reads for the two asymptotic cases

$$\sigma^2 = \begin{cases} \left(\frac{M}{M_*}\right)^{-4/3} & \text{for } n = 1, \\ 1 & \text{for } n = -3. \end{cases} \tag{2.39}$$

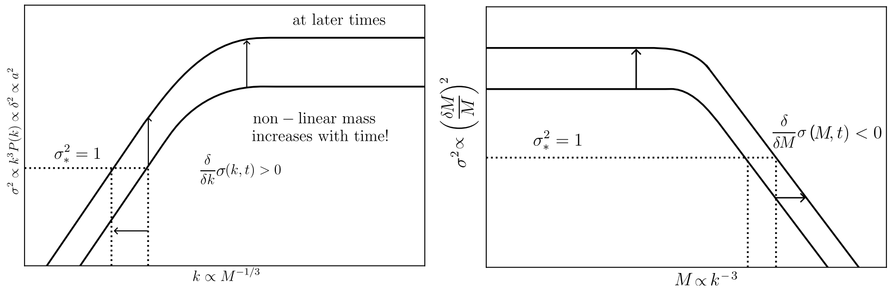


Figure 2.2: Variance of the density contrast (corresponding to a dimensionless power spectrum) as function of wave number (left) and as a function of mass contained within a (positive) perturbation (right). We see that first, power grows on small scales (large wave numbers) so that these scale first become non-linear and collapse via the influence of gravity. At later times, larger and larger scales collapse, giving rise to hierarchical structure formation ("bottom-up" growth of structures).

- The previous considerations allow to infer how cosmological structures grow. In k space, Eqs. (2.34) and (2.37) demonstrate that the variance is largest on small scales, i.e., the amplitude of small-scale fluctuations is largest. Hence these scales reach the non-linear mass first which causes them to collapse. As a result, structure forms "bottom-up" in CDM cosmologies. This hierarchical scheme of structure formation dictates that dwarf galaxies collapse and form before ellipticals, which in turn form earlier than galaxy clusters. Those sit atop the mass hierarchy as they have collapsed most recently in cosmic time. The reason for this can be easily understood by looking at the variance as a function of collapsing mass, Eq. (2.39). Gravity dictates that an overdensity continues to grow with time. The growth rate depends on the amount of matter: the more matter, the stronger the gravity, but the longer it takes to collapse. We can summarize, that the necessary and sufficient criterion for hierarchical structure formation is given by
$$\frac{\partial}{\partial M}\sigma(M,t) < 0. \tag{2.40}$$
- Another way of looking at this is that fluctuations on large scales are more subtle than fluctuations on small scales because the Universe is homogeneous on the largest scales according to the *cosmological principles* and as inferred from large-scale observations. Is there a deep reason why fluctuations are smaller on large scales? Let's look at the fluctuations in the gravitational potential,
$$\delta\Phi \sim \frac{GM}{R}\frac{\delta M}{M} \sim GM^{2/3}\bar{\rho}^{1/3}\frac{\delta M}{M} \tag{2.41}$$
since at any time $R \propto \bar{\rho}^{-1/3}$.
- Unless $\delta M/M \propto M^{-2/3}$, the potential fluctuations $\delta\Phi$ will diverge. Depending on the power-law index of $\delta M/M \propto M^{-\alpha}$, $\delta\Phi$ will

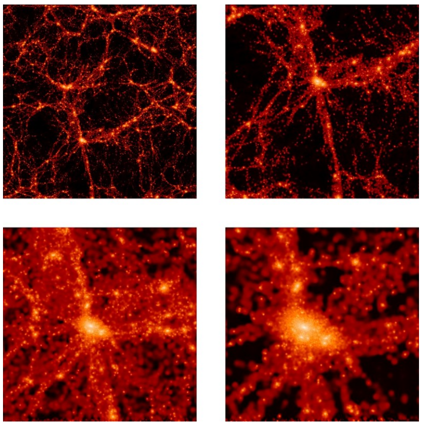
diverge on large scales (for $\alpha < 2/3$) or on small scales (for $\alpha > 2/3$). Hence, the most natural fluctuation spectrum is $\delta M/M \propto M^{-2/3}$, which avoids divergences. This can be related to a power spectrum in wave number space by considering $\delta\Phi \sim Gk\delta M$ and $M \propto R^3 \propto k^{-3}$. Taking the square root of Eq. (2.38) yields

$$\frac{\delta M}{M} \propto M^{-(n+3)/6} \quad \Rightarrow \quad \delta M \propto M^{-(n-3)/6} \propto k^{(n-3)/2}, \tag{2.42}$$

or

$$\delta\Phi \propto k^{(n-1)/2}. \tag{2.43}$$

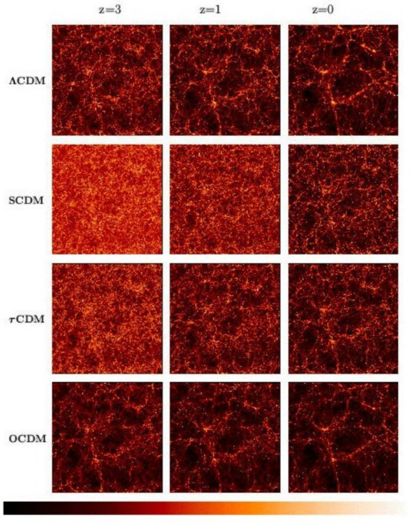
This shows that $n = 1$ is the characteristic spectral index index that avoids any unphysical divergence and corresponds to the so-called Harrison-Zel'dovich-Peebles spectrum of initial fluctuations.



2.2.3 Non-linear Evolution

- When the density contrast reaches unity, linear perturbation theory breaks down. For a correct treatment of the non-linear development, one has to resort to numerical simulations. They decompose the matter distribution into a large number of N particles whose velocities at some initial time are typically slightly perturbed according to some assumed power spectrum. This initial time or redshift is chosen such that linear evolution holds on all scales considered, even on the smallest resolved scales. For later evolution, the equations of motion for all particles are solved. In the following, the most popular numerical techniques shall be presented:

  1. Ideally, particles move under the influence of the gravity from all other particles, but direct summation of all the gravitational forces of $N-1$ particles on N particles becomes prohibitively time-consuming (because of the $(N-1)\times N \sim N^2$ computational operations per timestep) and the scheme attains numerical complexity of $\mathcal{O}(N^2)$ for every timestep. Several approximation schemes are therefore being employed.
  2. The particle-mesh (PM) algorithm computes the gravitational potential of the particle distribution on a grid (mesh) by solving Poisson's equation in Fourier space, making use of fast-Fourier techniques, thereby reducing the numerical complexity to $\mathcal{O}(N \log N)$. The gravitational forces are then given by the gradients of the potential at the particle positions. This technique has a spatial resolution limited by the size of the mesh cells which makes it impractical for many modern applications.



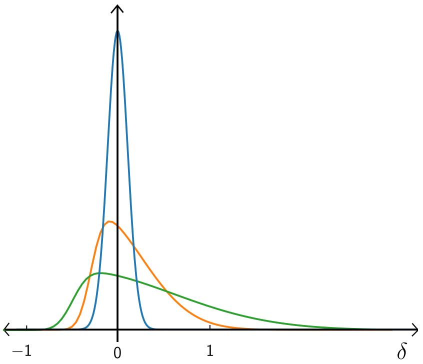
The VIRGO Collaboration 1996

Nonlinear structure evolution, simulated in different cosmologies (Virgo collaboration).

3. The particle-particle particle-mesh (P^{3}M) algorithm improves the PM technique by adding corrections for nearby particles which are determined by direct summation. Here, the numerical complexity is also $\mathcal{O}(N \log N)$ provided the number of particle-particle operations per timestep is kept constant. However, this is not the case for high-resolution "zoom" simulations of individually forming objects in cosmological environments.
4. Tree codes bundle distant particles into groups whose gravitational force on a particle is approximated as if they were point masses, or masses whose spatial distribution has a few low-order multipoles only, e.g. the monopole corresponding to a point mass, plus a dipole corresponding to a linear deformation, and so on. Depending on the solid angle that is subtended by the group on the sky seen by the particle, the "tree" is opened into its branches and leaves, i.e., higher-order multipoles of this group are considered. Alternatively, the monopole of that group, which is centered on its center-of-mass, is subdivided into the monopole moments of subgroups of the parent group, which subtend a smaller angle on the sky. It turns out that the numerical complexity of this technique is also $\mathcal{O}(N \log N)$. The particle tree is updated as the evolution proceeds.

- Non-linear evolution causes density-perturbation modes to couple: while modes of different wave lengths evolve independently during linear evolution, mode coupling in the non-linear evolution moves power from large to small scales as structures collapse. The effect on the power spectrum is that the amplitude on small scales is increased at the expense of intermediate scales. Large scales continue to evolve linearly and independently.

- Even if the original density perturbation field δ is Gaussian, it must develop non-Gaussianities during non-linear evolution. This is evident because $\delta \geq -1$ by definition, but can become arbitrarily large. An originally Gaussian distribution of δ thus becomes increasingly skewed as it develops a tail towards infinite δ.



The original density contrast follows a Gaussian distribution (blue; here we exaggerate the width for visual purposes). With increasing time (orange to green), the density contrast becomes more skewed towards the positive values as a result of gravitational collapse, which increases the density exponentially fast. Note that δ is bounded below because $\delta = (\rho - \rho_0)/\rho_0 > -1$.

- Typical behavior seen in numerical simulations shows the formation of "pancakes" (or cosmic walls) and filaments as predicted by the theory of Gaussian random fields. Gravitational fragmentation of filaments into individual lumps causes galaxy-sized dark matter overdensities to form, which are called halos. In the ΛCDM universe, those merge into galaxy groups which gradually stream towards the higher-density regions and larger mass concentrations at the intersections of filaments—galaxy clusters. Those form at the sites of constructive interference of long waves in the primordial fluctuations and are enhanced through gravitational collapse. Ongoing gravitational pull on the surrounding regions causes galaxy- and group-sized halos to continuously merge into clusters which sit atop the cosmic mass hierarchy of halos and thereby present the largest gravitationally collapsed objects to date. Giant voids form as matter accumulates in the walls and filaments of the cosmic network. Equivalently, the formation of voids can be considered to result from destructive interference of waves in the primordial fluctuations.

2.3 Spherical Collapse

2.3.1 Collapse of a Homogeneous Overdense Sphere

- In the previous chapter, we have considered the growth of the cosmic density field from small, linear perturbations, to large non-linear structures and use a power spectrum analysis to learn, which perturbations collapse first, and which grow later. We found that structures first collapses on small scales and then grows hierarchically from small to large scales respectively masses due to mergers of small objects. In fact, the distribution of the dark matter in the universe is not smooth but instead can be considered as being composed of individual so-called haloes, approximately spherical overdense clouds of dark matter which can reach highly non-linear densities in their centres. We will now follow the collapse of such an individual halo and consider an overdensity that first expands with the universe, then starts to lag behind as a result of the mutual gravitational attraction and finally collapse to form a halo.

- An approximate understanding of the parameters of such haloes and their relation to the dark-matter density contrast can be obtained by studying the dynamics of a spherical, homogeneous overdensity, leading to the so-called spherical collapse model. While realistic density perturbations are not spherical, considering an exact analytical solution that results from such an analysis nevertheless provides useful insights into non-linear collapse of more realistic situations. In particular the analysis (1) relates time (or redshift) at which the object collapses to its initial (linear) overdensity and (2) it maps the collapse time (redshift) to the final density of dark matter haloes that formed by collapse.

- The measured temperature anisotropies in the cosmic microwave background imply $\delta \ll 1$ at recombination. Thus, non-linear collapse happens at $a \gg a_{\rm rec}$, i.e., in the matter- or vacuum-dominated eras. We make the following assumptions in our analysis.
  - We consider a spherical perturbation that has initially a uniform overdensity.
  - The fluid is assumed to have zero pressure and is collisionless (i.e., the analysis applies to dark matter and not baryons). Later stages of baryonic collapse are different from that of dark matter since baryons additionally feel the pressure force, which causes the development of shocks in converging flows. However, since baryons only contribute

  $\sim 15\%$ of the total mass, they do not appreciably change the collapse of dark matter.
  - For simplicity, we set $\Omega = \Omega_{\rm m} = 1$, i.e., a flat matter-dominated universe. This can be generalized to cases with $\Omega_{\rm m0} \neq 1$ and $\Omega_\Lambda \neq 1$.

- We consider a sphere of mass M and proper radius R and assume that the universe outside the sphere remains spherically symmetric such that it exerts no gravitational force on the matter in the sphere. Since $M = \text{const.}$, we have

$$\frac{\mathrm{d}^2R}{\mathrm{d}t^2} = -\frac{GM}{R^2} , \tag{2.44}$$

which can be integrated to yield

$$\frac{1}{2}\left(\frac{\mathrm{d}R}{\mathrm{d}t}\right)^2 - \frac{GM}{R} = \Phi . \tag{2.45}$$

- We consider the gravitationally bound case, for which the energy per units mass is $\Phi < 0$ and which leads to collapse. Adopting $R = 0$ at $t = 0$, we can integrate this equation and obtain

$$t = \int_0^R \frac{\mathrm{d}r}{\sqrt{2\,(GM/r + \Phi)}} = \frac{A}{\sqrt{2|\Phi|}} \int_{\theta(0)}^{\theta(R)} \frac{\sin\theta \mathrm{d}\theta}{\sqrt{2/(1-\cos\theta) - 1}} , \tag{2.46}$$

where we suitably changed the integration variable, using the transformation $r = A(1 - \cos\theta)$, where $A = GM/(2|\Phi|)$. Employing trigonometric identities, we obtain

$$t = \frac{A}{\sqrt{2|\Phi|}} \int_{\theta(0)}^{\theta(R)} (1 - \cos\theta)\mathrm{d}\theta = \frac{A}{\sqrt{2|\Phi|}}(\theta - \sin\theta) . \tag{2.47}$$

Thus, the spherical collapse problem has the following parametric solution, which describes a cycloid,

$$R = A(1-\cos\theta) , \quad A = \frac{GM}{2|\Phi|} , \tag{2.48}$$

$$t = B(\theta - \sin\theta) , \quad B = \frac{GM}{(2|\Phi|)^{3/2}} . \tag{2.49}$$

- The solution is characterised by an initial expansion phase from $R = 0$ at $\theta = 0$. It reaches a maximum radius $R_{\rm ta} = 2A$ at $\theta_{\rm ta} = \pi$ at which it turns around and collapses back to $R = 0$ at $\theta_{\rm c} = 2\pi$. In principle, it re-expands for $\theta > 2\pi$ but in practice, other physical effects become important and complicate things. The corresponding times are $t_{\rm ta} = \pi B$ for the maximum (turn-around) radius and $t_{\rm c} = 2\pi B = 2t_{\rm ta}$ for collapse at $R = 0$.

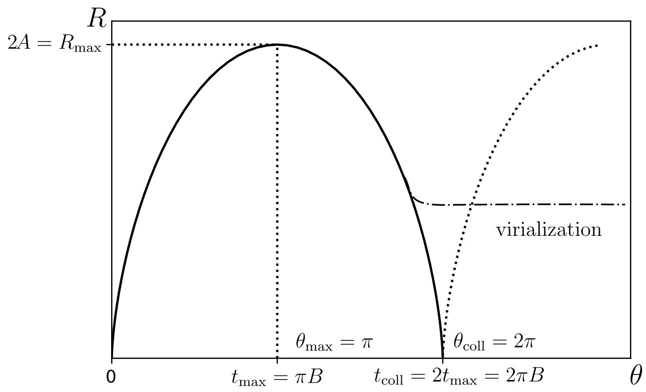


Figure 2.3: Cycloid solution of the spherical collapse problem (see text).

2.3.2 Connection to Linear Perturbation Theory

- The mean density inside the sphere is according to Eq. (2.48)

  $$\rho = \frac{M}{4\pi/3\,R^3} = \frac{3M}{4\pi A^3}\frac{1}{(1-\cos\theta)^3}\,, \tag{2.50}$$

  while the mean density of the background universe with $\Omega_{\rm m0} = 1$ is

  $$\bar\rho = \frac{3H^2}{8\pi G} = \frac{1}{6\pi G t^2} = \frac{1}{6\pi G B^2}\frac{1}{(\theta-\sin\theta)^2}\,, \tag{2.51}$$

  with $H = 2/(3t)$. The overdensity of the sphere (which is generally non-linear) can be obtained by combining these equations to yield

  $$1+\delta = \frac{\rho}{\bar\rho} = \frac{9}{2}\frac{(\theta-\sin\theta)^2}{(1-\cos\theta)^3}\,. \tag{2.52}$$

- To make the connection to linear perturbation theory, we consider the behaviour of the collapse at small t, which corresponds to small θ. Performing a Taylor series expansion of $\cos\theta$ and $\sin\theta$, we obtain

  $$\begin{aligned} 1+\delta &= 1+\frac{3}{20}\theta^2 + O(\theta^4)\,, && (2.53)\\ t &= \frac{B}{6}\theta^3 + O(\theta^5)\,. && (2.54) \end{aligned}$$

  Solving for θ gives (using $t_{\rm ta} = \pi B$)

  $$\theta = \left(\frac{6t}{B}\right)^{1/3} + \ldots = (6\pi)^{1/3}\left(\frac{t}{t_{\rm ta}}\right)^{1/3} + \ldots, \quad \text{for } t \ll t_{\rm ta}. \tag{2.55}$$

- Thus, $\theta \ll 1$ corresponds to $t \ll t_{\rm ta}$. Substituting Eq. (2.55) into Eq. (2.53) gives

  $$\delta = \frac{3}{20}(6\pi)^{2/3}\left(\frac{t}{t_{\rm ta}}\right)^{2/3} \ll 1, \quad \text{for } t \ll t_{\rm ta}. \tag{2.56}$$

This yields the scaling of the density contrast in the spherical collapse model, $\delta \propto t^{2/3} \propto a$ (since $t \propto a^{3/2}$ in the Einstein-de Sitter model), which is exactly the behaviour of the growing mode solution of linear perturbation theory. Note that the decaying mode solution is absent due to our choice of initial conditions at $t = 0$: we only considered the gravitationally bound case, for which the energy per units mass is $\Phi < 0$ and which leads to an increasing density as a result of the collapse.

- A corollary emerges from Eq. (2.56) that if the sphere has a uniform initial overdensity (δ_i) at some early time (t_i), then *all* interior spheres will have the same t_ta and hence the sphere remains uniform as it collapses!

- There is an important distinction between (1) the real overdensity (which will be fully studied in Section 2.3.3) and (2) the overdensity *extrapolated* according to linear theory,

$$\delta_\mathrm{lin} = \delta_\mathrm{i}\left(\frac{t}{t_\mathrm{i}}\right)^{2/3} = \frac{3}{20}(6\pi)^{2/3}\left(\frac{t}{t_\mathrm{ta}}\right)^{2/3} \quad \text{for all } t. \tag{2.57}$$

The maximum expansion radius at turnaround ($t = t_\mathrm{ta}$) is

$$\delta_\mathrm{lin}(t_\mathrm{ta}) = \frac{3}{20}(6\pi)^{2/3} \approx 1.062 \tag{2.58}$$

while the real (non-linear) overdensity is according to Eq. (2.52)

$$1 + \delta(t_\mathrm{ta}) = \frac{9\pi^2}{16} \approx 5.55. \tag{2.59}$$

- At collapse ($t = t_\mathrm{c} = 2t_\mathrm{ta}$), we have

$$\delta_\mathrm{c} \equiv \delta_\mathrm{lin}(t_\mathrm{c}) = \frac{3}{20}(12\pi)^{2/3} \approx 1.686. \tag{2.60}$$

In terms of the initial overdensity δ_i, collapse happens at time

$$t_\mathrm{c} = t_\mathrm{i}\left(\frac{\delta_\mathrm{c}}{\delta_\mathrm{i}}\right)^{3/2} \propto \delta_\mathrm{i}^{-3/2}, \tag{2.61}$$

$$1 + z_\mathrm{c} = (1 + z_\mathrm{i})\left(\frac{\delta_\mathrm{i}}{\delta_\mathrm{c}}\right) \propto \delta_\mathrm{i}\ , \tag{2.62}$$

since $t \propto a^{3/2} \propto (1 + z)^{-3/2}$. Thus, perturbations that are initially more overdense collapse earlier! Generally, $\delta_\mathrm{c} = \delta_\mathrm{c}(\Omega_\mathrm{m}, \Omega_\Lambda)$, but the dependence on Ω_m and Ω_Λ is weak so our result applies quite generally although it was derived for the Einstein-de Sitter model.

2.3.3 Final Density of a Collapsed Halo

- According to the spherical top hat collapse model, a uniform sphere collapses to a point of infinite density and then re-expands. In a realistic situation, the sphere contains inhomogeneities that generate tangential random velocities in the dark matter during collapse. This leads to an equilibrium configuration where the dark matter velocity dispersion balances its gravity. This relaxation process is called virialization.

- We assume that the *final* dark matter halo is in dynamical equilibrium and obeys the virial theorem

  $$2K_\mathrm{f} + V_\mathrm{f} = 0\,, \tag{2.63}$$

  where K denotes the total kinetic energy in random motions, V is the total gravitational binding energy, and we neglected the surface pressure term due to further infalling material.

- To calculate the total gravitational binding energy of a homogeneous sphere, we write down the masses of a shell and the sphere contained within it,

  $$\mathrm{d}m_\mathrm{shell} = 4\pi r^2 \rho \mathrm{d}r \qquad \text{and} \qquad m_\mathrm{interior} = \frac{4}{3}\pi r^3 \rho. \tag{2.64}$$

  The gravitational binding energy of a differential shell is given by

  $$\mathrm{d}V_\mathrm{f} = -G\frac{m_\mathrm{interior}\mathrm{d}m_\mathrm{shell}}{r}, \tag{2.65}$$

  which can be integrated to obtain the total gravitational binding energy,

  $$\begin{aligned} V_\mathrm{f} &= -G\int_0^M \frac{4\pi r^3 \rho}{3r}\mathrm{d}m_\mathrm{shell} = -G\frac{16}{3}\pi^2\rho^2\int_0^{R_\mathrm{f}} r^4 \mathrm{d}r \\ &= -G\frac{16}{15}\pi^2\rho^2 R_\mathrm{f}^5 = -\frac{3}{5}\frac{GM^2}{R_\mathrm{f}}. \end{aligned} \tag{2.66}$$

  In the last step, we eliminated ρ by adopting the density of a homogeneous sphere, $\rho = M/[(4/3)\pi R_\mathrm{f}^3]$.

- Hence, we have the kinetic and gravitational binding energies:

  $$K_\mathrm{f} = \frac{M}{2}\sigma_\mathrm{f}^2\,, \quad \text{and} \tag{2.67}$$

  $$V_\mathrm{f} = -\frac{3}{5}\frac{GM^2}{R_\mathrm{f}}\,, \tag{2.68}$$

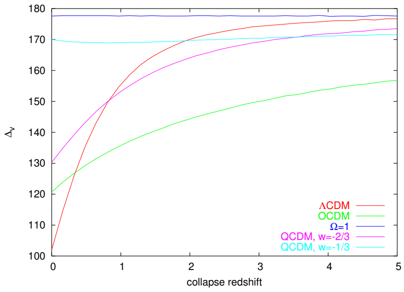
  where σ is the three-dimensional velocity dispersion and obtain the total energy

  $$E_\mathrm{f} = K_\mathrm{f} + V_\mathrm{f} = \frac{1}{2}V_\mathrm{f} = -\frac{3}{10}\frac{GM^2}{R_\mathrm{f}}\,. \tag{2.69}$$

- At turn-around, the sphere is at rest, i.e., $K_{\rm ta} = 0$. The total energy at turn-around is
$$E_{\rm ta} = V_{\rm ta} = -\frac{3}{5}\frac{GM^2}{R_{\rm ta}} \; . \tag{2.70}$$
Since dark matter is collisionless, the conservation of total energy during the collapse yields $E_{\rm f} = E_{\rm ta}$ and hence, $R_{\rm f} = R_{\rm ta}/2$.

- The final density is thus $\rho_{\rm f} = 8\rho(t_{\rm ta})$. Assuming that virialization happens at $t \approx t_{\rm c}$ and since $\bar{\rho} \propto t^{-2}$ and $t_{\rm c} = 2t_{\rm ta}$, the overdensity of the final halo is
$$1 + \delta_{\rm v} \equiv 1 + \delta_{\rm coll} = \frac{\rho_{\rm coll}}{\bar{\rho}\,(t_{\rm c}/t_{\rm ta})^{-2}} = 32\,[1 + \delta(t_{\rm ta})] = 18\pi^2 = 178 \; , \tag{2.71}$$
where $\rho(t_{\rm ta})/\bar{\rho} = 1 + \delta(t_{\rm ta})$ and we evaluated Eq. (2.52) at turn-around ($\theta = \pi$) so that $1 + \delta(t_{\rm ta}) = 9\pi^2/16$.

- Hence, the final halo density is
$$\rho_{\rm f} = (1 + \delta_{\rm v})\bar{\rho}(t_{\rm c}) = 18\pi^2\bar{\rho}(t_{\rm c}) \; . \tag{2.72}$$
$\delta_{\rm lin}$ and $\delta_{\rm v} \equiv \Delta_{\rm v}$ are widely used in cosmology to characterize dark matter haloes. $\delta_{\rm lin}$ is a useful concept to characterize the average over- or underdensity on a given scale in the initial conditions (in the linear regime) to predict whether it will expand into a void, stay at around the mean density, or even collapse into a halo (see Section 2.4.1). By contrast, $\delta_{\rm v} \equiv \Delta_{\rm v}$ is used to characterize the mean density within a collapsed halo and to estimate its mass, virial radius etc. Other popular choices are $\Delta_{\rm v} = 100, 200, 500$, where each definition has its merits and shortcomings.



Virial overdensity in different cosmologies as a function of the halo collapse redshift.

- These two parameters derived from the spherical collapse model, $\delta_{\rm coll}$ and $\Delta_{\rm v}$, are very widely used in cosmology for characterizing dark-matter haloes and their formation. Extending these calculations into more general cosmological models is surprisingly difficult and requires numerical solutions of the underlying differential equations. Approximations to the solutions for $\Omega_{\rm m} < 1$ are
$$\delta_{\rm coll} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \begin{cases} (1.0 + 0.0406\log_{10}\Omega_{\rm m}) & (\Omega_{\Lambda 0} = 0) \\ (1.0 + 0.0123\log_{10}\Omega_{\rm m}) & (\Omega_{\Lambda 0} = 1 - \Omega_{\rm m0}) \end{cases} \tag{2.73}$$
and
$$\Delta_{\rm v} = 9\pi^2 \begin{cases} \left[1 + 0.1210(\Omega_{\rm m} - 1) + \Omega_{\rm m}^{0.6756}\right] & (\Omega_{\Lambda 0} = 0) \\ \left[1 + 0.7076(\Omega_{\rm m} - 1) + \Omega_{\rm m}^{0.4403}\right] & (\Omega_{\Lambda 0} = 1 - \Omega_{\rm m0}) \end{cases} \tag{2.74}$$
where $\Omega_{\rm m}$ is the matter density parameter at the redshift of halo collapse.

2.4 The Halo Mass Function

2.4.1 The Press-Schechter Mass Function

- An important property of haloes is their distribution over mass, the so-called mass function, which yields the number density of haloes at redshift z within the mass range between M and $M+\mathrm{d}M$. The halo mass function is sensitive to cosmological parameters and especially well suited to test for deviations from our standard model of cosmology, which assumes a Gaussian process for setting the initial conditions of the density perturbations.

- A characteristic length scale $R(M)$ can be assigned to a region in the fluctuating density field that contains a mass M. The length scale is defined as the radius of a homogeneous sphere filled with the mean cosmic matter density,

$$\frac{4\pi}{3}R^3\rho_{\mathrm{cr}}\Omega_{\mathrm{m}} = M \;\Rightarrow\; R(M) = \left(\frac{3M}{4\pi\rho_{\mathrm{cr}}\Omega_{\mathrm{m}}}\right)^{1/3} , \tag{2.75}$$

where Ω_{m} and ρ_{cr} have to be evaluated at the redshift required. Upon gravitational collapse that region collapses to a halo and conserves its mass. However, the density is increased and the halo radius shrinks accordingly.

- To obtain a prescription for a halo of mass M emerging from the fluctuating density field, we consider the density contrast field filtered on the scale $R(M)$. We therefore use $\bar{\delta}$ as defined in Eq. (2.31), i.e. the density contrast convolved with a window function W_R which has a characteristic scale $R = R(M)$.

- It will be convenient to define a so-called "non-linear mass" M_* with a characteristic length scale $R(M_*) \equiv R_*$ that has a variance (Eq. 2.32) of the density contrast δ_{c}^2,

$$\sigma_{R_*}^2 = 4\pi\int_0^\infty \frac{k^2\mathrm{d}k}{(2\pi)^3}P(k)\hat{W}_{R_*}^2(k) = \delta_{\mathrm{c}}^2 , \tag{2.76}$$

Note that δ_{c} is the *linear* density contrast in the spherical collapse model extrapolated to the collapse time.

- For a Gaussian random field, the probability of finding at a given point $\boldsymbol{x}$ in space a filtered density contrast $\bar{\delta}(\boldsymbol{x})$ is

$$p(\bar{\delta}, a) = \frac{1}{\sqrt{2\pi\sigma_R^2(a)}}\exp\left[-\frac{\bar{\delta}^2(\boldsymbol{x})}{2\sigma_R^2(a)}\right] , \tag{2.77}$$

where we have explicitly noted that the variance σ will depend on time or equivalently on the scale factor a through the linear growth factor, $\sigma_R(a) = \sigma_R D_+(a)$ ($D_+(a) = a$ in the Einstein-de-Sitter model with $\Omega_{\mathrm{m}}(a) = 1$).

- Press & Schechter suggested that the probability of finding the filtered density contrast at or above the linear density contrast for spherical collapse, $\bar{\delta} > \delta_\mathrm{c}$, is equal to the fraction of the cosmic volume filled with haloes of mass M,

$$F(M,a) = \int_{\delta_\mathrm{c}}^{\infty} \mathrm{d}\bar{\delta} p(\bar{\delta},a) = \frac{1}{2}\mathrm{erfc}\left(\frac{\delta_\mathrm{c}}{\sqrt{2}\sigma_R(a)}\right) , \tag{2.78}$$

  where $\mathrm{erfc}(x)$ is the complementary error function. Obviously, this equation implies that the fraction of cosmic volume filled with haloes of fixed mass M is a highly sensitive function of the scale factor a, i.e., the strength of mass/density fluctuations on a certain scale.

- The distribution of haloes over masses M is simply $\partial F(M)/\partial M$, so we have to relate σ_R to M, which is accomplished by the characteristic radius $R(M)$,

$$\frac{\partial}{\partial M} = \frac{\mathrm{d}\sigma_R(a)}{\mathrm{d}M}\frac{\partial}{\partial\sigma_R(a)} = \frac{\mathrm{d}\sigma_R}{\mathrm{d}M}\frac{\partial}{\partial\sigma_R} , \tag{2.79}$$

  where we have inserted the variance σ_R on the scale R at the present epoch because this expression is valid for any epoch. Using

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{erfc}(x) = -\frac{2}{\sqrt{\pi}}\mathrm{e}^{-x^2} , \tag{2.80}$$

  we find

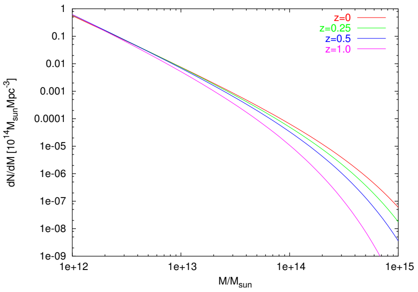
$$\left|\frac{\partial F(M)}{\partial M}\right| = \frac{1}{\sqrt{2\pi}}\frac{\delta_\mathrm{c}}{\sigma_R D_+(a)}\left|\frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M}\right|\exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_R^2 D_+^2(a)}\right) , \tag{2.81}$$

  where the absolute values have been added to ensure positiveness of the Press-Schechter mass function.

- The normalisation of the mass function is wrong, however. It is easy to see that

$$\int_0^\infty \left|\frac{\partial F(M)}{\partial M}\right| \mathrm{d}M = \frac{1}{2} \tag{2.82}$$

  the reason for this problem is quite subtle. The solution can be obtained with an elegant argument interpreting the statistics of halo formation in terms of a random walk (see Section 2.4.2). For for now, we will arbitrarily multiply the mass function by a factor of two.



Press-Schechter mass function for the ΛCDM model at four different redshifts

- This fraction of the cosmic volume filled with haloes of masses within $[M, M+\mathrm{d}M]$ is converted to a (comoving) number density of haloes by dividing with the mean volume $M/\bar{\rho}$ occupied by M

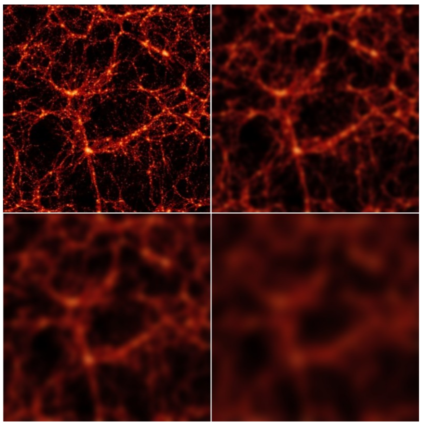
$$\begin{aligned} f(M,a)\mathrm{d}M &\equiv \frac{\partial n(M,a)}{\partial M}\mathrm{d}M = \frac{\bar{\rho}}{M}\left|\frac{\partial F(M)}{\partial M}\right|\mathrm{d}M \\ &= \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}\delta_\mathrm{c}}{\sigma_R D_+(a)}\left|\frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M}\right|\exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_R^2 D_+^2(a)}\right)\frac{\mathrm{d}M}{M} . \end{aligned} \tag{2.83}$$

- The Press-Schechter mass function (Eq. 2.83) has turned out to describe the mass distribution of dark-matter haloes in cosmological simulations remarkably well. Only recently have modifications been applied in order to improve its agreement with large, high-resolution simulations, or to take into account that halo collapse is not expected to proceed spherically, but elliptically.

2.4.2 Halo Formation as a Random Walk

- The normalisation problem, however, is embarrassing and needs to be resolved. The solution was given with an elegant argument interpreting the statistics of halo formation in terms of a random walk.

- Suppose the density-contrast field δ is given. A large sphere is centred on some point $\boldsymbol{x}$ and its radius gradually shrunk. For each radius R of the sphere, the density contrast $\bar{\delta}$ averaged within R is measured and monitored as a function of R. By choosing a window function W_R in the definition in Eq. (2.31) of $\bar{\delta}$ whose Fourier transform has a sharp cut-off in k space, $\bar{\delta}$ will undergo a random walk because decreasing R corresponds to adding shells in k space which are independent of the modes which are already included.

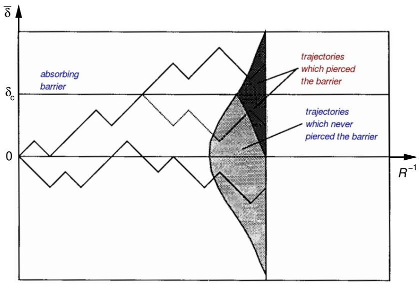


Progressive smoothing of the density field.

- $\bar{\delta}(\boldsymbol{x})$ is thus following a random trajectory. A halo is expected to be formed at $\boldsymbol{x}$ if $\bar{\delta}(\boldsymbol{x})$ reaches δ_c for some radius R. If $\bar{\delta}(\boldsymbol{x}) < \delta_\mathrm{c}$ for some radius, it may well exceed δ_c for a smaller radius. Or, if $\bar{\delta}(\boldsymbol{x}) \geq \delta_\mathrm{c}$ for some radius, it may well drop below δ_c for a smaller radius.

- For determining halo numbers correctly, it is thus necessary to count all points in space which are part of haloes of any mass. As R is shrunk around a point $\boldsymbol{x}$, that point must be counted as being part of a halo if there is a radius R for which $\bar{\delta}(\boldsymbol{x}) \geq \delta_\mathrm{c}$.

- In the terminology of the random walk, we need to introduce an *absorbing barrier* at δ_c such that points $\boldsymbol{x}$ with trajectories $\bar{\delta}(\boldsymbol{x})$ versus decreasing R which hit the barrier are removed from counting them as not being parts of haloes. To accomplish this, we follow the strategy of counting trajectories that do *not* make it into haloes such that the complement of that union represent trajectories of haloes. Hence, we define *allowed* trajectories that do *not* make it into haloes and *forbidden* trajectories that pierce the barrier *for some* R and thus form haloes.



Random walk with an absorbing barrier.

- A trajectory meeting the boundary has equal probability for moving above or below. For any *forbidden* trajectory continuing above the boundary, there is an *allowed* mirror trajectory continuing below it, and conversely. For any trajectory reaching a point $\bar\delta < \delta_\mathrm{c}$ exclusively along *allowed* trajectories, there is a path reaching its mirror point on the line $\bar\delta = \delta_\mathrm{c}$ exclusively along *forbidden* trajectories, and conversely. Thus, the probability for reaching a point $\bar\delta < \delta_\mathrm{c}$ along *allowed* trajectories exclusively below the barrier is the probability for reaching it along *any* trajectory, minus the probability for reaching its mirror point $\delta_\mathrm{c} + (\delta_\mathrm{c} - \bar\delta) = 2\delta_\mathrm{c} - \bar\delta$ along *forbidden* trajectories,

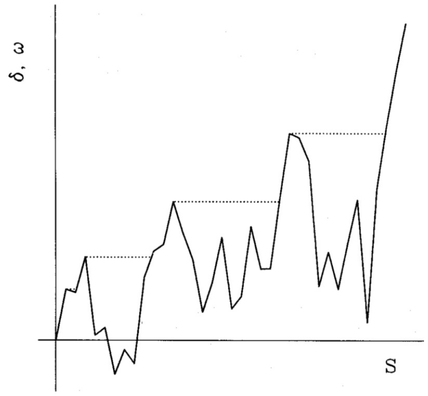
$$p_\mathrm{s}(\bar\delta)\mathrm{d}\bar\delta = \frac{1}{\sqrt{2\pi}\sigma_R}\left[\exp\left(-\frac{\bar\delta^2}{2\sigma_R^2}\right) - \exp\left(-\frac{(2\delta_\mathrm{c}-\bar\delta)^2}{2\sigma_R^2}\right)\right] , \quad (2.84)$$

where σ_R is the variance of $\bar\delta$ on the scale R, as before.

- Equation (2.84) is the probability distribution for the averaged density contrast to fall within $[\bar\delta, \bar\delta + \mathrm{d}\bar\delta]$ and *not* to exceed δ_c when averaged on *any* scale. The probability for $\bar\delta$ to *exceed* δ_c on some scale is thus

$$1 - P_\mathrm{s} = 1 - \int_{-\infty}^{\delta_\mathrm{c}} \mathrm{d}\bar\delta p_\mathrm{s}(\bar\delta) = \mathrm{erfc}\left(\frac{\delta_\mathrm{c}}{\sqrt{2}\sigma_R}\right) , \quad (2.85)$$

without the factor $1/2$ in Eq. (2.78). The rest of the derivation of the Press-Schechter mass function proceeds as before.



Trajectory of a halo in the S-ω plane. Increasing S means decreasing mass, and ω decreases with time.

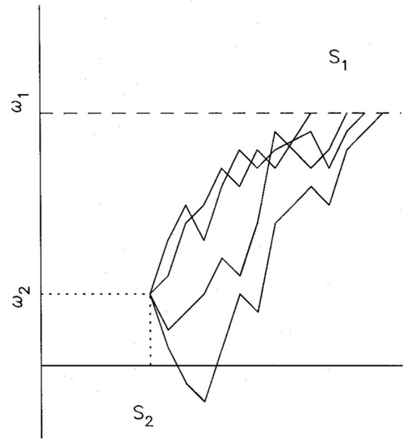
2.4.3 Extended Press-Schechter Theory

- Considering the random walk of the density contrast field when averaged over increasing or decreasing scales allows the statistics of haloes to be greatly extended. In order to simplify notation, we abbreviate $S := \sigma_R^2$.

- First, we note that we can either consider the barrier height δ_c to be constant while σ_R is increasing with time, or σ_R to be constant, while δ_c is decreasing with time, because only the ratio $\delta_\mathrm{c}/\sigma_R$ enters the relevant expressions. Thus, the barrier can be considered moving towards zero as time progresses,

$$\omega := \frac{\delta_\mathrm{c}}{D_+(a)} , \quad (2.86)$$

reflecting the fact that halo collapse becomes easier as structure formation proceeds. Since $\delta_\mathrm{c}(a)$ decreases monotonically with increasing time, it can uniquely be used instead of time. The evolution of a halo can now be expressed as a random walk in S as time proceeds, or ω decreases.



Trajectories of low-mass haloes at early time, forming a massive halo at a later time

- Second, we note that

$$
\begin{aligned}
-\frac{\partial P_\mathrm{s}}{\partial S}\mathrm{d}S &= -\frac{\partial}{\partial S}\int_{-\infty}^{\delta_\mathrm{c}} \mathrm{d}\bar{\delta}\, p_\mathrm{s}(\bar{\delta}) \\
&=: \; p_S(S,\omega)\mathrm{d}S = \frac{\omega}{\sqrt{2\pi S^3}}\mathrm{e}^{-\omega^2/2S}\mathrm{d}S \;, \qquad (2.87)
\end{aligned}
$$

  is the probability for $\bar{\delta}$ to hit the barrier δ_c for the first time when the variance is increased from S to $S+\mathrm{d}S$. It represents the fraction of mass in haloes of a mass M corresponding to the scale R.

- Consider now a trajectory passing through the barrier ω_2 for the first time at S_2, continuing to eventually pass through $\omega_1 > \omega_2$ at some $S_1 > S_2$. It represents a halo of mass M_1 corresponding to S_1 which, at a later time corresponding to ω_2, reaches mass $M_2 > M_1$ corresponding to S_2. The conditional probability for the halo to pass within $[S_1, S_1+\mathrm{d}S_1]$ at ω_1, starting from S_2 at ω_2 is, according to Eq. (2.87),

$$
p_{S_1}(S_1,\omega_1|S_2,\omega_2)\mathrm{d}S_1 = \frac{\omega_1-\omega_2}{\sqrt{2\pi}(S_1-S_2)^{3/2}}\exp\left[-\frac{(\omega_1-\omega_2)^2}{2(S_1-S_2)}\right]\mathrm{d}S_1 \qquad (2.88)
$$

  because the probability in Eq. (2.87) only needs to be transformed shifting the origin of trajectories from $(S,\omega)=(0,0)$ to $(S,\omega)=(S_2,\omega_2)$.

- From Eq. (2.88) and Bayes' theorem on conditional probabilities, we can straightforwardly derive the probability for a halo which for the first time reaches ω_1 at S_1 to reach ω_2 for the first time at S_2:

$$
\begin{aligned}
& p_{S_2}(S_2,\omega_2|S_1,\omega_1)\mathrm{d}S_2\, p_S(S_1,\omega_1)\mathrm{d}S_1 \\
= \; & p_{S_1}(S_1,\omega_1|S_2,\omega_2)\mathrm{d}S_1\, p_S(S_2,\omega_2)\mathrm{d}S_2 \\
\Rightarrow \; & p_{S_2}(S_2,\omega_2|S_1,\omega_1)\mathrm{d}S_2 \\
= \; & \frac{p_{S_1}(S_1,\omega_1|S_2,\omega_2)\mathrm{d}S_1\, p_S(S_2,\omega_2)\mathrm{d}S_2}{p_S(S_1,\omega_1)\mathrm{d}S_1} \\
= \; & \frac{1}{\sqrt{2\pi}}\left[\frac{S_1}{S_2(S_1-S_2)}\right]^{3/2}\frac{\omega_2(\omega_1-\omega_2)}{\omega_1} \\
\times \; & \exp\left[-\frac{(\omega_2 S_1-\omega_1 S_2)^2}{2S_1S_2(S_1-S_2)}\right]\mathrm{d}S_2 \;. \qquad (2.89)
\end{aligned}
$$

  This provides the conditional probability for a halo of mass M_1 to have merged to form a halo of mass between M_2 and $M_2+\mathrm{d}M_2$.

- The expected transition rate from S_1 to S_2 within the times t_1 and t_2 corresponding to ω_1 and ω_2 is determined by Eq. (2.89) taking the limit $\omega_2 \to \omega_1 =: \omega$,

$$\begin{aligned} &\frac{\mathrm{d}^2 p_{S_2}}{\mathrm{d}S_2\mathrm{d}\omega}(S_1 \to S_2|\omega)\mathrm{d}S_2\mathrm{d}\omega \qquad (2.90)\\ &= \frac{1}{\sqrt{2\pi}}\left[\frac{S_1}{S_2(S_1-S_2)}\right]^{3/2}\exp\left[-\frac{\omega^2(S_1-S_2)}{2S_1S_2}\right]\mathrm{d}S_2\mathrm{d}\omega\ . \end{aligned}$$

This gives the merger rate, i.e. the probability that, in the time interval corresponding to $\mathrm{d}\omega$, a halo of mass M_1 will merge with another halo of mass $M_2 - M_1$.

- We finally need to substitute the masses M_1 and M_2 for S_1 and S_2, and the time for ω. We wish to know the probability for a halo of mass M to accrete another halo of mass ΔM within the time interval $\mathrm{d}t$ at time t. The transformation is

$$\frac{\mathrm{d}^2 p_M}{\mathrm{d}\ln\Delta M\mathrm{d}t}(M_1 \to M_2|t) = \frac{\mathrm{d}S_2}{\mathrm{d}\ln\Delta M}\left|\frac{\mathrm{d}\omega}{\mathrm{d}t}\right|\frac{\mathrm{d}^2 p_{S_2}}{\mathrm{d}S_2\mathrm{d}\omega}\ . \qquad (2.91)$$

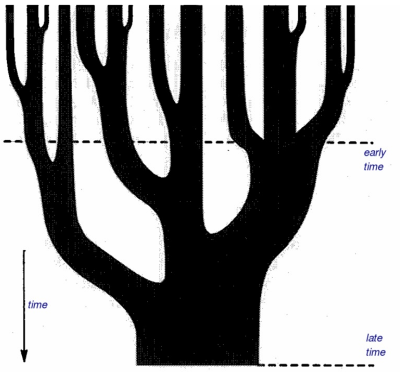
- By the definition of Eq. (2.86), the derivative of ω with respect to t is

$$\left|\frac{\mathrm{d}\omega}{\mathrm{d}t}\right| = \frac{\delta_\mathrm{c}}{D_+^2(a)}D_+'(a)\dot{a} = H\frac{\delta_\mathrm{c}}{D_+(a)}\frac{\mathrm{d}\ln D_+(a)}{\mathrm{d}\ln a}\ , \qquad (2.92)$$

where H is the Hubble parameter at scale factor a.

- Since $\Delta M = M_2 - M_1$, and S was introduced for σ_R^2, we have

$$\frac{\mathrm{d}S_2}{\mathrm{d}\ln\Delta M} = \Delta M\frac{\mathrm{d}\sigma_R^2(M_2)}{\mathrm{d}M_2}\ . \qquad (2.93)$$



A "merger tree", i.e. a graphical representation of the accretion history of a halo

- With expressions in Eqs. (2.92) and (2.93), the merger probability in Eq. (2.91) becomes

$$\begin{aligned} \frac{\mathrm{d}^2 p_M}{\mathrm{d}\ln\Delta M\mathrm{d}t} &= \sqrt{\frac{2}{\pi}}\frac{H\delta_\mathrm{c}}{\sigma_{R2}D_+}\frac{\mathrm{d}\ln D_+}{\mathrm{d}\ln a}\Delta M\frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M}(M+\Delta M)\\ &\times\left(1-\frac{\sigma_{R2}^2}{\sigma_R^2}\right)^{-3/2}\\ &\times\exp\left[-\frac{\delta_\mathrm{c}^2}{2\sigma_{R2}^2D_+^2}\left(1-\frac{\sigma_{R2}^2}{\sigma_R^2}\right)\right]\ , \qquad (2.94) \end{aligned}$$

where $\sigma_{R2} := \sigma_R(M_2) = \sigma_R(M+\Delta M)$.

- In much the same way, the random-walk interpretation of halo growth allows deducing halo-survival times and other interesting quantities related to halo growth.

2.5 Halo Density Profiles

2.5.1 General Remarks

- Generally, a self-gravitating system of particles does not have an equilibrium state. The virial theorem demands that its total energy ($E = K + V$) is minus half its potential energy (V),

  $$2K + V = E + K = 0 \quad \Rightarrow \quad K = -E = -\frac{V}{2}. \qquad (2.95)$$

  Since $V < 0$ for self-gravitating systems, any inevitable energy loss, e.g., through the ejection of a body by means of three-body encounters, makes the potential energy become more negative. As a result, the halo becomes more tightly bound, which in turn increases its energy loss because the dynamical timescale is reduced by this contraction according to

  $$t_{\mathrm{dyn}} \sim \left(\frac{R_{\mathrm{g}}^3}{GM}\right)^{1/2} \sim (G\rho)^{-1/2}\,, \qquad (2.96)$$

  where $R_{\mathrm{g}} = GM/v^2$ is the gravitational radius. Thus, any halo density profile must reflect a potentially long-lived, but transient state.

- Knowing global halo properties like their mass, their distribution in mass and redshift, and their growth over time, their internal density profiles are an important characteristic. We will discuss two widely used models for the density profiles.

2.5.2 Isothermal Sphere

- A simple analytic model for the density profile is the isothermal sphere, which is a spherically-symmetric, self-gravitating system of non-interacting particles whose kinetic energy is characterised by a constant "temperature" $T = \bar{m}\sigma^2/k_{\mathrm{B}}$ where σ denotes the three-dimensional velocity dispersion.

- The equations describing the isothermal sphere are thus the Euler equation of hydrostatic equilibrium,

  $$\frac{\mathrm{d}p}{\mathrm{d}r} = -\frac{GM(r)}{r^2}\rho\,, \qquad (2.97)$$

  and the equation of state for the ideal gas

  $$p = \frac{\rho}{\bar{m}}k_{\mathrm{B}}T\,, \qquad (2.98)$$

  where $\bar{m}$ is the mean mass of the particles constituting the sphere.

- Inserting Eq.(2.98) into Eq.(2.97) yields

$$\frac{k_\mathrm{B}T}{\bar{m}}\frac{\mathrm{d}\ln\rho}{\mathrm{d}r} = -\frac{G}{r^2}\int_0^r 4\pi\rho(r')r'^2\mathrm{d}r' \;, \tag{2.99}$$

where we have expressed the mass as an integral over the density. Differentiation with respect to r yields the second-order differential equation for ρ,

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}\ln\rho}{\mathrm{d}r}\right) = -\frac{4\pi G\bar{m}}{k_\mathrm{B}T}r^2\rho \;. \tag{2.100}$$

- One solution of Eq. (2.100) is singular and can be obtained by means of a power-law ansatz in r to yield

$$\rho_1(r) = \frac{\sigma^2}{2\pi Gr^2} \qquad \sigma^2 \equiv \frac{k_\mathrm{B}T}{\bar{m}} \;, \tag{2.101}$$

where σ is the (radially constant) velocity dispersion of the particles. The mass and circular velocity of the *singular isothermal sphere (SIS)* are given by

$$M(<r) = \frac{2\sigma^2}{G}r \quad \text{and} \quad v_\mathrm{c}^2 = \frac{GM(<r)}{r} = 2\sigma^2 \;. \tag{2.102}$$

- The solution to Eq. (2.100) depends on the boundary conditions. It turns out that there is a second solution, which has a finite central density ρ_0. To find this solution, we have to identify a characteristic length scale such that we can obtain a general solution in terms of dimensionless variables. The dimensional variables in Eq. (2.100) are G, ρ, and the combination $\sigma^2 = k_\mathrm{B}T/m$. Those can be combined to yield a length scale, $\sigma/\sqrt{G\rho}$, which represents the typical distance a particle travels in the central dynamical time. We define the *King radius* at which the density profile cores out,

$$r_0 \equiv \left(\frac{9\sigma^2}{4\pi G\rho_0}\right)^{1/2} \;, \tag{2.103}$$

and the dimensionless variables

$$\tilde{\rho} = \frac{\rho}{\rho_0} \;, \quad \text{and} \quad \tilde{r} = \frac{r}{r_0} \;. \tag{2.104}$$

- Equation (2.100) cast into dimensionless form reads

$$\frac{\mathrm{d}}{\mathrm{d}\tilde{r}}\left(\tilde{r}^2\frac{\mathrm{d}\ln\tilde{\rho}}{\mathrm{d}\tilde{r}}\right) = -9\tilde{r}^2\tilde{\rho} \;. \tag{2.105}$$

The (numerical) solution is obtained by integrating this differential equation outwards from $\tilde{r} = 0$ with the central boundary conditions $\tilde{\rho}(0) = 1$ and $\mathrm{d}\tilde{\rho}/\mathrm{d}\tilde{r} = 0$ (the second condition is necessary since $M(\tilde{r})$ vanishes at $\tilde{r} = 0$). The resulting second solution can be approximated by

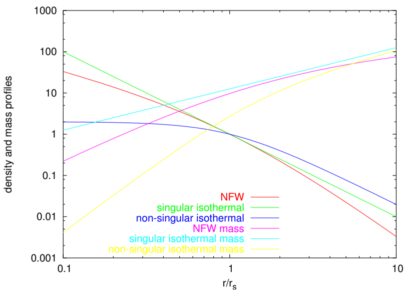
$$\tilde{\rho}_2(\tilde{r}) = \begin{cases} \left(1 + \tilde{r}^2\right)^{-3/2} & \tilde{r} \lesssim 3, \\ \frac{2}{9}\tilde{r}^{-2} & \tilde{r} \gtrsim 3, \end{cases} \tag{2.106}$$

i.e., the SIS is the asymptotic solution at large $\tilde{r}$. Note that by defining dimensionless variables, we have reduced the *family* of solutions with different densities and temperatures to a single solution for appropriately scaled variables.

- Both solutions have the advantage that they reproduce the flat rotation curves observed in spiral galaxies. The rotational velocity $v_{\rm rot}$ of a particle orbiting at radius r is determined by

$$v_{\rm rot}^2 = \frac{GM}{r} , \tag{2.107}$$

which is constant at $r \gg r_0$ for both density profiles of the isothermal sphere. However, the temperature within a stable "gas" sphere cannot be constant because particles would evaporate from it. Besides, the mass of the isothermal sphere diverges linearly as $r \to \infty$. To get a halo of finite mass, we must truncate it at some large radius by confining it with an external "pressure" that in practise is provided by the accretion of mass. The isothermal profile is thus at best an approximation for the inner parts of halos.



Singular and non-singular isothermal and NFW density and mass profiles.

2.5.3 Navarro-Frenk-White (NFW) Density Profile

- Numerical simulations of halo formation in the cold dark matter model consistently show density profiles like

$$\rho(r) = \frac{\rho_{\rm s}}{x(1+x)^2} , \quad x \equiv \frac{r}{r_{\rm s}} , \tag{2.108}$$

which have a characteristic scale radius $r_{\rm s}$ beyond which they fall off $\propto r^{-3}$, and within which the density profile flattens considerably.

- Using the identity $x/(1+x)^2 \equiv (1+x)^{-1} - (1+x)^{-2}$, the mass of such halos within radius r can easily be derived,

$$M(r) = 4\pi\rho_{\rm s}r_{\rm s}^3 \int_0^x \frac{x'{\rm d}x'}{(1+x')^2} = 4\pi\rho_{\rm s}r_{\rm s}^3 \left[\ln(1+x) - \frac{x}{1+x}\right] . \tag{2.109}$$

It rises $\propto x^2$ for small x and diverges logarithmically for $x \to \infty$. The divergence is not a fundamental problem because the halo profile must become invalid at the latest where ρ drops to the cosmic background density. In practise, the assumption of spherical symmetry starts to break down earlier, and becomes invalid at scales beyond the virial radius.

- The virial radius $r_{\rm vir}$ of a halo is often defined as the radius r_{200} enclosing a mean overdensity of 200 times the *critical* cosmic density,

$$\rho_{\rm cr}(a) \equiv \frac{3H^2(a)}{8\pi G}\ , \text{ where} \tag{2.110}$$

$$H^2(a) \equiv H_0^2\left[\Omega_{\rm m0}a^{-3} + \Omega_{\Lambda 0} + \Omega_{\rm K}a^{-2}\right] \tag{2.111}$$

is the Hubble function at late times, which are relevant for galaxies and clusters and which describes the expansion rate of the universe. This implies that the contribution of the cosmological constant Λ (and curvature K) to the critical density are included in the estimate of the reference density. Since $\Lambda = \text{const.}$, this has the advantage that the mass of a halo that has decoupled from the cosmic expansion can only grow by accretion but remains constant otherwise (at least at lates times when the cosmological constant dominates the energy density of the universe, $a > a_{\rm eq,\Lambda}$). The factor 200 is a rough approximation to the density contrast of $18\pi^2 \approx 178$ expected at virialisation in the spherical collapse model. This implies

$$M_{200}\left(\frac{4\pi}{3}r_{200}^3\right)^{-1} = 200\rho_{\rm cr}(a) = 200\frac{3H^2(a)}{8\pi G}\ , \tag{2.112}$$

where M_{200} is often identified with the total halo mass M. We obtain

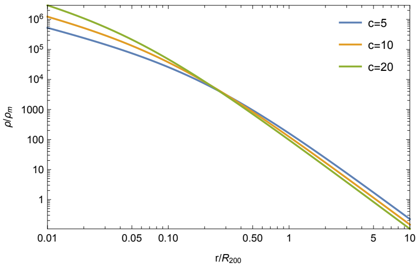
$$r_{200} = \left(\frac{GM_{200}}{100H^2}\right)^{1/3}\ . \tag{2.113}$$

- Other frequent definitions use the radius enclosing a mean overdensity of 200 times the *mean* matter density (i.e., without the Λ contribution),

$$M_{\rm 200m}\left(\frac{4\pi}{3}r_{\rm 200m}^3\right)^{-1} = 200\rho_{\rm cr}(a)\Omega_{\rm m}(a)\ . \tag{2.114}$$

This definition has the advantage that halos of the same mass but at different redshifts show the same amount of kinetic pressure contribution or velocity anisotropy as a function of radius, i.e., this definition is close to a dynamical definition of the virial radius. However, it requires the knowledge of the ab initio unknown cosmological parameter $\Omega_{\rm m0}$ and has the property that the halo mass increases at late times because of the redshift dilution of the mean matter density as $\rho_{\rm m} = \rho_{\rm m0}a^{-3}$ even in the absence of mass accretion! Sometimes, people prefer a redshift dependent overdensity $\Delta(a)$ from the solution of a spherical top-hat perturbation *at* the time of collapse, Eq. (2.74), rather than a constant overdensity threshold. While this property is easily calculable in simulations, the collapse time of a cluster is inaccessible in observations which jeopardises detailed comparisons of observations and theory.

- The ratio $c_{200} \equiv r_{200}/r_\mathrm{s}$ is called *concentration* of the halo . It turns out to be a function of halo mass and redshift and to depend on cosmological parameters. Generally, c_{200} is the higher the earlier halos form. This reflects the hierarchical growth of halos and implies that smaller halos form earlier when the mean background density was higher. As a result, these halos have a higher density at small scales in comparison to larger halos when radii are scaled to R_{200}. Given the halo mass M, the (virial) radius is given by Eq. (2.113), the concentration parameter gives $r_\mathrm{s} = r_{200}/c_{200}$, and the scale density ρ_s is then determined from Eq. (2.109) by the requirement that $M(r_{200}) = M_{200}$. Thus, the profile pf Eq. (2.108) is essentially determined by a single parameter, e.g. its mass.



NFW density profiles for different concentration parameters.

- This can be explicitly seen by writing down the virial mass

$$M_{200} = M(r_{200}) = 4\pi\rho_\mathrm{s} r_\mathrm{s}^3 \left[\ln(1 + c_{200}) - \frac{c_{200}}{1 + c_{200}}\right] . \qquad (2.115)$$

Setting $\rho_\mathrm{s} = \delta_\mathrm{s}\rho_\mathrm{cr}$ and using

$$r_\mathrm{s} = \frac{r_{200}}{c_{200}} = \left(\frac{GM_{200}}{c_{200}^3\, 100\, H^2}\right)^{1/3} , \qquad (2.116)$$

which follows from Eq. (2.113), enables us to characterize the scaling constant

$$\delta_\mathrm{s} = \frac{200}{3} c_{200}^3 \left[\ln(1 + c_{200}) - \frac{c_{200}}{1 + c_{200}}\right]^{-1} . \qquad (2.117)$$

Hence, the NFW profile of Eq. (2.108) can be rewritten and reads

$$\rho(r) = \frac{\delta_\mathrm{s}\rho_\mathrm{cr}}{x(1 + x)^2} , \quad x \equiv \frac{r}{r_\mathrm{s}} , \qquad (2.118)$$

where δ_s and r_s are defined in the preceding equations and we can use a power-law fit to the simulated concentration-mass relation at the scale of galaxy clusters,

$$c_{200} = 3.56 \times \left(\frac{M_{200}}{10^{15}\,\mathrm{M}_\odot}\right)^{-0.098} . \qquad (2.119)$$

Note that this relation has a substantial halo-to-halo scatter, i.e., halos at a given halo mass M_{200} show a broad distribution in c_{200} that is characterized by a large variance. NFW density profiles for different halo concentration parameters are shown in figure on the right.

- It is currently unclear how the density profile arises. Also, its slope near the core is being discussed.

Chapter 3

Evolution of the Baryonic Component

3.1 Non-radiative Physics

3.1.1 Adiabatic Processes and Entropy

- Before we enter a detailed discussion of the evolution of the baryonic component of a cluster, we review a few basic concepts. First, we start with the first law of thermodynamics (energy conservation):
$$\mathrm{d}q = T\mathrm{d}s = \mathrm{d}\epsilon + P\mathrm{d}\tilde{V}, \tag{3.1}$$
where $\tilde{V} \equiv \rho^{-1}$ is the specific volume, q is the heat per unit mass, s is the specific entropy, $\epsilon \equiv \varepsilon/\rho$ is the specific internal energy and ε is the internal energy density.

- The specific heat at constant volume
$$c_V \equiv \left(\frac{\partial q}{\partial T}\right)_V \tag{3.2}$$
is the amount of heat that must be added to raise the temperature of 1g of gas by 1K. At constant volume, the internal energy can only be changed by adding or releasing heat, $\mathrm{d}\epsilon = \mathrm{d}q$. If ϵ depends only on temperature (and not density), $\epsilon(\tilde{V}, T) = \epsilon(T)$, then
$$c_V \equiv \left(\frac{\partial q}{\partial T}\right)_V = \left(\frac{\partial \epsilon}{\partial T}\right)_V = \frac{\partial \epsilon}{\partial T} \tag{3.3}$$
implying
$$\mathrm{d}q = c_V\mathrm{d}T + P\mathrm{d}\tilde{V}. \tag{3.4}$$

- The pressure of a gas of particles with mean mass $\bar{m}$ is given by
$$P\tilde{V} = \frac{k_\mathrm{B}T}{\bar{m}} \implies P\mathrm{d}\tilde{V} = \frac{k_\mathrm{B}}{\bar{m}}\mathrm{d}T. \tag{3.5}$$
Using $\mathrm{d}q = c_V\mathrm{d}T + P\mathrm{d}\tilde{V}$, the specific heat at constant pressure is
$$c_P \equiv \left(\frac{\partial q}{\partial T}\right)_P = c_V + P\frac{\mathrm{d}\tilde{V}}{\mathrm{d}T} = c_V + \frac{k_\mathrm{B}}{\bar{m}}. \tag{3.6}$$
Changing the temperature at constant pressure requires more heat than at constant volume because some of the energy goes into $P\mathrm{d}\tilde{V}$ work.

- For reasons that become soon clear, we define the adiabatic index $\gamma = c_P/c_V$. The ionized plasma of the intra-cluster medium (ICM) is to very good approximation a monoatomic gas and has therefore 3 translational degrees of freedom. In this case, the specific energy is
$$\epsilon = \frac{3}{2}\frac{k_\mathrm{B}T}{\bar{m}} \implies c_V = \frac{3}{2}\frac{k_\mathrm{B}}{\bar{m}} \implies c_P = \frac{5}{2}\frac{k_\mathrm{B}}{\bar{m}} \implies \gamma = \frac{5}{3} \tag{3.7}$$

In general, the equation of state for an ideal gas is given by

$$\epsilon = \frac{1}{\gamma - 1}\frac{k_\mathrm{B}T}{\bar{m}} = \frac{1}{\gamma - 1}\frac{P}{\rho}. \tag{3.8}$$

- The total differential of the equation of state of an ideal gas is

$$\mathrm{d}\epsilon = \frac{1}{\gamma - 1}\left(\frac{\mathrm{d}P}{\rho} - \frac{P}{\rho^2}\mathrm{d}\rho\right). \tag{3.9}$$

For adiabatic ($\mathrm{d}q = \mathrm{d}s = 0$) changes, we can combine this with the first law of thermodynamics

$$\mathrm{d}\epsilon = -P\mathrm{d}\tilde{V} = \frac{P}{\rho^2}\mathrm{d}\rho \tag{3.10}$$

and find (after multiplying with ρ/P)

$$\frac{1}{\gamma - 1}\left(\frac{\mathrm{d}P}{P} - \frac{\mathrm{d}\rho}{\rho}\right) = \frac{\mathrm{d}\rho}{\rho}, \tag{3.11}$$

implying

$$\frac{\mathrm{d}P}{P} = \gamma\frac{\mathrm{d}\rho}{\rho} \implies P = P_0\left(\frac{\rho}{\rho_0}\right)^\gamma \equiv K\rho^\gamma. \tag{3.12}$$

- Thus, a polytropic equation of state ($P \propto \rho^\gamma$) defines the quantities

$$\begin{aligned} K &= \frac{P}{\rho^\gamma} = \frac{k_\mathrm{B}T}{\bar{m}\rho^{\gamma-1}}, \text{ and} & (3.13)\\ K_\mathrm{e} &= \frac{k_\mathrm{B}T_\mathrm{e}}{n_\mathrm{e}^{\gamma-1}} \propto K & (3.14)\end{aligned}$$

which are constants upon adiabatic changes. In the context of galaxy clusters, they are often referred to as “entropy”. This is especially the case for the observationally motivated K_e which can be conveniently computed with the X-ray observables $k_\mathrm{B}T_\mathrm{e}$ and n_e and has typical values of

$$K_\mathrm{e} \sim 100\left(\frac{k_\mathrm{B}T_\mathrm{e}}{1\ \mathrm{keV}}\right)\left(\frac{n_\mathrm{e}}{10^{-3}\ \mathrm{cm}^{-3}}\right)^{-2/3}\ \mathrm{keV\ cm}^2. \tag{3.15}$$

- To relate this cluster “entropy” to the proper thermodynamic entropy, we start with

$$\mathrm{d}\epsilon = -P\mathrm{d}\tilde{V} + T\mathrm{d}s \tag{3.16}$$

and consider adding or removing heat at constant ρ ($\mathrm{d}\tilde{V} = 0$),

$$T\mathrm{d}s = \mathrm{d}\epsilon = c_V\mathrm{d}T \tag{3.17}$$

implying

$$\mathrm{d}s = c_V\frac{\mathrm{d}T}{T} \implies s = c_V\ln T + \text{const.} \tag{3.18}$$

- Since $P \propto T$ at constant ρ, this implies that $s = c_V \ln P + \text{const.}$ We have just shown that adiabatic changes keep $P\rho^{-\gamma}$ constant, so these must be lines of constant entropy (in the P-ρ plane). Hence, for a single species gas, we can add a term $c_V \ln(\rho^{-\gamma})$ to the thermodynamic entropy s, which is a constant at constant ρ, but which ensures that the resulting expression for s is invariant for adiabatic changes:

$$s = c_V \ln(P\rho^{-\gamma}) + \text{const.} = \frac{k_\mathrm{B}}{(\gamma - 1)\bar{m}} \ln K + \text{const.} \tag{3.19}$$

or

$$s = c_V \ln\left(\frac{K}{K_0}\right) \quad \Longleftrightarrow \quad K = K_0 \exp\left(\frac{s}{c_V}\right). \tag{3.20}$$

3.1.2 Basic Conservation Equations

3.1.2.1 Preliminaries

- A physical system can be described at different levels: with quantum physics, at the classical particle level, or in the continuous fluid level. A fluid is a macroscopic description of a physical system that is characterized by its mass density ρ, temperature T, and velocity $\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w}$. Here $\boldsymbol{v} \equiv \langle \boldsymbol{u} \rangle$ is the mean velocity in the local fluid element and $\boldsymbol{w}$ is the random velocity component that defines the temperature. The equipartition theorem of classical mechanics states that each degree of freedom i that can be excited has energy $k_\mathrm{B}T/2$:

$$\left\langle \frac{1}{2}\bar{m}w_i^2 \right\rangle = \frac{1}{2}k_\mathrm{B}T \implies \left\langle |\boldsymbol{w}|^2 \right\rangle = \frac{3k_\mathrm{B}T}{\bar{m}}. \tag{3.21}$$

Fluid elements move and change their density and temperature, but particles random walk from one fluid element to another only slowly, through a diffusion process.

- A system can be well described as a fluid, if the particle mean free path is much shorter than the characteristic system size, $\lambda_\mathrm{mfp} \ll L$. In an ionized plasma, the electron's effective interaction radius r_e is to order of magnitude given by an energy balance between the electrostatic potential of an ion of charge Ze and the thermal energy of an electron:

$$\frac{Ze^2}{r_\mathrm{e}} \sim m_\mathrm{e} w_\mathrm{e}^2 \sim k_\mathrm{B} T_\mathrm{e}, \tag{3.22}$$

where e and m_e are the electron charge and mass, respectively. In the (unphysical) case of dominating large-angle electron scatterings, the electron mean free path is given by

$$\begin{aligned}\lambda_{\mathrm{mfp}} &= \frac{1}{n\sigma} \sim \frac{1}{n\pi r_{\mathrm{e}}^2} \sim \frac{1}{\pi n}\left(\frac{k_{\mathrm{B}}T_{\mathrm{e}}}{Ze^2}\right)^2 &(3.23)\\ &\sim 5.4\times 10^{23}\left(\frac{n}{10^{-3}\ \mathrm{cm}^{-3}}\right)^{-1}\left(\frac{k_{\mathrm{B}}T_{\mathrm{e}}}{6\ \mathrm{keV}}\right)^2\ \mathrm{cm}, &(3.24)\end{aligned}$$

where we have assumed $Z = 1$. This would imply an enormous electron mean free path of 175 kpc in massive galaxy clusters.

- Careful calculations (which we will perform in Section 3.2.5) yield a result that is shorter by $\sim (\ln\Lambda)^{-1}$ because of the effects of distant interactions. Here, $\Lambda = b_{\max}/b_{\min}$ is the ratio of the largest-to-smallest impact parameter of such an interaction. The typical impact parameter in a large-angle deflection constitutes the minimum impact parameter, $b_{\min} \sim r_{\mathrm{e}}$, and $b_{\max} \sim \lambda_{\mathrm{D}}$ since the plasma becomes neutral on scales larger than the Debye length, $\lambda_{\mathrm{D}} = \sqrt{k_{\mathrm{B}}T/(4\pi n_{\mathrm{e}}Ze^2)}$. Hence, we obtain to order of magnitude

$$\begin{aligned}\ln\Lambda &\sim \ln\frac{\lambda_{\mathrm{D}}}{r_{\mathrm{e}}} \sim \ln\sqrt{\frac{(k_{\mathrm{B}}T)^3}{n_{\mathrm{e}}4\pi Z^3e^6}}\\ &\sim 35 - \frac{1}{2}\ln\left(\frac{n_{\mathrm{e}}}{10^{-2}\mathrm{cm}^{-3}}\right) + \frac{3}{2}\ln\left(\frac{k_{\mathrm{B}}T}{\mathrm{keV}}\right) &(3.25)\\ &\sim 39 - \frac{1}{2}\ln\left(\frac{n_{\mathrm{e}}}{10^{-3}\mathrm{cm}^{-3}}\right) + \frac{3}{2}\ln\left(\frac{k_{\mathrm{B}}T}{6\ \mathrm{keV}}\right). &(3.26)\end{aligned}$$

- This implies an electron mean free path of

$$\begin{aligned}\lambda_{\mathrm{mfp}} &\sim \frac{1}{\pi n\ln\Lambda}\left(\frac{k_{\mathrm{B}}T_{\mathrm{e}}}{Ze^2}\right)^2 &(3.27)\\ &\sim 1.4\times 10^{22}\left(\frac{n}{10^{-3}\ \mathrm{cm}^{-3}}\right)^{-1}\left(\frac{k_{\mathrm{B}}T_{\mathrm{e}}}{6\ \mathrm{keV}}\right)^2\ \mathrm{cm}, &(3.28)\end{aligned}$$

which is of order 4.6 kpc. Hence, $\lambda_{\mathrm{mfp}} \ll L_{\mathrm{cluster}}$ and we can treat the bulk of the ICM as a fluid. However, on small scales or toward the cluster outskirts, this is not true and we have to consider plasma processes.

3.1.2.2 The Distribution Function and the Boltzmann Equation

- We define the distribution function $f(\boldsymbol{x},\boldsymbol{u},t)$ such that $f(\boldsymbol{x},\boldsymbol{u},t)\mathrm{d}^3x\mathrm{d}^3u$ is the probability of finding a particle in the phase space volume $\mathrm{d}^3x\mathrm{d}^3u$ centered on position $\boldsymbol{x}$, velocity $\boldsymbol{u}$ at time t. Integrating over all phase space yields the total number of particles

$$N = \int\int f(\boldsymbol{x},\boldsymbol{u},t)\mathrm{d}^3x\mathrm{d}^3u. \qquad (3.29)$$

Since particles are neither created nor destroyed, continuity implies

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{x}, \boldsymbol{u}, t) = \frac{\partial f}{\partial t} + \dot{\boldsymbol{x}} \cdot \boldsymbol{\nabla} f + \dot{\boldsymbol{u}} \cdot \boldsymbol{\nabla}_u f = \left.\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\mathrm{c}}, \tag{3.30}$$

where the term $\mathrm{d}f/\mathrm{d}t|_{\mathrm{c}}$ represents discontinuous motions of particles through phase space as a result of collisions. While collisions happen at a fixed point in space, they can instantaneously change particle velocities and thus cause particles to jump in phase space. Substitution $\dot{\boldsymbol{x}} = \boldsymbol{u}$ and $\dot{\boldsymbol{u}} = \boldsymbol{g}$ leads to the Boltzmann equation

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{x}, \boldsymbol{u}, t) = \frac{\partial f}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} f + \boldsymbol{g} \cdot \boldsymbol{\nabla}_u f = \left.\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\mathrm{c}} \tag{3.31}$$

which describes the evolution of the phase space distribution function $f(\boldsymbol{x}, \boldsymbol{u}, t)$.

- In the fluid limit ($\lambda_{\mathrm{mfp}} \ll L$), the collision term causes $f(\boldsymbol{u})$ to approach a Maxwellian velocity distribution while locally conserving mass, momentum, and energy. This property allows us to coarse grain the information in phase space and distill the essential pieces of information, namely how density, mean velocity, and velocity dispersion change as a function of $\boldsymbol{x}$ and t. In practice, this is done by taking the appropriate moments of the Boltzmann equation and integrating over velocity space, d^3u. We identify the mass density

$$\rho = \rho(\boldsymbol{x}, t) = \int \bar{m} f(\boldsymbol{x}, \boldsymbol{u}, t)\, \mathrm{d}^3 u. \tag{3.32}$$

The mass-weighted average of some quantity q at position $\boldsymbol{x}$ is given by

$$\langle q \rangle = \frac{1}{\rho} \int q \bar{m} f(\boldsymbol{x}, \boldsymbol{u}, t)\, \mathrm{d}^3 u. \tag{3.33}$$

3.1.2.3 Mass Conservation – Continuity Equation

- We multiply Eq. (3.31) by $\bar{m}$ and integrate over d^3u to get

$$\frac{\partial}{\partial t} \int \bar{m} f \mathrm{d}^3 u + \sum_{i=1}^{3} \frac{\partial}{\partial x_i} \int \bar{m} f u_i \mathrm{d}^3 u + \bar{m} \int \sum_{i=1}^{3} \frac{\partial}{\partial u_i} (g_i f) \mathrm{d}^3 u = \int \bar{m} \left.\frac{\partial f}{\partial t}\right|_{\mathrm{c}} \mathrm{d}^3 u. \tag{3.34}$$

Here we assume that the force $\boldsymbol{g}$ does not depend on velocity $\boldsymbol{u}$.

- Using the definitions of Eqs. (3.32) and (3.33) and applying Gauss' divergence theorem, this simplifies to

$$\frac{\partial}{\partial t}\rho(\boldsymbol{x},t)+\sum_{i=1}^{3}\frac{\partial}{\partial x_i}(\rho\langle u_i\rangle)+\bar{m}\int_{\partial\Omega} f(\boldsymbol{g}\cdot\boldsymbol{n})\mathrm{d}^2A_u=0, \quad (3.35)$$

  where $\boldsymbol{n}$ is the normal vector of the differential surface element d^2A_u. The right-hand side vanishes because of local mass conservation: collisions do not create or destroy particles at a fixed position, they can only discontinuously shift them in velocity space. Since the velocity can be split into a mean and a random component, $\boldsymbol{u}=\boldsymbol{v}+\boldsymbol{w}$, we have $\langle\boldsymbol{u}\rangle\equiv\boldsymbol{v}$ in the second term. Assuming that $f\to 0$ for $|\boldsymbol{u}|\to\infty$, the third term also vanishes on taking the limit of the integration boundary to infinity. We hence obtain

$$\frac{\partial\rho}{\partial t}+\boldsymbol{\nabla}\cdot(\rho\boldsymbol{v})=0. \quad (3.36)$$

- Taking the volume integral extending over the entire space and applying Gauss' divergence theorem, we can demonstrate that the total mass M of the system is conserved:

$$\begin{aligned}\frac{\partial}{\partial t}\int_\Omega \rho\mathrm{d}^3x+\int_\Omega\boldsymbol{\nabla}\cdot(\rho\boldsymbol{v})\mathrm{d}^3x &= 0,\\ \frac{\partial M}{\partial t}+\lim_{\partial\Omega\to\infty}\int_{\partial\Omega}\rho(\boldsymbol{v}\cdot\boldsymbol{n})\mathrm{d}^2A &= \frac{\mathrm{d}M}{\mathrm{d}t}=0. \quad (3.37)\end{aligned}$$

  In the last line, we exchanged the total for the partial time derivatives since M depends neither on position nor on velocity.

3.1.2.4 Momentum Conservation

- We multiply Eq. (3.31) by $\bar{m}\boldsymbol{u}$ and integrate over d^3u to get

$$\begin{aligned}\frac{\partial}{\partial t}\int\bar{m}u_jf\mathrm{d}^3u+\sum_{i=1}^{3}\frac{\partial}{\partial x_i}\int\bar{m}fu_ju_i\mathrm{d}^3u&\\ +\bar{m}\int\sum_{i=1}^{3}g_iu_j\frac{\partial f}{\partial u_i}\mathrm{d}^3u=\int\bar{m}u_j\left.\frac{\partial f}{\partial t}\right|_{\mathrm{c}}d^3u.&\end{aligned} \quad (3.38)$$

- The first term is $\partial(\rho v_j)/\partial t$ and second term is

$$\sum_{i=1}^{3}\frac{\partial}{\partial x_i}(\rho\langle u_ju_i\rangle)=\sum_{i=1}^{3}\frac{\partial}{\partial x_i}(\rho v_jv_i+\rho\langle w_jw_i\rangle). \quad (3.39)$$

  To simplify the third term, we use the identity

$$\frac{\partial}{\partial u_i}(u_jf)\equiv u_j\frac{\partial f}{\partial u_i}+\delta_{ij}f \quad (3.40)$$

and obtain

$$\sum_{i=1}^{3} \bar{m} g_i \int \left[\frac{\partial}{\partial u_i}(u_j f) - \delta_{ij} f \right] d^3u = -\sum_{i=1}^{3} g_i \delta_{ij} \int \bar{m} f \mathrm{d}^3 u = -\rho g_j, \tag{3.41}$$

where $\delta_{ij} \equiv 1$ for $i = j$, 0 for $i \neq j$ and the first term in the bracket vanishes because of Gauss' divergence theorem. The right-hand side of Eq. (3.38) vanishes because collisions conserve momentum.

- We get the result

$$\frac{\partial}{\partial t}(\rho v_j) + \sum_{i=1}^{3} \frac{\partial}{\partial x_i}(\rho v_j v_i + \rho \langle w_j w_i \rangle) = \rho g_j. \tag{3.42}$$

The diagonal terms of $\langle w_i w_j \rangle$ are generally much larger than the off-diagonal terms since random velocities in different directions are almost uncorrelated. It is useful to split the $\rho \langle w_i w_j \rangle$ term into an (isotropic) contribution from pressure, P, and a contribution from the anisotropic viscous stress tensor, $\bar{\mathbf{\Pi}}$,

$$P \equiv \frac{1}{3} \rho \langle |\boldsymbol{w}|^2 \rangle, \tag{3.43}$$

$$\Pi_{ij} \equiv P \delta_{ij} - \rho \langle w_i w_j \rangle, \tag{3.44}$$

which is a symmetric and traceless tensor.

- The final result is the Navier-Stokes equation

$$\frac{\partial}{\partial t}(\rho v_j) + \sum_{i=1}^{3} \frac{\partial}{\partial x_i} \left(\rho v_i v_j + P \delta_{ij} - \Pi_{ij} \right) = \rho g_j, \text{ or} \tag{3.45}$$

$$\frac{\partial}{\partial t}(\rho \boldsymbol{v}) + \boldsymbol{\nabla} \cdot \left(\rho \boldsymbol{v}\boldsymbol{v} + P \bar{\mathbf{1}} - \bar{\mathbf{\Pi}} \right) = \rho \boldsymbol{g}. \tag{3.46}$$

Taking the volume integral and applying Gauss' divergence theorem (as in Eq. 3.37), we can demonstrate that the total momentum, $\boldsymbol{p} = \int \rho \boldsymbol{v} \mathrm{d}^3 x$ is conserved in the absence of an external force field that acts as a source of momentum.

- To simplify this equation, we rewrite the first two terms in Eq. (3.46),

$$\begin{aligned} & \frac{\partial}{\partial t}(\rho \boldsymbol{v}) + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{v}\boldsymbol{v}) \\ = \; & \dot{\rho} \boldsymbol{v} + \rho \dot{\boldsymbol{v}} + \rho \boldsymbol{v}(\boldsymbol{\nabla} \cdot \boldsymbol{v}) + \rho (\boldsymbol{v} \cdot \nabla) \boldsymbol{v} + \boldsymbol{v}(\boldsymbol{v} \cdot \boldsymbol{\nabla} \rho) \\ = \; & \boldsymbol{v} \left[\dot{\rho} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{v}) \right] + \rho \left[\dot{\boldsymbol{v}} + (\boldsymbol{v} \cdot \boldsymbol{\nabla}) \boldsymbol{v} \right] = \rho \left[\dot{\boldsymbol{v}} + (\boldsymbol{v} \cdot \boldsymbol{\nabla}) \boldsymbol{v} \right]. \end{aligned} \tag{3.47}$$

In the last step, we have used the continuity equation (3.36).

- If we use this result in combination with the momentum equation (3.46), we get

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v} = \boldsymbol{g} - \frac{1}{\rho}\boldsymbol{\nabla} P + \frac{1}{\rho}\boldsymbol{\nabla}\cdot\bar{\boldsymbol{\Pi}}. \tag{3.48}$$

  Viscosity acts to oppose shearing motion and interpenetration.

- To make progress, we adopt an ansatz for the viscous stress tensor and assume a "Newtonian fluid", i.e., we assume that Π_{ij} is linearly proportional to the velocity gradient, $\partial v_i/\partial x_j$. The most general symmetric tensor that is linear in $\partial v_i/\partial x_j$ is

$$\Pi_{ij} = \eta D_{ij} + \xi\delta_{ij}(\boldsymbol{\nabla}\cdot\boldsymbol{v}), \text{ where} \tag{3.49}$$

$$D_{ij} = \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - \frac{2}{3}\delta_{ij}(\boldsymbol{\nabla}\cdot\boldsymbol{v}) \tag{3.50}$$

  is the deformation tensor that vanishes for uniform expansion or contraction. η and ξ are the coefficients of shear and bulk viscosity, respectively and have units of g cm^{-1} s^{-1}. The term ηD_{ij} represents resistance to shearing motion and $\xi\delta_{ij}(\boldsymbol{\nabla}\cdot\boldsymbol{v})$ represents resistance to changes in volume.

3.1.2.5 Energy Conservation

- To obtain the internal energy equation, we multiply Eq. (3.31) by $\bar{m}\boldsymbol{u}^2$ and integrate over d^3u. Making use of Gauss' divergence theorem and of the fact that collisions conserve energy (as well as mass and momentum) we get (after a similar procedure)

$$\begin{aligned}\rho\frac{\mathrm{d}\epsilon}{\mathrm{d}t} &= \rho\left(\frac{\partial\epsilon}{\partial t} + \boldsymbol{v}\cdot\boldsymbol{\nabla}\epsilon\right)\\ &= \frac{\partial}{\partial t}(\rho\epsilon) + \boldsymbol{\nabla}\cdot(\rho\epsilon\boldsymbol{v}) = -P\boldsymbol{\nabla}\cdot\boldsymbol{v} + \Psi - \boldsymbol{\nabla}\cdot\boldsymbol{Q}.\end{aligned} \tag{3.51}$$

  In the second step, we have used the continuity equation (3.36). Here, ϵ is the specific internal energy (defined in Eq. 3.8), Ψ is the viscous dissipation rate, and $\boldsymbol{Q}$ is the conductive heat flux:

$$\epsilon \equiv \frac{1}{2}\left\langle|\boldsymbol{w}|^2\right\rangle, \tag{3.52}$$

$$\Psi \equiv \sum_{i,j=1}^{3}\Pi_{ij}\frac{\partial v_i}{\partial x_j} = \bar{\boldsymbol{\Pi}} : \boldsymbol{\nabla}\boldsymbol{v}, \tag{3.53}$$

$$\boldsymbol{Q} \equiv \frac{1}{2}\rho\left\langle\boldsymbol{w}|\boldsymbol{w}|^2\right\rangle. \tag{3.54}$$

  Ψ represents conversion of bulk motion of the fluid into internal energy via viscous dissipation. It is the viscous analog of heating by $P\mathrm{d}V$ work.

- If the distribution of the random velocity component, $\boldsymbol{w}$, is symmetric about zero, then $\boldsymbol{Q}$ vanishes. If the distribution is skewed, then hot particles drift relative to cold particles and produce a heat flux in the direction of the drift. In most cases, a temperature gradient produces a conductive flux (see also Section 3.2.5),

  $$\boldsymbol{Q} = -\chi \boldsymbol{\nabla} T, \quad \text{with } \chi \simeq 6 \times 10^{-7}\, T^{5/2}\, \text{erg s}^{-1}\text{cm}^{-1}\text{K}^{-1}. \tag{3.55}$$

  We will also use $\kappa = \chi T/P$ instead of χ for convenience (κ has units of cm^2 s^{-1}, i.e., of a diffusion coefficient). However, if $\boldsymbol{Q}$ is uniform, heat flowing out is replaced by heat flowing in. Hence, the local thermal energy changes only if $\boldsymbol{\nabla} \cdot \boldsymbol{Q} \neq 0$.

- One can derive an equivalent equation for the entropy by considering the first law of thermodynamics,

  $$\mathrm{d}\epsilon = -P\mathrm{d}\tilde{V} + T\mathrm{d}s = \frac{P}{\rho^2}\mathrm{d}\rho + T\mathrm{d}s. \tag{3.56}$$

  Combining this with the internal energy equation (3.51), we get

  $$\frac{P}{\rho^2}\frac{\mathrm{d}\rho}{\mathrm{d}t} + T\frac{\mathrm{d}s}{\mathrm{d}t} = -\frac{P}{\rho}\boldsymbol{\nabla}\cdot\boldsymbol{v} - \frac{1}{\rho}\boldsymbol{\nabla}\cdot\boldsymbol{Q} + \frac{1}{\rho}\Psi. \tag{3.57}$$

  Using the continuity equation, $\mathrm{d}\rho/\mathrm{d}t = -\rho\boldsymbol{\nabla}\cdot\boldsymbol{v}$, the first terms on both sides cancel each other and we obtain,

  $$\rho T\frac{\mathrm{d}s}{\mathrm{d}t} = -\boldsymbol{\nabla}\cdot\boldsymbol{Q} + \Psi. \tag{3.58}$$

  This shows explicitly, that heat conduction and viscous friction change the entropy. Equivalently, if these processes are absent, specific entropy is conserved.

3.1.3 Buoyancy Instabilities

- We are now interested in studying adiabatic hydrodynamic perturbations about an atmosphere in hydrostatic equilibrium. The starting point are conservation equations of mass, momentum, and internal energy (or equivalently entropy) in the absence of viscosity and magnetic fields:

  $$\begin{aligned}
  \frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\boldsymbol{v}) &= 0, && (3.59)\\
  \frac{\partial\boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v} &= -\frac{\boldsymbol{\nabla}P}{\rho} + \boldsymbol{g}, && (3.60)\\
  \rho T\frac{\mathrm{d}s}{\mathrm{d}t} &= -\boldsymbol{\nabla}\cdot\boldsymbol{Q}, && (3.61)
  \end{aligned}$$

  where $\rho(t,\boldsymbol{x})$ and $\boldsymbol{v}(t,\boldsymbol{x})$ are the density and velocity of the cosmic fluid at position $\boldsymbol{x}$ and time t, $\boldsymbol{g}$ is a conservative force field per unit mass (such as the gravitational acceleration), P is the thermal pressure, T is the temperature, s is the entropy per unit mass and $\mathrm{d}/\mathrm{d}t = \partial/\partial t + \boldsymbol{v}\cdot\boldsymbol{\nabla}$ is a Lagrangian time derivative.

- In linear theory, there are two families of modes corresponding to whether pressure or gravity acts as the restoring force. This becomes clear from the right-hand side of the momentum equation (3.60). The pressure modes represent the familiar sound waves which propagate at the sound speed, $c_\mathrm{s} = \sqrt{\gamma P/\rho}$, where P is the thermal pressure, ρ is the mass density, and γ is the ratio of specific heat capacities. The pressure modes radiate out of any given region on the sound crossing time.

- Gravity modes have the important property that they can be trapped. They can induce vortical motions that feed into a turbulent cascade with important consequences for the thermal cluster state. To see this in detail, we will now derive the dispersion relation of a local g-mode. We will carry out a Wentzel-Kramers-Brillouin (WKB) perturbation analysis on a hydrostatic background. To this end, we assume that the background plasma is thermally stratified in the presence of a uniform gravitational field in the vertical direction, $\boldsymbol{g} = -g\boldsymbol{e}_z$ ($\boldsymbol{e}_z$ is the unit vector in the z direction). Hence, the force balance implies $\mathrm{d}P_0/\mathrm{d}z = -\rho_0 g$ and $\boldsymbol{v}_0 = \boldsymbol{0}$ (the subscript 0 denotes background quantities). The background heat flux $\boldsymbol{Q}_0 = -\chi\boldsymbol{e}_z\mathrm{d}T_0/\mathrm{d}z$ requires us to demand $\boldsymbol{\nabla}\cdot\boldsymbol{Q}_0 = 0$ in order for the initial equilibrium to be in steady state which implies a temperature that varies at most linearly with height. Although this steady state assumption is formally required, we note that as long as the time scale for the evolution of the system is longer than the local dynamical time, the general features of the instability described here are unlikely to depend critically on this steady state assumption.

- To proceed, we perturb the stratified plasma and split the dynamical quantities into background values and small perturbations: $\rho = \rho_0 + \delta\rho$, $\boldsymbol{v} = \delta\boldsymbol{v}$, $P = P_0 + \delta P$, and $s = s_0 + \delta s$. In a first step, we evaluate the time derivative of the entropy and obtain to first order in the perturbed quantities

$$\begin{aligned}\frac{\partial s}{\partial t} &= \frac{1}{\gamma-1}\frac{k_\mathrm{B}}{\bar{m}}\frac{\partial(\ln P\rho^{-\gamma})}{\partial t}\\ &= \frac{1}{\gamma-1}\frac{k_\mathrm{B}}{\bar{m}}\left(\frac{1}{P_0}\frac{\partial\delta P}{\partial t} - \frac{\gamma}{\rho_0}\frac{\partial\delta\rho}{\partial t}\right).\end{aligned}\tag{3.62}$$

- Next, we insert the expressions for $\rho = \rho_0 + \delta\rho$, etc. into the conservation equations of mass, momentum, and entropy (Eqs. 3.59 to 3.61) and obtain the perturbation equations to first order in the perturbed quantities:

$$\frac{\partial \delta\rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho_0 \delta\boldsymbol{v}) = 0, \tag{3.63}$$

$$\frac{\partial \delta\boldsymbol{v}}{\partial t} - \frac{\delta\rho}{\rho_0^2}\boldsymbol{\nabla} P_0 + \frac{\boldsymbol{\nabla}\delta P}{\rho_0} = 0, \tag{3.64}$$

$$\frac{1}{\gamma-1}\left(\frac{\partial \delta P}{\partial t} - \frac{\gamma k_{\mathrm{B}} T_0}{\bar{m}}\frac{\partial\delta\rho}{\partial t}\right) + \rho_0 T_0 (\delta\boldsymbol{v}\cdot\boldsymbol{\nabla}) s_0 = -\boldsymbol{\nabla}\cdot\delta\boldsymbol{Q} \tag{3.65}$$

where we have used Eq. (3.62) and $\boldsymbol{g} = \boldsymbol{\nabla} P_0/\rho_0$ in Eq. (3.64).

- We can decompose all dynamical variables ($\delta\rho$, $\delta\boldsymbol{v}$, δs, δP, $\delta\boldsymbol{Q}$) into plane waves,

$$\delta\rho(\boldsymbol{x}, t) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\, \delta\hat{\rho}(\boldsymbol{k}, \omega)\, \mathrm{e}^{-\mathrm{i}\omega t + \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{3.66}$$

  introducing the Fourier amplitudes $\delta\hat{\rho}(\boldsymbol{k}, \omega)$ which obey algebraic equations rather than partial differential equations. Formally, plane waves form an orthonormal system on a homogeneous background. As long as the perturbations of our stratified hydrostatic background are small, a decomposition into plane waves is complete. However, as we will see, the growth rate of the perturbations depends in general on position (i.e., height in the gravitational potential), which renders this approach inaccurate after some time because the wave vector will start to depend on position and different wave vectors are not any more linearly independent.

- The wave vector is defined as $\boldsymbol{k} = k_x \boldsymbol{e}_x + k_y \boldsymbol{e}_y + k_z \boldsymbol{e}_z$ and we define $k_\perp^2 = k_x^2 + k_y^2$ to be the wave vector perpendicular to the local gravitational field. The *WKB assumption* requires $kH \gg 1$, where H is the local scale height of the system (which is given) and $k = |\boldsymbol{k}|$. Hence you have to choose the wave lengths (and wave numbers) that are on small enough scales (relative to the cluster scale height) for the overall stratification not to matter!

- We get

$$\begin{aligned} -\mathrm{i}\omega\delta\hat{\rho} + (\delta\hat{\boldsymbol{v}}\cdot\boldsymbol{\nabla})\rho_0 + \mathrm{i}\rho_0 \boldsymbol{k}\cdot\delta\hat{\boldsymbol{v}} &= 0, & (3.67)\\ -\mathrm{i}\omega\delta\hat{\boldsymbol{v}} - \frac{\delta\hat{\rho}}{\rho_0^2}\boldsymbol{\nabla} P_0 + \mathrm{i}\boldsymbol{k}\frac{\delta\hat{P}}{\rho_0} &= 0, & (3.68)\\ \mathrm{i}\omega\frac{\gamma}{\gamma-1}P_0\frac{\delta\hat{\rho}}{\rho_0} + \rho_0 T_0(\delta\hat{\boldsymbol{v}}\cdot\boldsymbol{\nabla})s_0 &= -\mathrm{i}\boldsymbol{k}\cdot\delta\hat{\boldsymbol{Q}} & (3.69)\end{aligned}$$

To derive the third equation, we used the *Boussinesq approximation* which filters out time scales faster than the sound crossing time. Comparing the Fourier transform of the two terms in the parenthesis of Eq. (3.65) and using the dispersion relation of sound waves, the Boussinesq approximation requires

$$\delta\hat{P}=\frac{\omega^2}{k^2}\delta\hat{\rho}\overset{!}{\ll}c_\mathrm{s}^2\delta\hat{\rho}=\frac{\gamma k_\mathrm{B}T_0}{\bar{m}}\delta\hat{\rho}\quad\Rightarrow\quad\omega\overset{!}{\ll}kc_\mathrm{s}. \tag{3.70}$$

This implies that we effectively drop the δP term in the energy equation (but not in the momentum equation).

- We define a displacement vector $\boldsymbol{\xi}=\mathrm{i}\delta\hat{\boldsymbol{v}}/\omega$ (with $\xi=|\boldsymbol{\xi}|$) and use this to perform an order of magnitude analysis of the terms in Eq. (3.67):

$$\begin{array}{ccccccc}\dfrac{\delta\hat{\rho}}{\rho_0} & + & \dfrac{1}{\rho_0}(\boldsymbol{\xi}\cdot\boldsymbol{\nabla})\rho_0 & - & \dfrac{\boldsymbol{k}\cdot\delta\hat{\boldsymbol{v}}}{\omega} & = & 0\\ \dfrac{\delta\hat{\rho}}{\rho_0} & \sim & \dfrac{\xi}{H} & \ll & k\xi & \text{since} & kH\gg 1\end{array} \tag{3.71}$$

which follows from the WKB approximation. Thus, the last term dominates over the first two terms and leaves us with the near incompressibility condition:

$$\boldsymbol{k}\cdot\delta\hat{\boldsymbol{v}}=0\quad\Rightarrow\quad k_\perp\delta\hat{v}_\perp+k_z\delta\hat{v}_z=0 \tag{3.72}$$

- As we will now see, this conditions enables us to project the momentum equation (3.68) into a purely vertical and perpendicular equation, respectively:

$$-\mathrm{i}\omega\boldsymbol{k}\cdot\delta\hat{\boldsymbol{v}}=0 = -\frac{\delta\hat{\rho}}{\rho_0}gk_z-\mathrm{i}k^2\frac{\delta\hat{P}}{\rho_0}, \tag{3.73}$$

$$\frac{\delta\hat{P}}{\rho_0} = \omega\frac{\delta\hat{v}_\perp}{k_\perp}. \tag{3.74}$$

- Combining Eqs. (3.69), (3.73), (3.74) and neglecting heat flux perturbations for simplicity ($\boldsymbol{k}\cdot\delta\hat{\boldsymbol{Q}}=0$), we obtain the dispersion relation for gravity waves,

$$\omega^2=N^2\frac{k_\perp^2}{k^2},\qquad N^2=\frac{g}{\gamma}\frac{\partial\ln K}{\partial z}. \tag{3.75}$$

Here, $K=P\rho^{-\gamma}$ and N denotes the *Brunt-Väisälä frequency*.

- This dispersion relation has two important consequences:

  1. For a stably stratified atmosphere where the entropy is increasing outward ($\partial s/\partial z>0$ or $\partial K/\partial z>0$), ω is positive and the displaced fluid parcel oscillates with the Brunt-Väisälä frequency around the equilibrium position. If the entropy is decreasing outward, we have an unstable situation: displacing high-entropy gas in such an atmosphere upwards causes it to rise further until the entropy profile is inverted and stably stratified, defining a new equilibrium.

2. Since $k_\perp \leq |\boldsymbol{k}|$, g-modes have a maximum possible frequency of $\omega_{\max} = N$ at which point $k_\perp = |\boldsymbol{k}|$ and $k_z = 0$. If the Brunt-Väisälä frequency is a decreasing function of height z, g-modes of a given frequency ω will be confined/trapped below the height at which $N(z) = \omega$.

The first result can also be obtained by thermodynamic considerations only and is known as the Schwarzschild criterion for convective instability. We show its derivation for completeness in Appendix A.2.

3.1.4 Vorticity

- Among other sources, g-modes can also generate vorticity that decays to smaller eddies and drives a turbulent cascade, as we will see. To lowest order in the (small) Mach number ($\mathcal{M} = v/c_\mathrm{s}$) of the velocity perturbations, the flow is incompressible, $\boldsymbol{\nabla}\cdot\boldsymbol{v} = 0$. This implies that the velocity field in this approximation is a pure vortex field, $\boldsymbol{v} = \boldsymbol{\nabla}\times\boldsymbol{A}$ (where $\boldsymbol{A}$ is a vector potential). In other words, the excitation of g-modes leads to the generation of vorticity, $\boldsymbol{\omega} = \boldsymbol{\nabla}\times\boldsymbol{v}$. We will now derive an evolution equation for vorticity in the case of an ideal inviscid fluid, which has viscous forces that are much smaller than inertial forces. We only consider conservative external forces per unit mass, $\boldsymbol{g} = -\boldsymbol{\nabla}\Phi$.

- If we apply the curl operator to the momentum equation (3.60) and adopt the definition of vorticity, $\boldsymbol{\omega} = \boldsymbol{\nabla}\times\boldsymbol{v}$, we obtain

$$\frac{\partial\boldsymbol{\omega}}{\partial t} + \boldsymbol{\nabla}\times[(\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v}] = \frac{1}{\rho^2}\boldsymbol{\nabla}\rho\times\boldsymbol{\nabla}P \tag{3.76}$$

since $\boldsymbol{\nabla}\times\nabla\phi \equiv 0$ where ϕ is a scalar field. Using the identity

$$(\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v} \equiv \frac{1}{2}\boldsymbol{\nabla}(\boldsymbol{v}^2) - \boldsymbol{v}\times\boldsymbol{\omega}, \tag{3.77}$$

we can rewrite the second term in Eq. (3.76) as follows:

$$\begin{aligned}\boldsymbol{\nabla}\times[(\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v}] &= -\boldsymbol{\nabla}\times(\boldsymbol{v}\times\boldsymbol{\omega}) \\ &= -(\boldsymbol{\omega}\cdot\boldsymbol{\nabla})\boldsymbol{v} + \boldsymbol{\omega}(\boldsymbol{\nabla}\cdot\boldsymbol{v}) + (\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{\omega} \qquad (3.78)\end{aligned}$$

since $\boldsymbol{\nabla}\cdot\boldsymbol{\omega} \equiv 0$.

- Hence the evolution equation for vorticity reads

$$\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} = (\boldsymbol{\omega}\cdot\boldsymbol{\nabla})\boldsymbol{v} - \boldsymbol{\omega}(\boldsymbol{\nabla}\cdot\boldsymbol{v}) + \frac{1}{\rho^2}\boldsymbol{\nabla}\rho\times\boldsymbol{\nabla}P \tag{3.79}$$

where $\mathrm{d}/\mathrm{d}t \equiv \partial/\partial t + \boldsymbol{v}\cdot\boldsymbol{\nabla}$ is the Lagrangian time derivative. In the limit of small velocities ($\mathcal{M} \ll 1$) this evolution equation reads to linear order

$$\frac{\partial\boldsymbol{\omega}}{\partial t} = \frac{1}{\rho^2}\boldsymbol{\nabla}\rho\times\boldsymbol{\nabla}P. \tag{3.80}$$

- Hence, vorticity is conserved in subsonic, polytropic ($P = P(\rho)$) flows. Conversely, vorticity production is associated with departures between surfaces of constant density and those of constant pressure. Given that the pressure gradient is essentially in the vertical direction (as defined by the local gravitational field), the term $\boldsymbol{\nabla}\rho \times \boldsymbol{\nabla}P$ and hence the generated vorticity will lie principally in the horizontal plane. This facilitates the identification of g-modes.

3.1.5 Turbulence

> "Big whirls have little whirls that feed on their velocity,
> and little whirls have lesser whirls and so on to viscosity."
> – *Lewis Fry Richardson*

- We start with the Navier-Stokes equation (3.47) for an incompressible fluid ($\boldsymbol{\nabla}\cdot\boldsymbol{v} = 0$) and obtain

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v} = \boldsymbol{g} - \frac{1}{\rho}\boldsymbol{\nabla}P + \nu\Delta\boldsymbol{v}, \tag{3.81}$$

  where we used Eqs. (3.49) and (3.50) and defined the kinematic viscosity $\nu = \eta/\rho$. From left to right, the terms have the following meaning: 1) rate of change of $\boldsymbol{v}$, 2) advective transport, 3) external force (e.g., gravity), 4) pressure force, 5) viscous dissipation term.

- To order of magnitude, the viscosity is the particle mean free path times the thermal particle velocity, $\nu \sim \lambda_{\mathrm{mfp}} v_{\mathrm{th}}$ and has the units $\mathrm{cm}^2\ \mathrm{s}^{-1}$: thermal particles moving a mean free path collide and randomize their velocities, which we call generation of heat, or dissipation of kinetic energy. It turns out that the particle mean free path is the typical length over which the fluid can communicate changes in its shear stress. A fluid with a longer mean path length therefore more easily opposes changes to its local shear velocity, i.e., is more viscous.

- We compare the time scales for advection, t_{adv}, and for viscous dissipation, t_{diss}:

$$t_{\mathrm{adv}} = \frac{L}{v} \quad \text{and} \quad t_{\mathrm{diss}} = \frac{L^2}{\nu}, \tag{3.82}$$

  where L and v are characteristic length and velocity scales of the (macroscopic) system. We define the Reynolds number to be the ratio of dissipative-to-advective time scale,

$$\mathrm{Re} = \frac{t_{\mathrm{diss}}}{t_{\mathrm{adv}}} = \frac{Lv}{\nu} = \frac{L}{\lambda_{\mathrm{mfp}}}\frac{v}{v_{\mathrm{th}}}. \tag{3.83}$$

  This shows that Re is the product of the ratios of macroscopic-to-microscopic length and velocity scales.

- Note that the assumption of an incompressible flow

$$\boldsymbol{v}(\boldsymbol{x},t) = \int \hat{\boldsymbol{v}}(\boldsymbol{k},\omega)\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}\mathrm{d}^3k\mathrm{d}\omega, \tag{3.84}$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{v} = 0 \implies \boldsymbol{k}\cdot\hat{\boldsymbol{v}} = 0 \tag{3.85}$$

  does not allow for longitudinal disturbances (sound waves), but only for rotational flows, so-called "eddies" and implies subsonic velocities (since supersonic velocities would cause the formation of shocks, which necessarily have $\boldsymbol{\nabla}\cdot\boldsymbol{v} \neq 0$).

- If $\mathrm{Re} \gg 1$, advection is much faster than dissipation which cannot stabilize the dynamical growth. The vortical fluid motions interact non-linearly and turbulence sets in. In three dimensions, energy is being fed into the turbulent cascade on the macroscopic "injection scale" L with a typical velocity v. Energy is being transported from large to small scales as large eddies break up into smaller eddies, thereby conserving vorticity in the absence of the baroclinic term. The energy transport to small scales continues until the energy is dissipated through the production of viscous heat on the microscopic "viscous" scale, λ_{visc}, which is of order the particle mean free path. The scales in between, for $\lambda_{\mathrm{visc}} < \lambda < L$, are called the "inertial range". In two dimensions, however, small eddies merge to form larger eddies and energy flows from small to large scales along an "inverse cascade".

- Let λ be the size of an eddy and v_λ the typical rotational velocity across the eddy. The energy flow through that scale is the product of kinetic energy and the eddy turnover rate on that scale,

$$\dot{\epsilon} \approx \left(\frac{v_\lambda^2}{2}\right)\left(\frac{v_\lambda}{\lambda}\right) \approx \frac{v_\lambda^3}{\lambda}. \tag{3.86}$$

  In the inertial range, the energy flow must be independent on scale, $\dot{\epsilon} = v^3/L = \text{const.}$ because energy must not accumulate anywhere in steady state: the only channel for the energy to be transferred is through non-linear interactions with other eddies and hence, we obtain the velocity scaling from Eq. (3.86)

$$\dot{\epsilon} = \frac{v^3}{L} \approx \frac{v_\lambda^3}{\lambda} \qquad \implies \qquad v_\lambda \approx v\left(\frac{\lambda}{L}\right)^{1/3}. \tag{3.87}$$

- The largest eddies assume the highest velocities (and thus the highest kinetic energies), but the smallest eddies have the highest vorticity

$$|\boldsymbol{\omega}| \approx \frac{v_\lambda}{\lambda} \approx \frac{v}{(\lambda^2 L)^{1/3}}. \tag{3.88}$$

  Since the overall vorticity is approximately conserved this implies that turbulence becomes more and more intermittent on smaller scales, i.e., less volume is filled with turbulent eddies.

- To compute the power spectrum of eddy velocity, $v_\lambda \approx (\dot{\epsilon}\lambda)^{1/3}$, we write down the correlation function which scales as

  $$\xi_v \propto v_\lambda^2 \propto (\dot{\epsilon}\lambda)^{2/3}. \tag{3.89}$$

  Note that the kinetic energy on a scale λ scales exactly as the correlation function, $\epsilon \propto v_\lambda^2 \propto \lambda^{2/3}$. The correlation function ξ_v is the Fourier transform of the velocity power spectrum, $\xi_v \propto k^3 P_v$ (see Section 2.2.1), which inherits the scaling (using $k = 2\pi/\lambda$)

  $$P_v \propto \lambda^3 \xi_v \propto k^{-3} \left(\dot{\epsilon} k^{-1}\right)^{2/3} \propto \dot{\epsilon}^{2/3} k^{-11/3}. \tag{3.90}$$

  The power per linear and logarithmic interval in k-space scale as

  $$P_v k^2 \mathrm{d}k \propto \dot{\epsilon}^{2/3} k^{-5/3} \mathrm{d}k, \quad \text{and} \tag{3.91}$$
  $$P_v k^3 \mathrm{d}\ln k \propto \dot{\epsilon}^{2/3} k^{-2/3} \mathrm{d}\ln k, \tag{3.92}$$

  which is the Kolmogorov turbulence spectrum of driven turbulence.

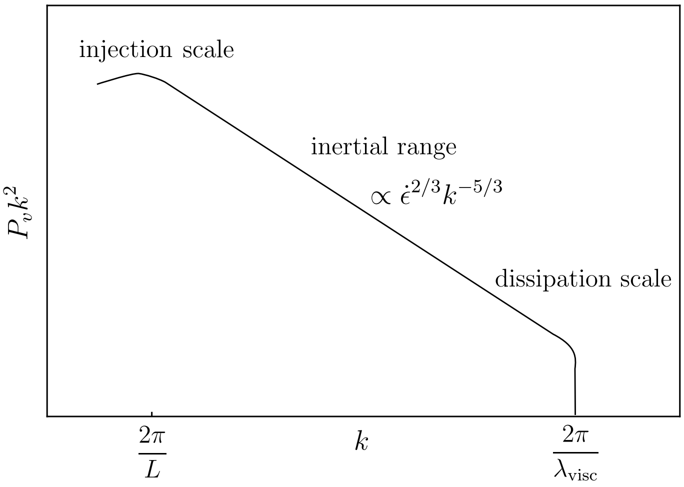


Figure 3.1: The turbulent power spectrum in Fourier space. Stirring rotational motions gives rise to large-scale eddies. The scale L at which stirring happens is called the injection scale. Non-linearly interacting eddies break up into smaller and smaller eddies in the inertial range, giving rise to a turbulent cascade until motions are dissipated at the viscous (Kolmogorov) scale $\lambda_{\rm visc}$. The power spectrum follows the Kolmogorov slope in the inertial range.

- In contrast to driven subsonic (Kolmogorov) turbulence, in clusters we encounter decaying turbulence: a merger injects kinetic energy on scales $L \sim r_{\rm c}$, which will successively decay after a few eddy turnover time scales L/v. The possible implications of turbulence in clusters is mainly

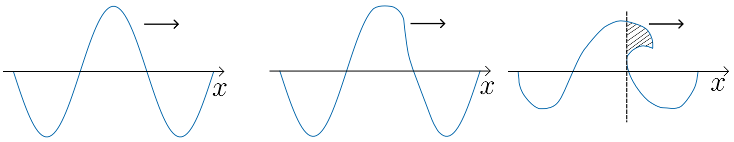
  1. mixing of metals that have been injected by galactic winds,

2. explaining the magnetization in clusters: driving a turbulent dynamo of either primordial magnetic fields or field that was injected by AGNs or galactic winds in proto-clusters,
3. shredding AGN bubbles and mixing relativistic components with the ICM: buoyantly rising bubbles (filled with light relativistic material) induce Kelvin-Helmholtz instabilities at their interface with the ICM; this feeds into a turbulent cascade that implies turbulent diffusion and energy dissipation, eventually causing bubbles to resolve,
4. heating cool cores and possible arresting the over-cooling in them, provided the coupling efficiency of $P\mathrm{d}V$ work to the turbulent cascade is high and the dissipation is volume filling and thermally stable.

3.1.6 Shocks

3.1.6.1 General Considerations

- Imagine the propagation of a sound wave with finite amplitude. The sound speed is higher at higher temperature as $c_\mathrm{s}^2 \propto k_\mathrm{B}T$, so that the crest of the wave gradually overtakes the colder trough ($T \propto \rho^{\gamma-1}$). When faster moving gas overtakes slower moving gas, we would obtain a multivalued solution that is inconsistent with the hydrodynamic equations:

  Instead, we get a discontinuous change of density and velocity, a so-called “shock”. This steepening happens even for $\gamma = 1$ because of the non-linear dependence on the velocity in the equations. Shocks can also be produced by any supersonic compressible disturbance (or through non-linear interactions of subsonic compressible modes). This can result from a supernova explosion within a galaxy, by gas accreting super-sonically onto a cluster, or if two galaxy clusters merge to form a larger entity. In general, a shock wave is

  1. propagating faster than the “signal speed” for compressible waves (i.e., the fastest linear eigenmode of the system which is the sound speed c_s in a hydrodynamic fluid or the fast magnetosonic mode in the high-beta[1] magnetohydrodynamic plasma of a galaxy cluster),

---

[1] The plasma beta factor is the thermal-to-magnetic pressure ratio, $\beta = P_\mathrm{th}/P_B$. In

2. producing an irreversible change of the fluid state, i.e., an increase in entropy, and
3. can either be caused by a pressure-driven compressive disturbance, results from non-linear wave interactions, or is caused by supersonic collisions of two streams of fluids (e.g., as a result of their mutual gravitational interactions).

- In most cases, a shock involves a "discontinuous" change of the fluid properties over a scale proportional to the effective mean free path λ_{eff}. In "collisional" shocks, the mean free path λ_{mfp} is set by Coulomb-force mediated particle-particle collisions. In a "collisionless" plasma (which is of relevance for galaxy clusters) electromagnetic waves scatter charged particles on a much faster timescale than particle-particle interactions take place and thus reduce λ_{eff} by many orders of magnitude, $\lambda_{\mathrm{eff}} \ll \lambda_{\mathrm{mfp}}$, so that we are dealing here with "collisionless" shocks.

- In order to understand general properties at fluid discontinuities, we are now considering the conservation laws of mass, momentum, and internal energy in the absence of external gravitational forces and conductive heat flux (which act on time scale that are much longer in comparison to the transition times at shocks or tangential discontinuities),

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{v}) &= 0, && (3.93)\\ \frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \boldsymbol{\nabla})\boldsymbol{v} &= -\frac{\boldsymbol{\nabla} P}{\rho} + \frac{1}{\rho}\boldsymbol{\nabla} \cdot \bar{\boldsymbol{\Pi}}, && (3.94)\\ \frac{\partial}{\partial t}(\rho\epsilon) + \boldsymbol{\nabla} \cdot (\rho\epsilon\boldsymbol{v}) &= -P\boldsymbol{\nabla} \cdot \boldsymbol{v} + \Psi. && (3.95) \end{aligned}$$

- Since we are interested how the total energy density changes in a given volume, $\rho\boldsymbol{v}^2/2 + \rho\epsilon$, we are supplementing the internal energy equation (3.95) with a conservation law of $\rho\boldsymbol{v}^2/2$. To this end, we consider

$$\frac{\partial}{\partial t}\left(\frac{\rho \boldsymbol{v}^2}{2}\right) = \frac{\boldsymbol{v}^2}{2}\frac{\partial \rho}{\partial t} + \rho\boldsymbol{v} \cdot \frac{\partial \boldsymbol{v}}{\partial t} \tag{3.96}$$

and substitute Eqs. (3.93) and (3.94) to get

$$\frac{\partial}{\partial t}\left(\frac{\rho \boldsymbol{v}^2}{2}\right) = -\frac{\boldsymbol{v}^2}{2}\boldsymbol{\nabla} \cdot (\rho\boldsymbol{v}) - \rho\boldsymbol{v} \cdot (\boldsymbol{v} \cdot \boldsymbol{\nabla})\boldsymbol{v} - \boldsymbol{v} \cdot \boldsymbol{\nabla} P + \boldsymbol{v} \cdot (\boldsymbol{\nabla} \cdot \bar{\boldsymbol{\Pi}}). \tag{3.97}$$

---

the bulk of the intra-cluster plasma $\beta \sim 100$, i.e., magnetic fields are on first sight dynamically not relevant while they can affect the thermodyanmics in subtle ways as we will learn later on.

Using the identity $(\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v} \equiv \boldsymbol{\nabla}\boldsymbol{v}^2/2$, we obtain an equation for the conservation of kinetic energy density

$$\frac{\partial}{\partial t}\left(\frac{\rho \boldsymbol{v}^2}{2}\right) + \boldsymbol{\nabla}\cdot\left(\frac{1}{2}\rho \boldsymbol{v}^2 \boldsymbol{v}\right) = -\boldsymbol{v}\cdot\boldsymbol{\nabla}P + \boldsymbol{v}\cdot(\boldsymbol{\nabla}\cdot\bar{\boldsymbol{\Pi}}). \qquad (3.98)$$

3.1.6.2 Jump Conditions

- Consider a propagating fluid discontinuity in the rest frame of the discontinuity. Fluid moves from upstream to downstream. We denote the *upstream conditions* by ρ_1, v_1, T_1 and *downstream conditions* by ρ_2, v_2, T_2.

- We would like to derive the relations (also known as “jump conditions”) between ρ_1, v_1, T_1 and ρ_2, v_2, T_2 for a steady-state, plane-parallel geometry of a fluid discontinuity such as a shock. First, we assume that the velocity is perpendicular to the surface of the discontinuity. While this may seem to be a substantial loss of generality, it captures the main effect of discontinuities as we will see by generalizing this simplification in the last part of this section. As we will also see, there are two types of discontinuities:

    1. shocks that are characterized by a mass flux through their interface, and
    2. tangential discontinuities which have *no* mass flux through their interface.

- Within the shock front or “transition layer” on a scale of λ_{eff}, viscous effects are important and cause the shock in the first place, i.e., dissipate kinetic energy and thus generate heat and entropy. However, outside the layer, viscous effects are small on scales $L \gg \lambda_{\mathrm{eff}}$. We will derive conservation equations of the form

  $$\frac{\mathrm{d}}{\mathrm{d}x}Q(\rho, v, P) = 0 \implies Q(\rho, v, P) = \mathrm{const.} \qquad (3.99)$$

  and although Q involves viscous terms, we can ignore these outside the shock zone and can derive jump conditions from equations without viscosity terms.

- We assume steady state ($\partial/\partial t = 0$) and plane-parallel geometry ($\partial/\partial y = \partial/\partial z = 0$, $\partial/\partial x = \mathrm{d}/\mathrm{d}x$). The conservation laws of

Eqs. (3.93), (3.94), (3.95), and (3.98) simplify to

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}(\rho v) &= 0, && (3.100)\\ v\frac{\mathrm{d}v}{\mathrm{d}x} &= -\frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}x} + \frac{1}{\rho}\frac{\mathrm{d}}{\mathrm{d}x}\left\{\left(\frac{4}{3}\eta+\xi\right)\frac{\mathrm{d}v}{\mathrm{d}x}\right\}, && (3.101)\\ \frac{\mathrm{d}}{\mathrm{d}x}(\rho\epsilon v) &= -P\frac{\mathrm{d}v}{\mathrm{d}x} + \left(\frac{4}{3}\eta+\xi\right)\left(\frac{\mathrm{d}v}{\mathrm{d}x}\right)^2, && (3.102)\\ \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}\rho v^2 v\right) &= -v\frac{\mathrm{d}P}{\mathrm{d}x} + v\frac{\mathrm{d}}{\mathrm{d}x}\left\{\left(\frac{4}{3}\eta+\xi\right)\frac{\mathrm{d}v}{\mathrm{d}x}\right\}. && (3.103) \end{aligned}$$

- The equation for mass conservation (Eq. 3.100) gives

$$\rho v = \text{const.} \implies \rho_1 v_1 = \rho_2 v_2 = j \implies [\rho v] = 0, \qquad (3.104)$$

where j is the current density and we changed the meaning of the brackets, [...], for the rest of this subsection: they now indicate differences between the up- and downstream quantities. Note, that the up- and downstream velocities, v_1 and v_2, are measured in the frame of the discontinuity!

- Using

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\rho v^2\right) = \rho v\frac{\mathrm{d}v}{\mathrm{d}x} + v\frac{\mathrm{d}}{\mathrm{d}x}(\rho v) \overset{(3.104)}{=} \rho v\frac{\mathrm{d}v}{\mathrm{d}x} \qquad (3.105)$$

allows Eq. (3.101) to be rewritten as

$$\begin{aligned} \rho v\frac{\mathrm{d}v}{\mathrm{d}x} + \frac{\mathrm{d}P}{\mathrm{d}x} - \frac{\mathrm{d}}{\mathrm{d}x}\left\{\left(\frac{4}{3}\eta+\xi\right)\frac{\mathrm{d}v}{\mathrm{d}x}\right\} & \\ = \frac{\mathrm{d}}{\mathrm{d}x}\left\{\rho v^2 + P - \left(\frac{4}{3}\eta+\xi\right)\frac{\mathrm{d}v}{\mathrm{d}x}\right\} &= 0 && (3.106)\\ \implies \left[\rho v^2 + P - \left(\frac{4}{3}\eta+\xi\right)\frac{\mathrm{d}v}{\mathrm{d}x}\right] &= 0 && (3.107) \end{aligned}$$

This demonstrates that within the transition zone (where η, ξ, and $\mathrm{d}v/\mathrm{d}x$ are non-zero) $\rho v^2 + P \neq \text{const}$. However, in the pre- and post-shock zones, η, ξ, and $\mathrm{d}v/\mathrm{d}x$ are negligible, implying

$$\left[\rho v^2 + P\right] = 0. \qquad (3.108)$$

- In principle, we could use Eq. (3.107) to follow the behavior in the transition zone, i.e., to understand how entropy is generated. But on scales $L < \lambda$ the fluid description breaks down and we have to reside to kinetic theory (or use plasma particle-in-cell codes to understand the nonlinear behavior of the heating process). From now on, we neglect viscosity effects in the bulk.

- Adding Eqs. (3.102) and (3.103) yields (for the region outside the transition zone)

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}x}\left\{ v\left(\frac{1}{2}\rho v^2 + \rho\epsilon\right) + Pv\right\} = \frac{\mathrm{d}}{\mathrm{d}x}\left\{\rho v\left(\frac{1}{2}v^2 + \epsilon + \frac{P}{\rho}\right)\right\} \\ &= \left(\frac{1}{2}v^2 + \epsilon + \frac{P}{\rho}\right)\frac{\mathrm{d}}{\mathrm{d}x}(\rho v) + \rho v \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}v^2 + \epsilon + \frac{P}{\rho}\right). \end{aligned} \tag{3.109}$$

  Since $\mathrm{d}(\rho v)/\mathrm{d}x = 0$ and $\rho v \neq 0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}v^2 + \epsilon + \frac{P}{\rho}\right) \implies \left[\frac{1}{2}v^2 + \epsilon + \frac{P}{\rho}\right] = 0. \tag{3.110}$$

- Summarizing, we have the Rankine-Hugoniot jump conditions for a plane-parallel shock in the shock rest frame:

$$[\rho v] = 0, \tag{3.111}$$

$$\left[\rho v^2 + P\right] = 0, \tag{3.112}$$

$$\left[\frac{1}{2}v^2 + \epsilon + \frac{P}{\rho}\right] = 0. \tag{3.113}$$

  Independent of the complicated physics within the transition layer, these conditions simply follow from the conservation laws. The first follows from mass conservation, the second from mass and momentum conservation, and the third from mass and total energy conservation.

- Using $\epsilon_i = P_i/\{\rho_i(\gamma_i - 1)\}$, we can rewrite the energy jump condition to get

$$\frac{1}{2}v_1^2 + \frac{\gamma_1}{\gamma_1 - 1}\frac{P_1}{\rho_1} = \frac{1}{2}v_2^2 + \frac{\gamma_2}{\gamma_2 - 1}\frac{P_2}{\rho_2} \tag{3.114}$$

  for a single-species gas that is described by a polytropic equation of state. In principle, $\gamma_1 \neq \gamma_2$, since a shock can e.g., dissociate molecules, or raise T so that previously inaccessible degrees of freedom become accessible.

3.1.6.3 Tangential Discontinuities

- $[\rho v] = 0$ allows for two types of solutions. The first type is clearly $\rho_1 v_1 = \rho_2 v_2 = 0$ and since ρ_1 and ρ_2 are non-zero, we have

$$v_1 = v_2 = 0, \tag{3.115}$$

$$P_1 = P_2 \implies [P] = 0 \tag{3.116}$$

  which follows from Eq. (3.112). The constancy of the normal component of the velocity across such an interface implies that there is no mass flux through a tangential discontinuity. If additionally the tangential velocity is also continuous, a special discontinuity is present which is called a contact discontinuity.

- At a tangential discontinuity, there can be an arbitrary jump of density, that however needs to be compensated by the same jump of T, but in the opposite direction!

3.1.6.4 Shock Mach Number

- The other type of solution requires $\rho_1 v_1 \neq 0$ so that we have a mass flux through this type of discontinuity that we call a "shock".

- We define a dimensionless number that characterizes the shock strength, the Mach number as the ratio of shock speed to upstream sound speed $c_1^2 = \gamma P_1/\rho_1$,

$$\mathcal{M}_1 \equiv \frac{v_1}{c_1} = \sqrt{\frac{\rho_1 v_1^2}{\gamma P_1}} = \sqrt{\frac{\bar{m} v_1^2}{\gamma k_\mathrm{B} T_1}}, \tag{3.117}$$

  which can be interpreted as a ratio of ram pressure $(\rho_1 v_1^2)$-to-thermal pressure in the pre-shock gas or equivalently a ratio of kinetic-to-thermal energy density.

- We can rewrite the Rankine-Hugoniot jump conditions in terms of $\mathcal{M}_1$ (and assuming $\gamma_1 = \gamma_2 = \gamma$, which is applicable for the ionized plasma of the ICM)

$$\frac{\rho_2}{\rho_1} = \frac{v_1}{v_2} = \frac{(\gamma+1)\mathcal{M}_1^2}{(\gamma-1)\mathcal{M}_1^2+2} \xrightarrow{\gamma=1} \mathcal{M}_1^2 \tag{3.118}$$

$$\frac{P_2}{P_1} = \frac{\rho_2 k_\mathrm{B} T_2}{\rho_1 k_\mathrm{B} T_1} = \frac{2\gamma\mathcal{M}_1^2 - (\gamma-1)}{\gamma+1} \xrightarrow{\gamma=1} \mathcal{M}_1^2 \tag{3.119}$$

$$\frac{T_2}{T_1} = \frac{\left[(\gamma-1)\mathcal{M}_1^2+2\right]\left[2\gamma\mathcal{M}_1^2-(\gamma-1)\right]}{(\gamma+1)^2\mathcal{M}_1^2} \xrightarrow{\gamma=1} 1 \tag{3.120}$$

  Note, the brackets in these equations retrieve their usual meaning.

- Those relations simplify for strong shocks ($\mathcal{M}_1 \gg 1$), yielding

$$\frac{\rho_2}{\rho_1} = \frac{v_1}{v_2} \approx \frac{\gamma+1}{\gamma-1} = 4, \tag{3.121}$$

$$P_2 \approx \frac{2\gamma}{\gamma+1}\mathcal{M}_1^2 P_1 = \frac{2}{\gamma+1}\rho_1 v_1^2 = \frac{3}{4}\rho_1 v_1^2, \tag{3.122}$$

$$k_\mathrm{B} T_2 \approx \frac{2\gamma(\gamma-1)}{(\gamma+1)^2} k_\mathrm{B} T_1 \mathcal{M}_1^2 = \frac{2(\gamma-1)}{(\gamma+1)^2}\bar{m} v_1^2 = \frac{3}{16}\bar{m} v_1^2, \tag{3.123}$$

  where we used a non-relativistic ideal gas ($\gamma = 5/3$) in the last equalities.

- In the shock rest frame, the post-shock kinetic and thermal specific energies are ($\gamma = 5/3$, $\mathcal{M} \gg 1$)

$$\frac{1}{2}v_2^2 \approx \frac{1}{32}v_1^2, \tag{3.124}$$

$$\frac{3}{2}\frac{k_\mathrm{B}T_2}{\bar{m}} \approx \frac{9}{32}v_1^2 = \frac{9}{16}\left(\frac{1}{2}v_1^2\right). \tag{3.125}$$

So roughly half of the pre-shock kinetic energy is converted to thermal energy (in the shock rest frame). The total specific energy ϵ_tot of the post-shock gas is

$$\epsilon_\mathrm{tot,2} = \frac{1}{2}v_2^2 + \frac{3}{2}\frac{k_\mathrm{B}T_2}{\bar{m}} \approx \frac{10}{16}\left(\frac{1}{2}v_1^2\right) = \frac{5}{8}\epsilon_\mathrm{kin,1} = \frac{5}{8}\epsilon_\mathrm{tot,1} \tag{3.126}$$

because in a strong shock, the upstream thermal energy is negligible in comparison to the kinetic energy. Hence, $\epsilon_\mathrm{tot,2}$ is lower than $\epsilon_\mathrm{tot,1}$ (in the shock rest frame) because of the $P\mathrm{d}V$ work done by pressure and viscosity on the post-shock gas in compressing its volume. Note that this $P\mathrm{d}V$ term is absent in the rest frame of the post-shock gas.

- The post-shock Mach number is

$$\mathcal{M}_2 \equiv \frac{v_2}{c_2} = \frac{v_1}{c_1}\frac{v_2}{v_1}\frac{c_1}{c_2} = \mathcal{M}_1\frac{v_2}{v_1}\left(\frac{T_1}{T_2}\right)^{1/2}. \tag{3.127}$$

This simplifies in the strong-shock limit, yielding

$$\mathcal{M}_2 \approx \mathcal{M}_1\frac{\gamma-1}{\gamma+1}\left[\frac{(\gamma+1)^2}{2\gamma(\gamma-1)\mathcal{M}_1^2}\right]^{1/2} = \left(\frac{\gamma-1}{2\gamma}\right)^{1/2} \approx 0.45. \tag{3.128}$$

A shock converts supersonic gas into denser, slower moving, higher pressure, subsonic gas.

3.1.6.5 Shock Adiabatic Curve

- The shock increases the specific entropy of the gas by an amount

$$\begin{aligned} s_2 - s_1 &= c_V \ln\left(\frac{P_2}{\rho_2^\gamma}\right) - c_V \ln\left(\frac{P_1}{\rho_1^\gamma}\right) \\ &= c_V \ln\left(\frac{P_2}{P_1}\right) - c_V\gamma \ln\left(\frac{\rho_2}{\rho_1}\right) = c_V \ln\left(\frac{K_2}{K_1}\right). \end{aligned} \tag{3.129}$$

Hence, the shock shifts the gas to a higher adiabatic curve that is uniquely labeled by $K = P\rho^{-\gamma}$: gas can move adiabatically along an adiabatic curve while changes in entropy move it from one adiabatic curve to another.

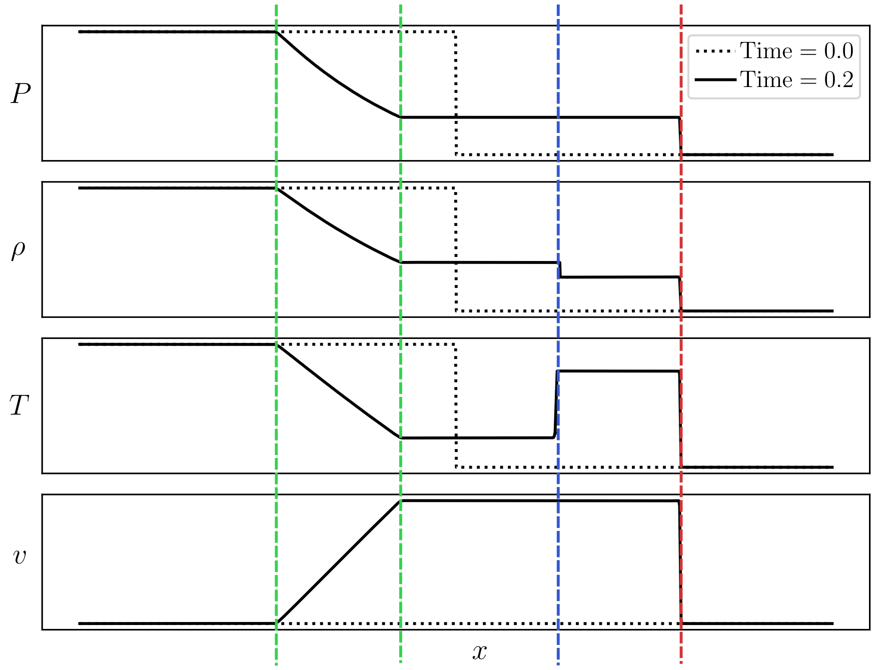


Figure 3.2: The figure shows the propagation of a shock wave. At $t = 0$, there is an initial discontinuity in the state variables P, ρ, T that separates two piecewise constant values (the left, L, and right, R, states), which meet in the middle of the x axis. The initial pressure jump pushes on the gas to the right and causes a shock to develop that propagates to the right. The solution is described by a Riemann problem and shown at $t = 0.2$. The solution is subdivided into different regions, separated by coloured, vertical lines. From right to left, we identify the unperturbed initial state and the shocked gas, separated by the shock (red dashed line). The shock compresses the gas and manifests as discontinuity in all state variables. The contact discontinuity (blue dashed line) separates the gas that was left and right of the initial discontinuity. The mass flux to the right causes a rarefaction wave moving to the left (i.e., the states in between the two green dashed lines) across which the gas is adiabatically expanded. The initial conditions for this example are $P_{\mathrm{L}} = 1$, $P_{\mathrm{R}} = 0.18$, $\rho_{\mathrm{L}} = 1$, $\rho_{\mathrm{L}} = 0.25$ and $v_{\mathrm{L}} = v_{\mathrm{R}} = 0$.

- With the definition of the current density $j = \rho_1 v_1 = \rho_2 v_2 = \text{const.}$, we obtain for Eq. (3.112)

$$[\rho v^2 + P] = \left[\frac{j^2}{\bar{m}} V + P\right] = 0 \implies \frac{j^2}{\bar{m}} V_1 + P_1 = \frac{j^2}{\bar{m}} V_2 + P_2. \tag{3.130}$$

- Hence, the slope of the shock adiabatic curve in the P-V diagram is given by

$$-\frac{j^2}{\bar{m}} = -\frac{P_2 - P_1}{V_1 - V_2}. \tag{3.131}$$

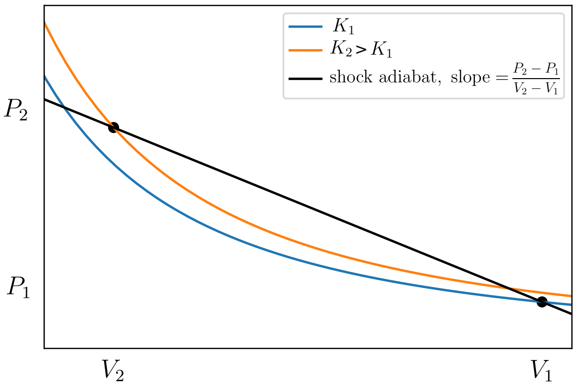


Figure 3.3: We show two adiabatic curves in the $P - V$ diagram labeled by their entropic values K_1 and K_2. A shock increases the entropy and directly connects the low-entropy state at (V_1, P_1) to the high-entropy state (V_2, P_2). The connecting line is called the "shock adiabat" and shows a characteristic slope given by the negative squared mass flux, $-j^2/\bar{m}$, see equation (3.131).

3.1.6.6 Oblique Shocks

- So far, our specific discussion about shocks has been constrained to plane-parallel shock geometries, i.e., we only considered a fluid velocity that was aligned with the shock normal. In general, the fluid can impact the shock at some oblique angle. We define a velocity component parallel to the shock normal, $v_{\|} \equiv \boldsymbol{v} \cdot \boldsymbol{n}$, as well as a perpendicular component, $v_{\perp}$.

- The momentum conservation equation (3.46) defines a momentum current through a unit surface area with normal vector $\boldsymbol{n}$ (neglecting viscosity outside the shock transition layer and splitting the flux into a normal and a perpendicular component),

$$\rho \boldsymbol{v}(\boldsymbol{v} \cdot \boldsymbol{n}) + P\boldsymbol{n}. \tag{3.132}$$

  The momentum current has to be continuous across the shock in order for the forces that are acting on both sides of the shock on the gas to be identical. In our case, $\boldsymbol{n}$ coincides with the shock normal and points along $\boldsymbol{e}_x$. Continuity of the x, y, and z components of the momentum current yields

$$[\rho v_x^2 + P] = 0, \tag{3.133}$$
$$[\rho v_x v_y] = 0, \tag{3.134}$$
$$[\rho v_x v_z] = 0. \tag{3.135}$$

- At a shock $j = \rho v_x \neq 0$ and $\rho \neq 0$ so that we get

$$[v_y] = 0 \quad \text{and} \quad [v_z] = 0, \tag{3.136}$$

i.e., the tangential velocities are continuous across the shock. Thus, only the parallel velocity component is modified at a shock according to $v_{\|,2} = v_{\|,1}\rho_1/\rho_2$ while the perpendicular component remains invariant, $v_{\perp,1} = v_{\perp,2} = v_{\perp}$. This implies a refraction of the (oblique) flow toward the shock surface.

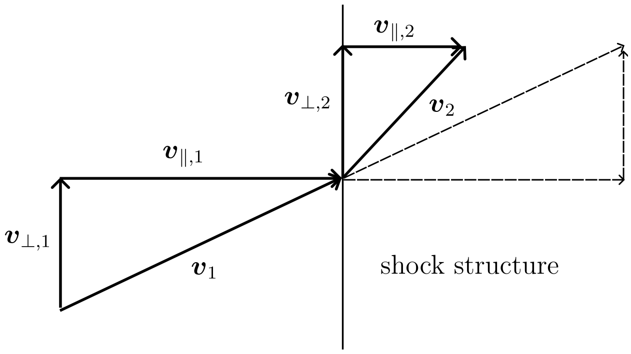


Figure 3.4: Geometry of an oblique shock. While the normal component $v_{\|}$ of the velocity gets decelerated at the shock (motions are converted to internal energy by the shock), the perpendicular velocity component $\boldsymbol{v}_{\perp}$ remains invariant across the shock transition.

- If the flow impinges with a constant angle at a shock, it is deflected by the same amount everywhere along the shock surface. Consider now a curved shock: this implies a changing angle between $\boldsymbol{v}$ and $\boldsymbol{n}$ along the shock surface and hence, the shock transition causes a different amount of "shock deflection" of the velocity field.

- As a result, there is shear injected at a shock because two infinitesimally separated points on the shock surface experience a different amount of deflection. This implies subsonic (solenoidal) turbulence in the post-shock regime: the vorticity is injected at the curvature radius of the shock and cascades down in scale to the Kolmogorov scale where it gets dissipated. Hence, in a curved shock, there is eventually more kinetic energy dissipated into heat in comparison to an oblique shock without curvature that experiences the same amount of ram pressure.

3.1.7 Entropy Generation by Accretion

3.1.7.1 Philosophical Detour

- The Uncertainty Principle is $\Delta p_x \Delta x = h$, and statistical mechanics counts the number of states with $h^{-3}\mathrm{d}^3x\mathrm{d}^3p$. Hence the phase space density of cluster gas is

$$f \sim \frac{h^3 \mathrm{d}^6 N}{\mathrm{d}^3 x \mathrm{d}^3 p} \sim n \left(\frac{h}{m_\mathrm{p} v}\right)^3$$
$$\sim 6 \times 10^{-35} \left(\frac{n}{10^{-3}\ \mathrm{cm}^{-3}}\right)\left(\frac{v}{10^3\ \mathrm{km\ s}^{-1}}\right)^{-3} \propto K^{-3/2}. \qquad (3.137)$$

If this was unity, we would deal with a degenerate gas. Instead, this is extremely small, making it the least degenerate (non-relativistic) gas in the Universe or equivalently, the highest entropy gas (of equilibrium systems)!

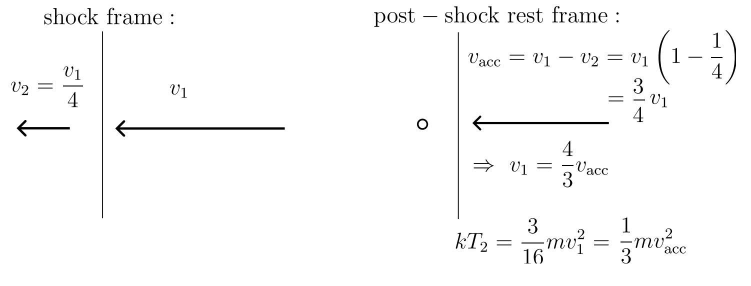
3.1.7.2 Smooth Accretion

- One way to approach the problem of gravitationally driven generation is through spherically symmetric models of smooth accretion in which gas passes through an accretion shock as it enters the cluster. If the incoming gas is cold, then the cluster accretion shock is the sole source of cluster entropy. If instead, the incoming gas has been heated before passing through the accretion shock, the Mach number is smaller and the cluster entropy level reflects both, the amount of pre-heating and entropy production at the accretion shock.

- We first consider the case of cold accretion (P and K of the incoming gas are negligible) which implies the strong-shock regime. Conveniently, we transform our Rankine-Hugoniot jump conditions to the rest frame of the post-shock gas, i.e., the cluster rest frame.

  shock frame: $$v_2 = \frac{v_1}{4} \qquad (3.138)$$

  post-shock rest frame: $$v_\mathrm{acc} = v_1 - v_2 = v_1\left(1 - \frac{1}{4}\right) = \frac{3}{4} v_1$$
  $$kT_2 = \frac{3}{16}\bar{m} v_1^2 = \frac{1}{3}\bar{m} v_\mathrm{acc}^2 \qquad (3.139)$$



- Note that the location of the accretion shock does *not* move outward with $v_\mathrm{acc}/4$ because the gravitational attraction of the cluster potential causes causes it to fall onto the cluster. Depending on that infall rate, the accretion shock can either be stationary or move slowly outward.

- Suppose that mass accretes in a series of concentric shells, each with a baryon fraction f_b, that initially comove with the Hubble flow as in the spherical collapse model of Section 2.3. In this simple model, a shell that initially encloses total mass M reaches zero velocity at the turnaround radius r_{ta} and falls back through an accretion shock at radius r_{acc} in the neighborhood of the virial radius $r_{ta}/2$. The system of governing equations are

$$\begin{aligned} \dot{M}_g &= 4\pi r_{acc}^2 \rho_1 v_{acc} = f_b \dot{M}, && (3.140)\\ v_{acc}^2 &= \frac{2GM}{r_{ta}}, \quad \text{(assuming } \Omega_\Lambda = 0\text{)}, && (3.141)\\ k_B T_2 &= \frac{1}{3}\bar{m} v_{acc}^2, \quad \bar{m} = \mu m_p, && (3.142)\\ \rho_2 &= 4\rho_1. && (3.143) \end{aligned}$$

Here, ρ_1 is the pre-shock density, ρ_2 and T_2 are the post-shock density and temperature, $r_{acc} = r_{ta}/2$ is the accretion radius. In the post-shock frame, the post-shock thermal energy equals the pre-shock ram pressure (+ initial thermal energy that we neglect here) and Eq. (3.142) implies

$$\epsilon_2 = \frac{3}{2}\frac{k_B T_2}{\bar{m}} = \frac{3}{2}\frac{1}{3} v_{acc}^2 = \frac{v_{acc}^2}{2} = \epsilon_{1,\,kin}. \qquad (3.144)$$

- Using Eqs. (3.142) and (3.143), we can compute the post-shock entropy that is produced by smooth accretion

$$K_{2,sm} \equiv \frac{k_B T_2}{\bar{m}\rho_2^{2/3}} = \frac{v_{acc}^2}{3(4\rho_1)^{2/3}}. \qquad (3.145)$$

Combining Eqs. (3.140) and (3.141) yields

$$\dot{M}_g = 4\pi r_{acc}^2 \rho_1 \left(\frac{GM}{r_{acc}}\right)^{1/2} \implies \rho_1 = \frac{\dot{M} f_b}{4\pi r_{acc}^{3/2}\sqrt{GM}}. \qquad (3.146)$$

- Inserting this into Eq. (3.145) yields

$$\begin{aligned} K_{2,sm} &= \frac{v_{acc}^2}{3(4\rho_1)^{2/3}} = \frac{1}{3}\left[\frac{\pi (GM)^2}{f_b \dot{M}}\right]^{2/3} \\ &= \frac{1}{3}\left(\frac{\pi G^2}{f_b}\right)^{2/3} \left(\frac{d\ln M}{d\ln t}\right)^{-2/3} (Mt)^{2/3}. && (3.147) \end{aligned}$$

Because the entropy generated at the shock front increases monotonically with time, such an ideal, smoothly accreting cluster never convects but rather accretes shells of baryons as if they were onion skins. It is useful to cast Eq. (3.147) into dimensionless form using a characteristic cluster entropy K_{200},

$$K_{200} \equiv \frac{k_\mathrm{B}T_{200}}{\bar{m}(200 f_\mathrm{b}\rho_\mathrm{cr})^{2/3}} = \frac{1}{2}\left[\frac{2\pi}{15}\frac{G^2 M_{200}}{f_\mathrm{b}H(z)}\right]^{2/3} . \quad (3.148)$$

Note that we have adopted here the critical density of the Universe, ρ_cr, and the characteristic temperature of a singular isothermal sphere, $k_\mathrm{B}T_{200}$,

$$\rho_\mathrm{cr} \equiv \frac{3H^2(z)}{8\pi G}, \quad (3.149)$$

$$k_\mathrm{B}T_{200} = \frac{GM_{200}\bar{m}}{2r_{200}} = \frac{\bar{m}}{2}\left[10GM_{200}H(z)\right]^{2/3} . \quad (3.150)$$

- We effectively define a radial coordinate corresponding to the amount of gas accreted at t divided by that at the present time, t_0:

$$\eta \equiv \frac{M_\mathrm{g}(t)}{f_\mathrm{b}M_{200}(t_0)} \quad (3.151)$$

and cast Eq. (3.147) into dimensionless form

$$\frac{K_{2,\mathrm{sm}}}{K_{200}} = \frac{2}{3}\left(\frac{15}{2}\frac{H_0}{M_{200}}\right)^{2/3}\left(\frac{\mathrm{d}\ln M}{\mathrm{d}\ln t}\right)^{-2/3}(Mt)^{2/3}$$
$$= \frac{2}{3}\left(\frac{15}{2}H_0 t_0\right)^{2/3}\left(\frac{\mathrm{d}\ln \eta}{\mathrm{d}\ln t}\right)^{-2/3}\left(\frac{\eta t(\eta)}{t_0}\right)^{2/3} . \quad (3.152)$$

Thus, the entropy profile due to smooth accretion of cold gas depends entirely on the mass accretion history $M(t)$, and the entropy profiles of objects with similar $M(t)$ should be self-similar with respect to K_{200}.

- Extended Press-Schechter theory or numerical simulations show that clusters in the mass range 10^{14}-$10^{15}\,\mathrm{M}_\odot$ grow roughly as $M(t) \propto t$ to $M(t) \propto t^2$ in the concordance model. Inserting these growth times $t \propto M^{0.5...1}$ into Eq. (3.152) yields

$$K_{2,\mathrm{sm}} \propto M_\mathrm{g}^{1...4/3} . \quad (3.153)$$

- Throughout a cluster, M_g encompassed by a given radius is approximately $M_\mathrm{g} \propto r$ (which is exact for the singular isothermal sphere, see Eq. 2.102). We obtain the following radial entropy profile

$$K_{2,\mathrm{sm}} \propto r^{1...4/3} \quad (3.154)$$

which compares well with numerical simulations $K_{2,\mathrm{sm}} \propto r^{1.1}$.

3.1.7.3 Hierarchical Merging

- In real clusters the accreting gas is lumpy and not smooth which transforms the nature of entropy generation. Incoming gas that is bound to accreting sublumps of matter enters the cluster with a wide range of densities. There is no well-defined accretion shock but rather a complex network of shocks as different lumps of infalling gas mix with the intracluster medium of the main cluster.

- Numerical codes employing different numerical techniques (Eulerian grid codes or Lagrangian "particle" codes) all agree on the baseline profile in non-radiative gas simulations for $r > 0.1r_{200}$,

  $$K_{\rm sim} = 1.32K_{200}\left(\frac{r}{r_{200}}\right)^{1.1}. \tag{3.155}$$

  For $r < 0.1r_{200}$, there is more dispersion among the simulated clusters and the answer depends somewhat on the numerical technique, with grid codes showing an elevated entropy core due to efficient "entropy mixing".

- Despite the complexity of shock structure in hierarchical accretion, the simulated entropy profiles resemble that from smooth accretion models, but with an important distinction: the normalization of the smooth models is higher. The likely reason is that smooth accretion maximizes the entropy production as smoothing does not change the accretion velocity but reduces the mean density of accreting gas lumps. Since the post-shock entropy scales as $K_2 \propto \upsilon_{\rm acc}^2\rho_1^{-2/3}$, a smaller (smoothed) density implies larger entropy everywhere.

- This may be of physical relevance: consider the case of preheating gas before it falls into clusters. Heating causes the gas to adiabatically expand and smoothes the density field of the preshock gas. This could then explain elevated entropy profiles of low-temperature clusters relative to the baseline profile.

3.1.8 Cluster Scaling Relations

- In order to use clusters as cosmological probes, we need to relate the different observables to a functional that is sensitive to cosmology. Traditionally this is obtained by using the mass function.

- The main assumption underlying this approach is the choice of an average density of the matter so that this implicitly defines a cluster "radius" by

  $$M_\Delta = \frac{4}{3}\pi r_\Delta^3 \Delta\rho_{\rm cr}(z), \quad \Delta = 100\dots500, \tag{3.156}$$

which also relates the temperature to this definition, $T \sim T_\Delta$.

- Cautionary remarks: when considering X-ray emission, we encounter ρ_X, T_X which is not necessarily identical to $\bar{\rho} = \Delta\rho_\mathrm{cr}$ and T_Δ as it is degenerate with observational biases. Not accounting for these would break self-similarity as e.g., the presence of a clumped multiphase medium may bias T_X towards the dense, colder phase with a higher X-ray emissivity. On the other hand, we encounter similar problems when defining a three-dimensional "radius" from a projected, non-isotropic density distribution (or X-ray emissivity $j_\mathrm{X} \propto \rho^2$). Careful mock observations of simulated clusters are necessary in either case.

3.1.8.1 Cosmologist's Ideal Cluster

- Recall the definition of the *critical* density at scale factor a and today are defined as

$$\rho_\mathrm{cr}(a) \equiv \frac{3H^2(a)}{8\pi G}\,, \quad \rho_\mathrm{cr0} \equiv \frac{3H_0^2}{8\pi G}\,, \quad \text{and} \tag{3.157}$$

$$H^2(a) \equiv \left(\frac{\dot{a}}{a}\right)^2 \equiv H_0^2 E^2(a) = H_0^2\left[\Omega_\mathrm{r0}a^{-4} + \Omega_\mathrm{m0}a^{-3} + \Omega_{\Lambda 0} + \Omega_\mathrm{K}a^{-2}\right] \tag{3.158}$$

is the Hubble function that derives from Friedmann's equation (2.5) and describes the expansion rate of the universe and the cosmological parameters are defined in Eq. (2.8). Note that scale factor and redshift are related via $a = 1/(1+z)$.

- We assume hydrostatic equilibrium and obtain the scaling

$$\frac{k_\mathrm{B}T}{m} \sim v^2 \sim \frac{GM_\Delta}{r_\Delta} \propto M_\Delta^{2/3}\rho_\mathrm{cr}^{1/3} \tag{3.159}$$

which immediately yields the temperature-mass scaling

$$T_\Delta \propto M_\Delta^{2/3}\rho_\mathrm{cr}^{1/3}(z) \propto [M_\Delta E(z)]^{2/3}. \tag{3.160}$$

- We assume that clusters are self-similar objects that only scale with $M_\Delta = M_\mathrm{tot}$. Consequently, the gas fraction, $f_\mathrm{gas}(< r_\Delta) = M_\mathrm{gas}/M_\mathrm{tot}$, and stellar mass fraction, $f_\star(< r_\Delta) = M_\star/M_\mathrm{tot}$, do not scale with cluster mass. Here, $M_\mathrm{tot} = M_\mathrm{tot}(< r_\Delta) = M_\mathrm{dm} + M_\mathrm{gas} + M_\star$ is the gravitational mass. We get the gas and stellar mass scaling,

$$M_\mathrm{gas} = \int_0^{r_\Delta} \rho_\mathrm{gas}\mathrm{d}V \approx M_\Delta f_\mathrm{gas} \propto M_\Delta, \tag{3.161}$$

$$M_\star \approx M_\Delta f_\star \propto M_\Delta \implies N_\mathrm{gals} \propto M_\Delta. \tag{3.162}$$

Especially $N_\mathrm{gals} \propto M_\Delta$ assumes a fair sampling of the luminosity function which is not anymore the case on group scales with $\mathcal{O}(10)$ galaxies.

- To obtain the Sunyaev-Zel'dovich scaling relation, we integrate the Compton-y parameter over the cluster's cross section, A,

$$\begin{aligned} Y_{\rm sph} &= \int_A y \mathrm{d}A = \frac{\sigma_T}{m_{\rm e}c^2}\int_V n_{\rm e} k_{\rm B} T_{\rm e} \mathrm{d}V, & (3.163)\\ Y_{\rm sph} &\propto M_{\rm gas} T_\Delta \propto M_\Delta^{5/3} E(z)^{2/3}. & (3.164)\end{aligned}$$

- Finally, the X-ray scaling relation is obtained by assuming that free-free emission (a two-body process) is the dominating radiative process. In this case, the emissivity per unit volume is $j_{\rm X} \propto n_{\rm e} n_{\rm ion} T^{1/2}$ and we obtain the following scaling of X-ray luminosity with cluster mass,

$$\begin{aligned} L_{\rm X} &\propto \int_V n_{\rm e} n_{\rm ion}\sqrt{k_{\rm B} T_{\rm e}} \mathrm{d}V \propto M_{\rm gas}\rho_{\rm cr} T_\Delta^{1/2}, & (3.165)\\ L_{\rm X} &\propto M_\Delta^{4/3} E(z)^{1/3} \propto T_\Delta^2 E(z)^{-1}. & (3.166)\end{aligned}$$

3.1.8.2 Real Clusters

- Observational scaling relations do not follow the self-similar prediction above. One finds

$$\begin{aligned} L_{\rm X} &\propto T_{\rm X}^3, & (3.167)\\ \frac{\mathrm{d}}{\mathrm{d}M}\left(\frac{M_{\rm gas}}{M}\right) &> 0, & (3.168)\\ \frac{\mathrm{d}}{\mathrm{d}M}\left(\frac{M_\star}{M}\right) &< 0 & (3.169)\end{aligned}$$

  where M is some observational proxy for M_Δ. $Y(M)$ and $T_{\rm X}(M)$ are roughly in agreement with the self-similar prediction. It appears that gas physics modifies these simple scale-invariant models and imposes new scales to the otherwise scale-free gravity!

- Consider a simple cored model for the gas distribution of the ICM:

$$\rho(r) = \begin{cases} \text{const.} & \text{for } r < r_{\rm c},\\ \rho_\Delta\left(\dfrac{r}{r_\Delta}\right)^{-2} & \text{for } r > r_{\rm c}. \end{cases} \qquad (3.170)$$

  We define the cluster concentration parameter $c = r_\Delta/r_{\rm c}$, implying $\rho(r_{\rm c}) \equiv \rho_{\rm c} = c^2\rho_\Delta$.

- If bremsstrahlung (free-free emission) is the dominating emission process, we obtain an X-ray luminosity in this model,

$$\begin{aligned} L_{\rm X} &\propto \int_0^R \rho^2 T^{1/2} r^2 \mathrm{d}r, & (3.171)\\ \frac{\mathrm{d}L_{\rm X}}{\mathrm{d}\ln r} &\propto r^3\rho^2 T^{1/2} \sim \begin{cases} r^3 & \text{for } r < r_{\rm c},\\ r^{-1}T^{1/2} & \text{for } r > r_{\rm c}. \end{cases} & (3.172)\end{aligned}$$

Thus, the contribution to the X-ray luminosity per logarithmic bin in radius increases steeply toward r_c (because of the larger available volume) and then drops beyond r_c (after realizing that $T \sim r^{-1/2}$ in the peripheral cluster parts). The radii around r_c dominate L_X and thus, we expect

$$L_X \propto \rho_c^2 T^{1/2} r_c^3 \approx \rho_\Delta^2 c r_\Delta^3 T^{1/2}. \qquad (3.173)$$

Using Eqs. (3.160) and (3.161), we obtain

$$L_X \propto c M_\Delta^{4/3} E(z)^{1/3} \propto c T_\Delta^2 E(z)^{-1} \qquad (3.174)$$

- If c is independent of mass, we recover Eq. (3.166). However, gas physics modifies c so that it assumes a mass dependence. There have been three classes of models suggested to explain the deviation from scale invariance:

  1. "Pre-heating" the gas by supernova winds or some other feedback mechanism *before* falling into clusters imprints an "entropy floor" in the gas distribution – a minimum entropy level $K_{\rm min}$ below which it cannot fall. The clusters' central entropy is $K_0 \propto T\rho_0^{-2/3} \propto Tc^{-4/3}$. If all clusters have $K_0 = K_{\rm min} = \text{const.}$, then

     $$c \propto T^{3/4} K_{\rm min}^{-3/4} \implies L_X \propto T^{2.75}. \qquad (3.175)$$

     Thus, an entropy floor leads to larger core (relative to r_Δ), $r_c = r_\Delta/c \propto (K_{\rm min}/T)^{3/4} r_\Delta$, which is larger for smaller clusters (lower T) and thus to a steeper L_X-T relation close to the observations.

  2. An alternative possibility is that the gas gets heated after falling into the cluster, potentially through feedback by active galactic nuclei (AGN). This is however energetically much more expensive: to reach the same entropy as in the pre-heating case, one needs more energy if it is injected at the center by a factor

     $$\frac{k_B T_{\rm center}}{k_B T_{\rm pre-heat}} \sim \frac{K_{\rm center}}{K_{\rm pre-heat}} \left[\frac{n_{\rm center}}{n_{\rm pre-heat}}\right]^{2/3} \sim 10^2. \qquad (3.176)$$

     Here, we adopted typical values for $n_{\rm center} \sim 2 \times 10^{-3}\,\text{cm}^{-3}$ and $n_{\rm pre-heat} \sim 10\bar{n} \sim 2 \times 10^{-6}\,\text{cm}^{-3}$. However, AGNs can provide this energy (see Section 3.2.2) if the energy can be effectively coupled into the ICM.

  3. Cooling out the low-entropy gas at the cluster center and fueling central star formation selectively removes the low-entropy gas. The gas at larger radii (and on higher adiabatic curves) flows in adiabatically and replaces the condensed

gas which leads to the formation of an entropy floor. This process is observed to happen, but the star formation rate is only 10% of what would be needed to explain the steeper L_{X}-T slope.

3.2 Radiative Physics

- Observed cluster scaling relations do not obey self-similar predictions. Hence we have to take more realistic physics such as radiative cooling and star formation into consideration. The nonlinearity of the problem requires numerical simulations that represent a formidable computational challenge. This requires numerical codes that simulate three-dimensional hydrodynamics in simulations that span an enormous range in scales and track a plethora of physical processes. Typically, we simulate a periodic box of side length L that contains a representative volume of the universe and is large enough to host enough objects of interest to provide a sufficiently large statistical sample.

  - $L \gtrsim 300\,\mathrm{Mpc} \sim 10^{27}\,\mathrm{cm}$. This large volume is necessary in order to get a few sites of constructive interference of long wavelength modes which evolve into a few massive ($M \sim 10^{15}\,\mathrm{M}_\odot$) galaxy clusters.
  - $l \sim 30\,\mathrm{kpc} \sim 10^{23}\,\mathrm{cm}$. The simulation needs to resolve the diameter l of the stellar content of galaxies by at least 10 resolution elements. Such a Eulerian mesh would then have $[L/(0.1l)]^3 \sim 10^{15}$ individual cells—too many elements even for current state-of-the-art simulations. A solution to this problem consists by either introducing adaptive grid-refinement capabilities in Eulerian codes (which increase the numerical resolution where needed, i.e., inside collapsing objects) or Lagrangian simulation frameworks that discretize the simulated mass rather than simulation space.
  - $l_\star \sim 3\,\mathrm{pc} \sim 10^{19}\,\mathrm{cm}$. The star forming regions have typical sizes of 3 pc. The resulting dynamical range of the simulation volume, $(L/l_\star)^3 \sim 10^{24}$, is prohibitively large to be reliably included in first-principle, ab initio simulations. Instead, this requires a subgrid prescription of star formation physics to include the necessary dynamical back-reaction effects on the resolved larger scales.

- First, we will turn to the physics of cooling and condensation of baryons into stars (that will happen inevitably once the gas gets sufficiently dense). Numerical simulations show that this transforms a fraction of $f_\star/f_{\mathrm{gas}} \sim 25\% \ldots 50\%$ into stars without accounting for energy feedback. Since this is 5 to 10 times as much as observed in a cluster, we will then look at various "feedback processes" that were suggested to solve the "overcooling problem" of galaxy formation or equivalently the "cooling flow problem" in galaxy cluster evolution.

3.2.1 Radiative Cooling

- At high temperatures ($k_\mathrm{B}T \gtrsim 2$ keV) the light- and intermediate-mass elements of the ICM are fully ionized so that the only cooling process for them is free-free emission (thermal bremsstrahlung). Below $k_\mathrm{B}T \sim 2$ keV recombination-line cooling of heavy elements (Fe, ...) starts to dominate the cooling process (and the associated X-ray emission, assuming typical heavy element abundances relative to hydrogen, which are ~ 0.3 times those found in the Sun).

- The physics of bremsstrahlung emission is simple: electrons scatter off ions and are deflected in the Coulomb field of the ions. They radiate because of their acceleration and thus lose energy, i.e., they "cool".

- The spectral X-ray emissivity j_ν is defined as as the amount of energy emitted in photons of frequency ν per unit frequency interval dν, per unit time and per unit plasma volume, $j_\nu = \mathrm{d}^3E/(\mathrm{d}\nu\mathrm{d}t\mathrm{d}V)$. It must scale with the product of electron and ion number density (because it is a two-body interaction), with the time available for the scattering process, $t \sim l/\Delta v \sim l/\sqrt{k_\mathrm{B}T/\bar{m}}$, where Δv is the relative velocity of electron and ion, and with the Boltzmann factor for the distribution of energy at a given temperature. Hence we get

$$j_\nu = \frac{\mathrm{d}^3E}{\mathrm{d}\nu\mathrm{d}t\mathrm{d}V} = \tilde{C}\frac{n^2}{\sqrt{k_\mathrm{B}T}}\mathrm{e}^{-h\nu/k_\mathrm{B}T}, \quad \tilde{C} = \mathrm{const.} \tag{3.177}$$

- The volume emissivity is the integral of j_ν over frequency,

$$j \equiv \frac{\mathrm{d}^2E}{\mathrm{d}t\mathrm{d}V} = \int_0^\infty \frac{\mathrm{d}^3E}{\mathrm{d}\nu\mathrm{d}t\mathrm{d}V}\mathrm{d}\nu = \tilde{C}\frac{n^2}{\sqrt{k_\mathrm{B}T}}\frac{k_\mathrm{B}T}{h}\int_0^\infty \mathrm{e}^{-x}\mathrm{d}x \tag{3.178}$$

$$= Cn^2\sqrt{k_\mathrm{B}T} = 2.5\times10^{-23}\left(\frac{n_\mathrm{H}}{1\,\mathrm{cm}^{-3}}\right)^2\left(\frac{T}{10^8\,\mathrm{K}}\right)^{1/2}\frac{\mathrm{erg}}{\mathrm{cm}^3\,\mathrm{s}}, \tag{3.179}$$

for 0.3 times solar metallicity $Z_\odot$.

- Comparing the thermal energy content to the total (frequency-integrated) X-ray emissivity defines the cooling time

$$t_\mathrm{cool} = \frac{\varepsilon_\mathrm{th}}{|\dot{\varepsilon}_\mathrm{brems}|} = \frac{3nk_\mathrm{B}T}{2j} \tag{3.180}$$

$$\approx 2\left(\frac{k_\mathrm{B}T}{\mathrm{keV}}\right)^{1/2}\left(\frac{n_\mathrm{e}}{10^{-2}\,\mathrm{cm}^{-3}}\right)^{-1}\,\mathrm{Gyr}, \tag{3.181}$$

where $n = \rho/(\mu m_\mathrm{p})$, n_e is the electron number density, and the mean molecular weight of a fully ionized primordial gas with primordial hydrogen mass fraction $X = 0.76$ is $\mu = 0.588$, (see Appendix A.1).

- Hence in the centers of galaxy clusters, $t_{\rm cool}$ is smaller than the Hubble time: if gas in pressure equilibrium cools, it becomes denser and cools even faster. This is a run-away process that should lead to a large amount of cold gas and star formation—in conflict with observations. This is the famous "cooling flow problem".

- We can gain further insight if we rewrite $t_{\rm cool}$ in terms of the cluster entropy $K_{\rm e} \equiv k_{\rm B}Tn_{\rm e}^{-2/3}$. We define $t_0 = 2$ Gyr, $k_{\rm B}T_0 =$ keV, and $n_0 = 10^{-2}\,{\rm cm}^{-3}$, to obtain

$$t_{\rm cool} = t_0\left(\frac{k_{\rm B}T}{k_{\rm B}T_0}\right)^{1/2}\frac{n_0}{n_{\rm e}} = t_0\left(\frac{K_{\rm e}}{K_0}\right)^{3/2}\frac{k_{\rm B}T_0}{k_{\rm B}T}, \qquad (3.182)$$

  where $K_0 = 21.5\,{\rm keV\,cm}^2$ is a typical value for the central entropy in cool core clusters. Because $t_0 \ll t_{\rm Hubble} \approx 14\,{\rm Gyr}$ the cooling ICM needs additional (non-gravitational) energy injection that stabilizes it against the cooling catastrophe.

- This demonstrates that clusters with similar temperatures (or potential depths) have longer cooling times if the central entropy is larger. We can derive a critical entropy

$$K_{\rm c}(T) \approx 80\left(\frac{t_{\rm cool}}{14\,{\rm Gyr}}\right)^{2/3}\left(\frac{k_{\rm B}T}{\rm keV}\right)^{2/3}\,{\rm keV\ cm}^2, \qquad (3.183)$$

  that is large enough to avoid the cooling catastrophe in galaxy groups with $k_{\rm B}T \sim$ keV.

3.2.2 Cooling versus Heating

- We have seen that the cooling time in the core region of cool core clusters is smaller than the Hubble time which would imply a cooling catastrophe if not countered by energy feedback. To see how much feedback is needed, we first compute the cooling rate and redefine the X-ray emissivity as an energy cooling rate $\Lambda(T)$ according to

$$j = Cn_{\rm H}^2\sqrt{k_{\rm B}T} = \Lambda(T)n_{\rm H}^2, \quad \text{where} \qquad (3.184)$$

$$\begin{aligned}\Lambda(T) &= \Lambda_0\left(\frac{k_{\rm B}T}{k_{\rm B}T_0}\right)^{1/2}\\ &= 2.5\times10^{-23}\left(\frac{T}{10^8\,{\rm K}}\right)^{1/2}\frac{\rm erg\,cm^3}{\rm s}, \quad (Z = 0.3Z_\odot). \qquad (3.185)\end{aligned}$$

- We adopt a typical gas density profile as found in X-ray observations, the so-called beta profile which is simply a King profile with the outer slope parametrized by $\beta \approx 2/3\ldots1$:

$$n(r) = n_0\left[1+\left(\frac{r}{r_{\rm c}}\right)^2\right]^{-3\beta/2}. \qquad (3.186)$$

- We consider the X-ray luminosity as a proxy for the cooling luminosity. It is given by

$$\begin{aligned} L_\mathrm{X} &= \int_0^\infty j \mathrm{d}V = \Lambda_0 \sqrt{\frac{k_\mathrm{B}T}{k_\mathrm{B}T_0}}\, 4\pi \int_0^\infty n^2(r) r^2 \mathrm{d}r && (3.187)\\ &= \frac{4\pi}{3} r_\mathrm{c}^3 n_0^2 \Lambda_0 \sqrt{\frac{k_\mathrm{B}T}{k_\mathrm{B}T_0}} \times 3 \int_0^\infty \frac{x^2 \mathrm{d}x}{(1+x^2)^{3\beta}} && (3.188)\\ &= \frac{4\pi}{3} r_\mathrm{c}^3 n_0^2 \Lambda_0 \sqrt{\frac{k_\mathrm{B}T}{k_\mathrm{B}T_0}} \times \begin{cases} \frac{3\pi}{16} & \text{for } \beta = 1 \\ \frac{3\pi}{4} & \text{for } \beta = 2/3 \end{cases} && (3.189)\\ &\sim 10^{44} \left(\frac{r_\mathrm{c}}{100\,\mathrm{kpc}}\right)^3 \left(\frac{n_0}{10^{-2}\,\mathrm{cm}^{-3}}\right)^2 \left(\frac{k_\mathrm{B}T}{3\,\mathrm{keV}}\right)^{1/2} \mathrm{erg\ s}^{-1}, && (3.190) \end{aligned}$$

  where we adopted $\beta = 1$ in the last step. Note that to order of magnitude, it suffices to assume a homogeneous sphere with radius r_c and a density that is equal to that of the core region to calculate L_X. This corresponds to our finding $\mathrm{d}L_\mathrm{X}/\mathrm{d}\ln r \propto r^3 n^2 \sqrt{T}$ in Eq. (3.171) that radii around r_c dominate L_X.

- Hence, a successful feedback process has to heat the ICM at an average rate of 10^{44} erg s^{-1} to balance the cooling losses.

3.2.3 Feedback by Supernovae

The first obvious candidate for energy feedback are supernovae (SNe), i.e., exploding stars at the end of their lifetimes. There are two types of SNe: 1. core-collapse SNe (of type Ib,c or II) and 2. thermonuclear SNe (SNe of type Ia).

- **Core-collapse SNe.**
  - At the end of the lifetime of a massive star ($M \gtrsim 10\,\mathrm{M}_\odot$) it has used up its "fuel" (H, He, ...), i.e., the energy generated by nuclear burning and it cannot anymore balance the gravitational attraction. As a result, the core collapses and forms a black hole or a neutron star (pulsar). The envelope also collapses to nuclear densities which triggers an outward traveling shock that unbinds the envelope and ejects it. This enriches the surrounding medium with intermediate-mass elements, so-called "α elements" which can be built from α-particle nuclei (^{4}He) such as ^{16}O, ^{20}Ne, ^{24}Mg, ^{28}Si, ^{32}S, ^{36}Ar, ^{40}Ca, ^{48}Ti.
  - To estimate the effect of SNe heating on the ICM, we make three simplifications. We assume that 1. metals are fully mixed within the ICM, 2. neglect radiative losses, and 3. assume solar abundances. Since the metallicity Z of clusters is

typically $0.3Z_\odot$ and radiative losses cause a large fraction of this SNe energy to be radiated away, these numbers represent the absolute upper limit that SNe can contribute to the heating which is plausibly not reachable in the ICM.

- The mass fraction of α elements for a gas of solar abundance is
$$\frac{M_\alpha}{M_{\rm gas}} \approx 0.02. \tag{3.191}$$
Hence the supernova energy per α element that is created by the SN is given by
$$\frac{E_{\rm SN} m_{\rm p}}{M_\alpha} \sim \frac{10^{51}\,{\rm erg}\, m_{\rm p}}{10\,{\rm M}_\odot} \sim \frac{10^{51-24-34}}{2}\frac{\rm erg}{\rm nucleon} \sim 50\,\frac{\rm keV}{\rm nucleon}. \tag{3.192}$$
Mixing this energy into to the ICM (and neglecting radiative losses), we get
$$\frac{E_{\rm SN} m_{\rm p}}{M_{\rm gas}} \sim 1\,\frac{\rm keV}{\rm nucleon}. \tag{3.193}$$

- **Thermonuclear SNe.** The progenitor system of a type Ia supernova consists of a binary with at least one massive ($\approx$ 1 M$_\odot$) carbon-oxygen white dwarf:
  - The *single-degenerate scenario* assumes that the companion of the white dwarf is an evolved star. When the companion star becomes a red giant, it grows over its Roche volume and transfers mass to the white dwarf. White dwarfs are stabilized by the Fermi pressure of a degenerate electrons gas. This can only stabilize masses up to $1.4\,{\rm M}_\odot$ against gravity. When the companion star feeds the white dwarf beyond this limit, a thermonuclear runaway burning is eventually triggered, which explodes the white dwarf. This scenario appears to be ruled out for explaining the majority of type Ia supernovae.
  - Alternatively, the *double-degenerate scenario* assumes the existence of a binary consisting of two carbon-oxygen white dwarfs. At the end of their evolution, they merge and cause a thermonuclear runaway burning of carbon and oxygen in the more massive progenitor. The resulting type Ia supernovae explosion generates $\approx$ 1 M$_\odot$ ^{56}Ni, which decays radioactively into ^{56}Co and eventually to ^{56}Fe. This decay is responsible for the extraordinary brightness of type Ia supernovae ($\sim 10^{11}$ times more luminous in comparison to a star on the main sequence).

- Using the same assumptions as above, we obtain

$$\frac{M_{\mathrm{Fe}}}{M_{\mathrm{gas}}} \approx 0.001 \quad \text{(solar abundance)}, \tag{3.194}$$

$$\frac{E_{\mathrm{SN}} m_{\mathrm{p}}}{M_{\mathrm{Fe}}} \sim \frac{10^{51}\,\mathrm{erg}\, m_{\mathrm{p}}}{1\,\mathrm{M}_{\odot}} \sim 500\,\frac{\mathrm{keV}}{\mathrm{nucleon}}, \tag{3.195}$$

$$\frac{E_{\mathrm{SN}} m_{\mathrm{p}}}{M_{\mathrm{gas}}} \sim 0.5\,\frac{\mathrm{keV}}{\mathrm{nucleon}}. \tag{3.196}$$

- **Problems.** As we will now show, there are two problems with this hypothetical picture in which SNe provide the feedback energy: 1. the energetics is not sufficient and 2. the radiative losses are too strong to solve the “cooling flow problem”.

  1. For comparison we estimate the gravitational energy of a Milky Way-type galaxy and a massive galaxy cluster

  $$E_{\mathrm{gal}} \approx \frac{m_{\mathrm{p}}}{2} \upsilon_{\mathrm{gal}}^2 \approx 0.25 \left(\frac{\upsilon_{\mathrm{gal}}}{220\,\mathrm{km\,s^{-1}}} \right)^2 \mathrm{keV}, \tag{3.197}$$

  $$E_{\mathrm{cluster}} \approx \frac{m_{\mathrm{p}}}{2} \sigma_{\mathrm{cluster}}^2 \approx 8 \left(\frac{\sigma_{\mathrm{cluster}}}{1200\,\mathrm{km\,s^{-1}}} \right)^2 \mathrm{keV}. \tag{3.198}$$

  While SNe feedback can energetically modulate the star formation within galaxies, it is (by about an order of magnitude even for the unrealistically optimistic case) too weak in clusters to have any thermodynamic impact.

  2. In order to avoid radiative losses, SNe heating has to raise the entropy of the gas it heats to at least $\sim 100\,\mathrm{keV\,cm^2}$, see Eq. (3.183). An evenly distributed thermal energy input of order 1 keV would thus have to go into gas significantly less dense than $10^{-3}\,\mathrm{cm^{-3}}$ to avoid such losses, see Eq. (3.181). But gas near the centers of present-day cluster (not to mention the densities of the interstellar medium within galaxies where SNe occur) is denser than that with average densities $\bar{n}_{\mathrm{ISM}} \sim 1\,\mathrm{cm^{-3}}$, particularly at earlier times when most of the star formation happened. Hence, simulations that spread SNe feedback evenly into the ambient star-forming ISM that has a cooing time of $t_{\mathrm{cool}} \sim 20$ Myr in Eq. (3.181) cannot offset cooling for cosmological time scales (which is the relevant time scale for the evolution of clusters) and thus produce too many stars in the centers of clusters!

3.2.4 Feedback by Active Galactic Nuclei

- There is a compact region at the center of every galaxy that dominates the luminosity of its electromagnetic spectrum, the “active galactic nucleus” (AGN). Observationally, it is known that the AGN emission is caused by mass accretion onto a supermassive black hole (SMBH) which can launch relativistic outflows (so called “jets”). Particle acceleration in jets causes non-thermal radio synchrotron and γ-ray emission.

- The masses of SMBH at the centers of galaxies fall typically in the range of $10^6 \lesssim M_{\mathrm{SMBH}}/\mathrm{M}_\odot \lesssim 10^{10}$. Those SMBH masses are tightly correlated with the stellar mass in galactic bulges. Bulges are defined as the central spheroidal stellar component of a disk galaxy (“late types”) or the entire elliptical stellar distribution of ellipticals (“early types”), including the population of bright central galaxies (BCGs) in clusters.

- The mass of the stellar bulge and the SMBH obey the correlation

$$M_{\mathrm{SMBH}} \sim 0.005 M_{\mathrm{bulge}}, \tag{3.199}$$

so that we obtain typical masses for SMBHs at the centers of clusters according to

$$M_{\star,\mathrm{BCG}} \sim 10^{12}\,\mathrm{M}_\odot \quad \Rightarrow \quad M_{\mathrm{SMBH}} \sim 5\times 10^9 \mathrm{M}_\odot, \tag{3.200}$$

upon identifying $M_{\star,\mathrm{BCG}}$ with the bulge mass. This compares well with the latest mass measurement of the SMBH in M87 of $6 \times 10^9\,\mathrm{M}_\odot$ (M87 is the BCG in Virgo, our closest galaxy cluster with $D_{\mathrm{Virgo}} \sim 17$ Mpc).

- The accretion power onto the SMBH can be estimated by the release of gravitational energy with a radiative efficiency of $\eta \sim 0.1$,

$$E_{\mathrm{AGN}} \sim \eta M_{\mathrm{SMBH}} c^2 \sim 10^{63} \left(\frac{M_{\mathrm{SMBH}}}{5\times 10^9\,\mathrm{M}_\odot}\right) \mathrm{erg} \tag{3.201}$$

$$\frac{E_{\mathrm{AGN}} m_{\mathrm{p}}}{M_{\mathrm{gas}}} \sim \frac{10^{63}\,\mathrm{erg}\, m_{\mathrm{p}}}{10^{14}\,\mathrm{M}_\odot} \sim \frac{10^{63-14-24-33}}{2} \frac{\mathrm{erg}}{\mathrm{nucleon}} \sim 5\, \frac{\mathrm{keV}}{\mathrm{nucleon}}. \tag{3.202}$$

From the energetic viewpoint, this is a much more promising heating source in comparison to supernova feedback.

- The centers of many (if not all) cool core clusters with low-entropy gas whose cooling time is less than the age of the universe also contain AGNs. Relativistic jets from these AGNs inflate radio-emitting lobes of typical radii $r_{\mathrm{lobe}} \sim (10 \ldots 50)$ kpc and distances to the central AGN of $R \sim (20 \ldots 100)$ kpc. The location of these radio-emitting lobes coincide with cavities in X-ray maps. This suggests that the momentum of the relativistic outflow has been slowed down by the inertia of the ICM ($n_{\mathrm{ICM}} \sim 10^{-2}\,\mathrm{cm}^{-3}$, $k_{\mathrm{B}}T \sim 3\,\mathrm{keV}$) which got pushed away by the jet fluid in the process of inflating the lobes.

- **Emerging picture.** As the central gas is cooling, it can eventually form stars and feed the accretion disk of the AGN. The accreting gas has to lose its angular momentum through dissipating hydrodynamic processes (i.e., viscosity, turbulence, or shocks) which radiate away energy and (angular) momentum. Accreting a substantial amount of gas onto the supermassive black hole triggers the black hole-accretion disk system to launch a relativistic jet that is composed of cosmic rays (relativistic particle populations) and magnetic fields. Eventually, the jet momentum slows down due to the ram pressure of the ambient ICM and inflates lobes of relativistic plasma. As the jets terminate, the lobes detach from the ceasing outflows. Because of the relativistic filling, the lobes or bubbles are lighter than the heavier ambient ICM. Since these bubbles got injected at the bottom of the gravitational cluster potential, we have a convectively unstable situation and the bubbles start to rise buoyantly and subsonically in the stratified cluster atmosphere.

- The relativistic jets displace the ICM at the location of the cavities, i.e., they do PdV work against the ICM, as well as supply internal energy, U, to the cavities. Hence the total energy required to create the cavity is equal to its enthalpy,

  $$H = U + PV = \frac{1}{\gamma_b - 1} PV + PV = \frac{\gamma_b}{\gamma_b - 1} PV = 4PV, \qquad (3.203)$$

  where we used $\gamma_b = 4/3$ (assuming a relativistic filling of the radio emitting bubbles). Of this $4PV$, only $1PV$ is directly available for mechanical work on the surroundings while $3PV$ are stored as internal energy. Only a fraction of the mechanical energy resulting from the PdV work done on the surroundings is dissipated in the central regions, which have the smallest cooling time (since most of the temperature increase caused by weak shocks is provided adiabatically).

- There are two possibilities how to dissipate the remaining part of $3PV$ that is stored in the internal energy of the (presumably) relativistic particle population. (1) If these lobes rise buoyantly and unimpeded over several pressure scale heights, the relativistic lobe filling does PdV work on the surroundings, which expands the lobe and transfers the internal lobe energy adiabatically to mechanical energy of the ambient ICM. (2) However, *if* those cosmic rays were mixed into the ambient gas early on during their buoyant rise, there would be a promising process to transfer cosmic ray into thermal energy. Fast-streaming cosmic rays along the magnetic field excite Alfvén waves[2] through the "streaming

---

[2] An Alfvén wave is a low-frequency, transverse electromagnetic wave that propagates alongside or opposite to the direction of the magnetic field, see Section 3.3.2.6 for more detail.

instability". Scattering off of this wave field effectively limits the bulk speed of cosmic rays. Damping of these waves transfers cosmic ray energy and momentum to the thermal gas at a rate that scales with the cosmic ray pressure gradient. Hence, this could provide an efficient means of suppressing the cooling catastrophe of cooling cores, but only *if* the cosmic ray pressure gradient is sufficiently large.

- Hence, the work done by the two bubbles in one outburst (as supposed to the many accompanying SMBH growth, as implied in Eq. (3.201)) on the ambient ICM is

$$W = PV = 2 \times \frac{4}{3}\pi r^3 n_\mathrm{a} k_\mathrm{B} T \sim 10^{59}\,\mathrm{erg}, \tag{3.204}$$

  where we used a bubble radius $r = 20\,\mathrm{kpc}$, an ambient ICM density $n_\mathrm{a} = 10^{-2}\,\mathrm{cm}^{-3}$, and an ICM temperature $k_\mathrm{B}T = 3$ keV.

- There are (at least) three different ways to estimate the bubble's rise time, using 1. the sound crossing time, 2. the buoyant rise time, and 3. the time required for the ambient medium to refill the displaced volume as the bubble rises upward.

  1. The sound crossing time of the distance from the cavity center to the SMBH (using $\gamma_\mathrm{a} = 5/3$ for the ambient ICM) is given by

  $$t_\mathrm{s} = R\sqrt{\frac{\mu m_\mathrm{p}}{\gamma_\mathrm{a} k_\mathrm{B} T}} \approx 3.5 \times 10^7 \left(\frac{R}{40\,\mathrm{kpc}}\right)\left(\frac{k_\mathrm{B}T}{3\,\mathrm{keV}}\right)^{-1/2}\,\mathrm{yr}. \tag{3.205}$$

  2. To estimate the buoyancy time, we compute the buoyancy force acting upon the bubble

  $$\boldsymbol{F}_\mathrm{buoy} = -\boldsymbol{g}V(\rho_\mathrm{a} - \rho_\mathrm{b}), \tag{3.206}$$

     where $\boldsymbol{g}$ is the gravitational acceleration (assuming hydrostatic equilibrium of the ambient gas), V is the bubble volume, ρ_a and ρ_b denote the mass density of the ambient gas and the bubble, respectively. The ram pressure exerts a drag force on the bubble, oppositely directed to the rise velocity,

  $$\boldsymbol{F}_\mathrm{drag} = -\frac{C}{2}\sigma\rho_\mathrm{a} v^2 \frac{\boldsymbol{v}}{v}, \tag{3.207}$$

     where σ is the cross section of the bubble, C is the drag coefficient that depends on bubble geometry and Reynolds number (i.e., whether the flow is turbulent or laminar): $C \approx 0.6$ for a Mach number $\mathcal{M} \approx 0.7$. In equilibrium, the terminal velocity is obtained by balancing $|\boldsymbol{F}_\mathrm{buoy}|$ and $|\boldsymbol{F}_\mathrm{drag}|$, yielding

  $$v = \sqrt{\frac{2gV}{\sigma C}\frac{\rho_\mathrm{a} - \rho_\mathrm{b}}{\rho_\mathrm{a}}} \approx \sqrt{\frac{2gV}{\sigma C}}, \tag{3.208}$$

where we assumed $\rho_\mathrm{b} \ll \rho_\mathrm{a}$ in the last step. For a singular isothermal sphere (SIS), we can write down $g \approx v_\mathrm{c}^2/R = 2\sigma^2/R = 2k_\mathrm{B}T/(\mu m_\mathrm{p}R)$. With $\sigma = \pi r^2$ and $V = 4\pi r^3/3$, we obtain

$$t_\mathrm{buoy} \approx R\sqrt{\frac{\sigma C}{2gV}} \approx t_\mathrm{s}\sqrt{\frac{3C\gamma_\mathrm{a}}{16}\frac{R}{r}} \approx 0.6 t_\mathrm{s}\left(\frac{R}{2r}\right)^{1/2} . \qquad (3.209)$$

3. Finally, we can ask how the dense gas directly above the bubbles stays at its position. As the gas ahead of the bubble is denser than the surrounding gas due to displacement as a result of the bubble formation, it has negative buoyancy. According to Archimedes' principle, the net force per unit volume on overdense gas is $\delta\rho g$, where $\delta\rho$ is the difference between its density and that of the surrounding gas, and $g = GM(R)/R^2$ is the gravitational acceleration, where R is the distance to the cluster center. Hence, the acceleration of the gas is $a = g\delta\rho/\rho$, where ρ is the density of the gas. If this is not balanced, the gas is accelerated inwards and falls a distance r in the time

$$t_\mathrm{refill} \approx \sqrt{\frac{2r}{a}} = \sqrt{\frac{2r}{g}\frac{\rho}{\delta\rho}} = 2R\sqrt{\frac{r}{GM(R)}}, \qquad (3.210)$$

where we have assumed a negligible initial velocity of the gas and $\delta\rho/\rho = 0.5$. This time scale is equal to the time required to refill the volume as the bubble rises upward and can be simplified to give

$$t_\mathrm{refill} \approx 2R\sqrt{\frac{r}{GM(R)}} \approx t_\mathrm{s}\sqrt{\frac{2\gamma_\mathrm{a} r}{R}} \approx 1.3 t_\mathrm{s}\left(\frac{2r}{R}\right)^{1/2} . \qquad (3.211)$$

In the second step, we used the potential of the SIS, $\Phi_\mathrm{SIS} = GM/R = 2k_\mathrm{B}T/(\mu m_\mathrm{p})$.

- This demonstrates that all three estimates provide similar results (albeit with a different scaling in the ratio r/R.) We finally obtain the AGN heating rate by combining Eqs. (3.204) and (3.209),

$$L_\mathrm{AGN} \approx \frac{PV}{t_\mathrm{buoy}} \approx \frac{10^{59}\,\mathrm{erg}}{10^{15}\,\mathrm{s}} \approx 10^{44}\,\mathrm{erg\ s^{-1}} \approx L_\mathrm{X}, \qquad (3.212)$$

i.e., it is comparable to the X-ray "cooling" luminosity in Eq. (3.187).

- There are a number of open questions in this scenario which are currently being actively researched.

1. How is the accretion output *thermalized*? Several physical processes have been proposed to be responsible for the heating, including dissipation of mechanical energy by shocks or sound waves from the central AGN, escaping cosmic rays from the radio lobes excite Alfvén waves in the magnetized ICM which get damped and thereby heat the surrounding thermal plasma, turbulent mixing of the cold gas with the hotter phase which increases the net cooling time, redistribution of heat by buoyancy-induced turbulent convection, or an inward conductive transport of heat from the hot outer cluster regions.
2. Is the heating-cooling balance *stable* to local thermal perturbations? While turbulent or conductive heating are not stable (see Section 3.2.6), cosmic-ray Alfvén wave heating is stabilized around 1 keV, which coincides with the lower temperature floors observed at the centers of cool core clusters.
3. How is the accretion rate *tuned*? The Schwarzschild radius of a SMBH is

   $$r_{\mathrm{SMBH}} = \frac{2GM_{\mathrm{SMBH}}}{c^2} \approx 1.3 \times 10^{15}\,\mathrm{cm} \approx 1\,\text{light day}. \tag{3.213}$$

   On the contrary, cooling occurs on scales of about 30 kpc $\approx 10^{23}$ cm $\approx 10^8\, r_{\mathrm{SMBH}}$. How does the SMBH know about the cooling on scales 10^8 times larger and how does it inject just the right amount of energy to self-regulate the system? Remember, if cooling would proceed for several cooling times without being counteracted, we would get run-away cooling that forms a huge amount of stars, which are not observed. It occurs that gas accretion, jet lauching and propagation, and the associated heating rate are on average well regulated despite the large range of scales.

- At least Nature finds a way to solve all of these problems because observationally, SMBH activity accompanies transition to complexity when $t_{\mathrm{cool}} \lesssim 1$ Gyr.

3.2.5 Heat Conduction

3.2.5.1 Derivation

- A system can be in hydrostatic equilibrium, but out of thermal equilibrium. In the absence of viscosity, the entropy equation for smooth, differentiable flows reads

$$\rho T \frac{\mathrm{d}s}{\mathrm{d}t} = \boldsymbol{\nabla} \cdot (\kappa \boldsymbol{\nabla} T)\,. \tag{3.214}$$

Using $\mathrm{d}q = T\mathrm{d}s$ and $c_P \equiv (\mathrm{d}q/\mathrm{d}T)_P$, we get

$$c_P \mathrm{d}T = T\mathrm{d}s \quad \Rightarrow \quad \mathrm{d}s = c_P \mathrm{d}\ln T. \tag{3.215}$$

- Hence, we can rewrite Eq. (3.214) while assuming $\kappa = \text{const.}$,

  $$\rho c_P \frac{\mathrm{d}T}{\mathrm{d}t} = \kappa \boldsymbol{\nabla}^2 T \quad \text{or} \quad \frac{\mathrm{d}T}{\mathrm{d}t} = \chi \boldsymbol{\nabla}^2 T, \tag{3.216}$$

  where $\chi \equiv \kappa/\rho c_P$. This shows that the temperature of a smooth flow can only change as a result of thermal conduction if $\boldsymbol{\nabla} T \neq \mathbf{0}$ since the temperature gradient is the source of free energy.

- We now want to estimate the heat conductivity κ. To this end, we consider a system in thermal equilibrium with a temperature T and with particles moving randomly in all directions. ΔA denotes the area of a screen perpendicular to the x axis. The number of particles that fly per unit time with an *rms* velocity v through the screen from one side to the other is given by

  $$\frac{\Delta N}{\Delta t} = \frac{nv\Delta A}{6}, \tag{3.217}$$

  where the factor of 6 arises because on average, 1/3 of all particles fly along the x axis and of those, only 1/2 in either direction.

- The particle mean free path is $\lambda_{\mathrm{mfp}} = 1/(n\sigma)$ where σ is the collisional cross section. Particles at $x - \lambda_{\mathrm{mfp}}$ transport gas properties to x and vice versa. This is particularly important for gradients in gas properties that will be smoothed out as a result of such a transport.

- Hence, in the presence of a density gradient, $\partial n/\partial x \neq 0$, the net number of particles flying from the denser to the more dilute region is

  $$\frac{\Delta N}{\Delta t} = \frac{n(x+\lambda_{\mathrm{mfp}})v\Delta A}{6} - \frac{n(x-\lambda_{\mathrm{mfp}})v\Delta A}{6} \approx \frac{v\Delta A}{6}\frac{\partial n}{\partial x} 2\lambda_{\mathrm{mfp}}, \tag{3.218}$$

  where we have expanded the density field to first order and have assumed that the typical length of a gradient Δx is much larger than the mean free path, $\lambda_{\mathrm{mfp}} \ll \Delta x$.

- The diffusion coefficient that relates the particle current $j = \Delta N/(\Delta t \Delta A)$ to the number density gradient is given by

  $$\frac{\Delta N}{\Delta t \Delta A} \overset{!}{=} D\frac{\partial n}{\partial x} \quad \text{where} \quad D \equiv \frac{v\lambda_{\mathrm{mfp}}}{3} = \frac{v}{3n\sigma}. \tag{3.219}$$

  If the temperature changes along x (i.e., $\partial T/\partial x \neq 0$), the particles transport energy,

  $$\frac{\Delta E}{\Delta t \Delta A} = \frac{nv}{6}[E(x+\lambda_{\mathrm{mfp}}) - E(x-\lambda_{\mathrm{mfp}})] \tag{3.220}$$

  $$= \frac{nv\lambda_{\mathrm{mfp}}}{3}\left(\frac{\partial E}{\partial T}\frac{\partial T}{\partial x}\right) = \frac{nvc_V\lambda_{\mathrm{mfp}}}{3}\frac{\partial T}{\partial x}, \tag{3.221}$$

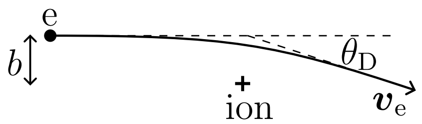
where c_V is the heat capacity at constant volume, see Eq. (3.3). Hence, we find

$$\frac{\Delta E}{\Delta t \Delta A} \overset{!}{=} \kappa \frac{\partial T}{\partial x} \quad \text{where} \quad \kappa = \frac{n v c_V \lambda_{\mathrm{mfp}}}{3} = \frac{v c_V}{3\sigma} = \frac{v k_{\mathrm{B}}}{2\sigma}, \tag{3.222}$$

where we used the heat capacity at constant volume $c_V = 3k_{\mathrm{B}}/2$ (Eq. 3.3) in the last step (assuming an ideal, monoatomic gas) and the heat conductivity κ has units of erg $\mathrm{cm}^{-1}\ \mathrm{s}^{-1}\ \mathrm{K}^{-1}$.

- Heat is conducted by electrons since they move faster than ions by $v_{\mathrm{e}}/v_{\mathrm{i}} = \sqrt{m_{\mathrm{i}}/m_{\mathrm{e}}} \approx 43\sqrt{Z}$ (assuming $T_{\mathrm{e}} = T_{\mathrm{i}}$ which applies to the ICM except for immediate post-shock regions). The electron mean free path is determined by the ion number density and the scattering cross section, implying $\lambda_{\mathrm{mfp}} = 1/(n_{\mathrm{i}}\sigma)$.

3.2.5.2 Coulomb Logarithm



An electron scattering in the Coulomb field of an ion.

- Let's first consider an electron scattering in the Coulomb field of an ion:

$$m_{\mathrm{e}} \frac{\mathrm{d}\boldsymbol{v}_{\perp,\mathrm{e}}}{\mathrm{d}t} = \boldsymbol{\nabla}_{\perp}\phi_{\mathrm{i}}. \tag{3.223}$$

In the limit of a small deflection angle, we have $\boldsymbol{e}_{\perp} \cdot \boldsymbol{v}_{\perp,\mathrm{e}} = v_{\perp,\mathrm{e}} \approx v_{\mathrm{e}}\,\theta_{\mathrm{d}}$.

1. If the deflection angle is small, $\theta_{\mathrm{d}} \ll 1$, we can approximate its value by computing the perpendicular impulse exerted by the ion's Coulomb field, integrating along the electron's unperturbed straight line trajectory (the "Born approximation")

$$\begin{aligned} m_{\mathrm{e}} v_{\mathrm{e}} \theta_{\mathrm{d}} &= \int_{-\infty}^{\infty} \boldsymbol{e}_{\perp} \cdot \boldsymbol{\nabla}_{\perp}\phi_i \mathrm{d}t = \int_{-\infty}^{\infty} \frac{\partial}{\partial b}\left(\frac{Ze^2}{\sqrt{b^2 + v_{\mathrm{e}}^2 t^2}}\right)\mathrm{d}t \\ &= \int_{-\infty}^{\infty} \frac{Ze^2 b \mathrm{d}t}{(b^2 + v_{\mathrm{e}}^2 t^2)^{3/2}} = \frac{Ze^2}{v_{\mathrm{e}}} \int_{-\infty}^{\infty} \frac{b^2 \mathrm{d}x}{b^3 (1 + x^2)^{3/2}} \\ &= \frac{Ze^2}{v_{\mathrm{e}} b} \left.\frac{x}{\sqrt{1 + x^2}}\right|_{-\infty}^{\infty} = \frac{2Ze^2}{b v_{\mathrm{e}}}, \end{aligned} \tag{3.224}$$

   where b is the impact parameter of the electron's trajectory and $\boldsymbol{e}_{\perp} \cdot \boldsymbol{\nabla}_{\perp} = \partial/\partial b$. Hence we obtain

$$\theta_{\mathrm{d}} = \frac{b_0}{b} \quad \text{for} \quad b \gg b_0 \equiv \frac{2Ze^2}{m_{\mathrm{e}} v_{\mathrm{e}}^2}. \tag{3.225}$$

2. If the dominant source of this electron deflection were a single large-angle scattering event in the Coulomb field of an

ion, then the relevant cross section would be $\sigma = \pi b_0^2$ (since all impact parameters $\lesssim b_0$ produce large-angle scatterings) and the mean deflection frequency ν_d and time t_d would be

$$\nu_\mathrm{d} = \frac{1}{t_\mathrm{d}} = n_\mathrm{i}\sigma v_\mathrm{e} = n_\mathrm{i}\pi b_0^2 v_\mathrm{e} \text{ (for large-angle scattering).} \tag{3.226}$$

- The cumulative, random-walk effects of many small-angle electron scatterings off ions produce a net deflection of order a radian in a shorter time. As the directions of the individual scatterings are random, the mean defection angle after many scatterings vanish, $\langle\theta\rangle = 0$. However, $\langle\theta^2\rangle$ will not vanish and we have

$$\langle\theta^2\rangle = \sum_{\text{all encounters}} \theta_\mathrm{d}^2 = \sum_{\text{all encounters}} \left(\frac{b_0}{b}\right)^2 . \tag{3.227}$$

- The number of encounters that occur with impact parameters between b and $b + \mathrm{d}b$ during time t is $\mathrm{d}N = n_\mathrm{i} v_\mathrm{e} t 2\pi b \mathrm{d}b$. Hence the mean square deflection angle accumulates up to

$$\langle\theta^2\rangle = \int_{b_\mathrm{min}}^{b_\mathrm{max}} \left(\frac{b_0}{b}\right)^2 \mathrm{d}N = n_\mathrm{i} 2\pi b_0^2 v_\mathrm{e} t \ln\left(\frac{b_\mathrm{max}}{b_\mathrm{min}}\right). \tag{3.228}$$

- While the integral diverges logarithmically, physics regularizes it quite naturally. The minimum impact parameter,

$$b_\mathrm{min} = \frac{Ze^2}{k_\mathrm{B}T}, \tag{3.229}$$

equals the radius where the Coulomb energy of the electron in the field of the ion vanishes, $E = mv^2/2 - Ze^2/b_\mathrm{min} \overset{!}{=} 0$. The maximum impact parameter is given by the maximum distance over which electric fields of individual particles can reach without being screened by the oppositely charged particles in a plasma. This is known as the Debye length,

$$b_\mathrm{max} = \lambda_\mathrm{d} = \sqrt{\frac{k_\mathrm{B}T}{4\pi n_\mathrm{e} Ze^2}}. \tag{3.230}$$

- Hence, we can define the Coulomb logarithm

$$\ln\Lambda = \ln\left(\frac{b_\mathrm{max}}{b_\mathrm{min}}\right) = \ln\sqrt{\frac{(k_\mathrm{B}T)^3}{4\pi n_\mathrm{e} Z^3 e^6}} \tag{3.231}$$

$$= 35 - \frac{1}{2}\ln\left(\frac{n_\mathrm{e}}{10^{-2}\mathrm{cm}^{-3}}\right) + \frac{3}{2}\ln\left(\frac{k_\mathrm{B}T}{\mathrm{keV}}\right). \tag{3.232}$$

- The value of t that implies $\langle\theta^2\rangle \approx 1$ is the deflection time t_d,

  $$\nu_\mathrm{d}^\mathrm{ei} = \frac{1}{t_\mathrm{d}^\mathrm{ei}} = n_\mathrm{i} 2\pi b_0^2 \upsilon_\mathrm{e} \ln\Lambda = \frac{8\pi n_\mathrm{i} Z^2 e^4}{m_\mathrm{e}^2 \upsilon_\mathrm{e}^3} \ln\Lambda \qquad (3.233)$$

  and $\ln\Lambda \approx 35\ldots40$ in the ICM. This deflection frequency is larger by a factor of $2\ln\Lambda$ than the frequency of Eq. (3.226), which is valid for a single large-angle scattering event.

- Back to our heat conductivity of electrons,

  $$\kappa = \frac{n_\mathrm{e}\upsilon_\mathrm{e} c_V \lambda_\mathrm{mfp}}{3} = \frac{n_\mathrm{e}\upsilon_\mathrm{e} c_V}{3\sigma n_\mathrm{i}}. \qquad (3.234)$$

  From Eq. (3.233), we can read off σ by remembering $\nu_\mathrm{d} = n_\mathrm{i}\sigma\upsilon_\mathrm{e}$:

  $$\sigma = 2\pi b_0^2 \ln\Lambda = \frac{8\pi Z^2 e^4 \ln\Lambda}{m_\mathrm{e}^2 \upsilon_\mathrm{e}^4}. \qquad (3.235)$$

  This yields the heat conductivity of electrons that are scattered by ions in a thermal gas,

  $$\kappa = \frac{n_\mathrm{e}\upsilon_\mathrm{e}}{3} c_V \frac{m_\mathrm{e}^2 \upsilon_\mathrm{e}^4}{8\pi n_\mathrm{i} Z^2 e^4 \ln\Lambda} = \frac{1}{3}\left(\frac{m_\mathrm{e}^2}{8\pi Z^2 e^4}\right)\left(\frac{n_\mathrm{e}}{n_\mathrm{i}}\right)\frac{c_V \upsilon_\mathrm{e}^5}{\ln\Lambda}. \qquad (3.236)$$

- The heat capacity at constant volume is $c_V = 3k_\mathrm{B}/2$ and the thermal electron velocity is $\upsilon_\mathrm{e} = \sqrt{2k_\mathrm{B}T_\mathrm{e}/m_\mathrm{e}}$. Inserting these expressions into Eq. (3.236) yields a value for the heat conductivity

  $$\begin{aligned} \kappa &= \frac{k_\mathrm{B}}{2}\left(\frac{m_\mathrm{e}^2}{8\pi Z^2 e^4}\right)\left(\frac{n_\mathrm{e}}{n_\mathrm{i}}\right)\left(\frac{2k_\mathrm{B}T_\mathrm{e}}{m_\mathrm{e}}\right)^{5/2}\frac{1}{\ln\Lambda} \\ &= 2.17\times10^{-6}\left(\frac{T}{1\,\mathrm{K}}\right)^{5/2}\left(\frac{\ln\Lambda}{1}\right)^{-1}\frac{\mathrm{erg}}{\mathrm{s\,K\,cm}} \\ &= 6.2\times10^{12}\left(\frac{T}{10^8\,\mathrm{K}}\right)^{5/2}\left(\frac{\ln\Lambda}{35}\right)^{-1}\frac{\mathrm{erg}}{\mathrm{s\,K\,cm}}, \end{aligned} \qquad (3.237)$$

  where we have used appropriate values for the Coulomb logarithm in cool core regions in clusters.

- The strong temperature dependence of κ is a consequence of the velocity dependence ($\kappa \propto T^{5/2} \propto \upsilon_\mathrm{e}^5$). Four powers of which derive from the cross section $\sigma \propto b_0^2$ where $b_0 \equiv 2Ze^2/(m_\mathrm{e}\upsilon_\mathrm{e}^2)$ results from balancing the kinetic energy with the potential energy during a scattering event and one power results from $\kappa \propto \upsilon_\mathrm{e}\lambda_\mathrm{mfp}/3$, so that the conductive heat flux scales as $\boldsymbol{Q} = \kappa\boldsymbol{\nabla}T \propto T^{7/2}$.

- The critical physics assumption behind this derivation of conduction is the random walk of electrons (along a magnetic field line); however plasma physics in form of electron scattering by Whistler waves modifies this in the weakly collisional plasma of the ICM!

3.2.6 Thermal Instability

3.2.6.1 Field Length

- Cool star forming clouds should only appear in systems whose size is *greater* than a critical length scale, known as the *Field length* below which thermal conduction smoothes out temperature inhomogeneities. Formally we would have to a Lagrangian perturbation analysis to derive this length scale. Instead, we will derive the Field length heuristically by considering thermal balance for a cool cloud of radius r embedded in a medium of temperature T.

- Electron thermal conduction sends energy into the cloud at a rate

$$\mathcal{H}_{\rm cond} \sim r^2 \kappa(T) \frac{T}{r} \sim \kappa_0 f_{\rm e} r \frac{T^{7/2}}{T_8^{5/2}}. \tag{3.238}$$

Here, $T_8 = 10^8$ K, $f_{\rm e}$ is a magnetic suppression factor that depends on the topology of magnetic field lines connecting our cloud of consideration, and we used the *Spitzer conductivity* (which assumes a value for the Coulomb logarithm of $\ln\Lambda = 35$),

$$\kappa = 6.2 \times 10^{12} \left(\frac{T}{10^8\,\mathrm{K}}\right)^{5/2} f_{\rm e} \frac{\mathrm{erg}}{\mathrm{s\ K\ cm}} = \kappa_0 f_{\rm e} \left(\frac{T}{T_8}\right)^{5/2}. \tag{3.239}$$

- Radiative cooling can radiate away energy at a rate

$$C_{\rm rad} \sim r^3 n_{\rm H}^2 \Lambda_0(T) \sim r^3 n_{\rm H}^2 \Lambda_0 \left(\frac{T}{T_8}\right)^{1/2}, \text{ with} \tag{3.240}$$

$$\Lambda_0(T) \approx 2.5 \times 10^{-23} \left(\frac{T}{T_8}\right)^{1/2} \frac{\mathrm{erg\ cm^3}}{\mathrm{s}}, \tag{3.241}$$

where we have used Eq. (3.185).

- Cooling and conduction are thus in approximate balance, $\mathcal{H}_{\rm cond} \sim C_{\rm rad}$, for systems with a radius of order the Field length

$$\lambda_{\rm F} \equiv \left[\frac{T\kappa(T)}{n_{\rm H}^2 \Lambda_0(T)}\right]^{1/2} = \left(\frac{\kappa_0 f_{\rm e} x_{\rm e}^2}{k_{\rm B} \Lambda_0 k_{\rm B}^2 T_8^2}\right)^{1/2} K_{\rm e}^{3/2} \tag{3.242}$$

$$\approx 6.6\,\mathrm{kpc} \left(\frac{K_{\rm e}}{20\,\mathrm{keV\ cm^2}}\right)^{3/2} f_{\rm e}^{1/2}, \tag{3.243}$$

where we have used $K_{\rm e} = k_{\rm B}T/n_{\rm e}^{2/3}$ and the square of the hydrogen number density is given by $n_{\rm H}^2 = X_{\rm H}^2 \rho^2 / m_{\rm p}^2 = n_{\rm e}^2 / x_{\rm e}^2$. Through a coincidence of scaling, the Field length is a function of entropy alone when free-free emission is the dominant cooling mechanism.

- We can translate this criterion in the entropy-radius plane by adopting $\lambda_{\mathrm{F}}(K) = r$. This yields a thermal stability threshold that obeys a scaling with radius of $K \propto r^{2/3} f_{\mathrm{e}}^{-1/3} = \lambda_{\mathrm{F}}^{2/3} f_{\mathrm{e}}^{-1/3}$. Gas that is below that threshold and resides within radius r constitutes a subsystem with $r > \lambda_{\mathrm{F}}$ (at constant K), i.e., the amount of entropy in the larger cloud is too small to support fast enough conduction that is necessary to prevent a cooling run-away, allowing multiphase gas to persist and star formation to proceed. Gas above the threshold resides in the region of thermal stability in which conduction is fast enough and leads to evaporation of a cool cloud and eventually homogeneity.

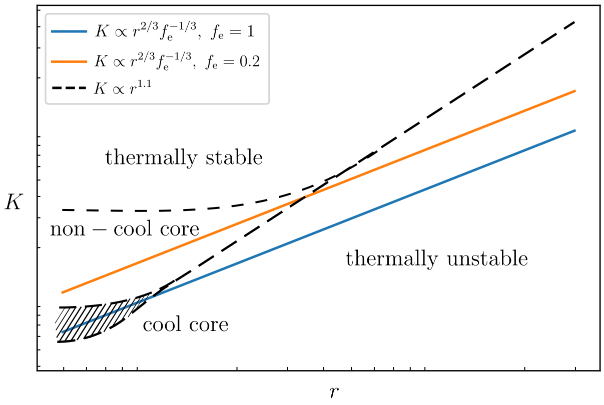


Figure 3.5: Criterion for thermal instability by considering cluster entropy profiles for cool core and non-cool core clusters (dashed) and the Field length by adopting $\lambda_{\mathrm{F}}(K) = r$ for two different heat flux suppression factors. The ICM becomes thermally unstable if it falls below this threshold (see text for details).

- In Section 3.1.7, we found that the entropy profile of the ICM at larger scales shows the behavior $K \propto r^{1.1}$. This leaves us with two possibilities of cluster states in reality (which appear to be dynamical attractor solutions of thermal stability considerations):

  1. Clusters can have an entropy profile that is elevated enough so that it stays always above the thermal stability threshold, see Fig. 3.5. As a consequence, the steeper entropy profile on larger scale necessarily needs to break at sufficiently large radii to join an elevated level of central entropy. This defines the class of *non-cool core clusters*.
  2. Clusters can have an entropy profile that continues to decrease toward smaller radii until it drops below the thermal stability threshold, see Fig. 3.5. There the gas is subject to thermal instability, and multiphase gas can form, potentially seeding star formation. This constitutes the class of *cool core clusters*.

3.2.6.2 Heating versus Cooling: a Visual Stability Analysis

- Formally one needs to do a perturbation analysis of the hydrodynamic equations (continuity, momentum, and entropy equations with thermal conduction), but here we will only sketch the concept and show the main ideas by introducing a visual stability analysis that adopts an energy balance argument. As a cautionary remark, this procedure can yield misleading answers, if the perturbative equations contain interesting dynamics (an example of which we will encounter in the next subsection).

- Because we allow for thermal conduction, the entropy of a fluid element is not any more conserved. Instead, we consider hydrostatic rearrangements that conserve the thermal pressure $P = nk_\mathrm{B}T$ and rewrite the energy deposition and cooling rates as functions of temperature and pressure. We consider radiative cooling (bremsstrahlung and free-free line emission, denoted by C_rad) and heating by conduction ($\mathcal{H}_\mathrm{cond}$), turbulent dissipation ($\mathcal{H}_\mathrm{turb}$), and Coulomb and hadronic interactions of cosmic rays with the thermal gas ($\mathcal{H}_\mathrm{hadr}$) and find the scaling of the volumetric cooling anad heating rates with the temperature, respectively:

$$C_\mathrm{rad} \propto r^3 n^2 \left[T^{1/2} + \Lambda_\mathrm{line}(T)\right] \propto P\left[T^{-1/2} + \frac{\Lambda_\mathrm{line}(T)}{T}\right], \quad (3.244)$$

$$\mathcal{H}_\mathrm{cond} \propto r^2 \kappa(T)\frac{T}{r} \propto r\,T^{7/2} \propto P^{-1/3}T^{23/6}, \quad (3.245)$$

$$\mathcal{H}_\mathrm{turb} \propto r^3 n \propto T^0, \quad (3.246)$$

$$\mathcal{H}_\mathrm{hadr} \propto r^3 n n_\mathrm{cr} \propto f_\mathrm{cr}\frac{P}{T}, \quad (3.247)$$

where $\Lambda_\mathrm{line}(T) \propto T^{\alpha_\mathrm{line}}$, $\alpha_\mathrm{line} < 1/2$ for $k_\mathrm{B}T < 2\,\mathrm{keV}$, and $f_\mathrm{cr} = n_\mathrm{cr}/n$. Assuming that $f_\mathrm{cr} \sim$ const., we find the stability properties of the different heating processes by considering the energy deposition and cooling rates as a function of temperature in Fig. 3.6. To this end, we start at a point of thermal equilibrium, where $C_\mathrm{rad} = \mathcal{H}_i$ (and $\mathcal{H}_i$ symbolizes one of the various heating processes). By perturbing the temperture upwards (i.e., moving from the thermal equilibrium to a higher temperature), we ask whether the volumetric heating rate is larger than the cooling rate. If this is the case (like for conductive heating), then the increased heating causes the temperture to increase further, implying an instability. The same applies for perturbing the temperature downwards. If the volumetric cooling rate is larger than the heating rate, then the increased cooling causes the temperature to decrease further, implying an instability.

- This clearly demonstrates that conductive and turbulent heating cannot be in stable equilibrium with radiative cooling. By contrast, a heating mechanism with an energy deposition rate that

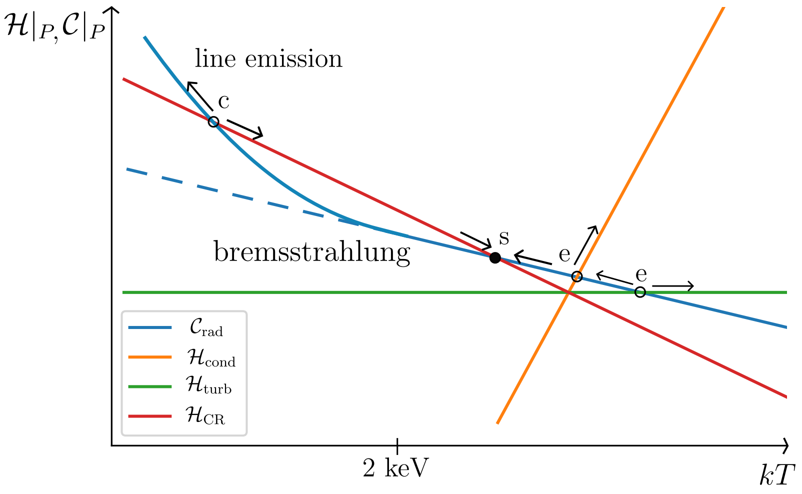


Figure 3.6: Visual stability analysis at constant pressure. Shown are radiative cooling and various energy deposition rates (thermal conduction, turbulent dissipation, cosmic ray streaming) as a function of temperature. We perturb the temperature up and down from an equilibrium of cooling and heating: if heating dominates over cooling, the ICM becomes hotter and vice versa. This causes either a stable fixed point (closed circle with “s”) or an unstable fixed point (open circle with “e” for evaporation and “c” for condensation) depending on whether the ICM is self-regulating or not upon perturbing the temperature of this fixed point.

scales with T^{-1} (such as cosmic ray heating through Coulomb and hadronic interactions) would allow for stable solutions. Hence, if $C_{\rm rad} \propto T^{\alpha_{\rm cool}}$ and $\mathcal{H} \propto T^{\alpha_{\rm heat}}$ then the condition for a stable thermal equilibrium is $\alpha_{\rm heat} < \alpha_{\rm cool}$.

- In practice, Coulomb and hadronic interactions of cosmic rays with the thermal gas are rather slow in the low-density ICM so that the normalization for this heating rate is far below the cooling rate and we cannot maintain an equilibrium with the radative cooling rate. A more promising process is cosmic ray heating via the gyro-resonant interaction with Alfvén waves. A streaming population of cosmic rays resonantly excites Alfvén waves, which experience damping through various collisionless plasma processes, thereby heating the surrounding plasma. The associated volumetric heating rate is

$$\mathcal{H}_{\rm cr} \sim r^3 |\boldsymbol{v}_{\rm A} \cdot \boldsymbol{\nabla} P_{\rm cr}| \qquad (3.248)$$

  where $\boldsymbol{v}_{\rm A} = \boldsymbol{B}/\sqrt{4\pi\rho}$ is the Alfvén velocity, $\boldsymbol{B}$ is the magnetic field, ρ is the mass density, and $P_{\rm cr}$ is the cosmic ray pressure.

- In order to estimate the ability of cosmic ray streaming heating to balance radiative cooling, we calculate the ratio of both rates:

$$\frac{\mathcal{H}_{\rm cr}}{C_{\rm cool}} = \frac{|\boldsymbol{v}_{\rm A}\cdot\boldsymbol{\nabla}P_{\rm cr}|}{n^2\Lambda(T)} \sim \frac{1}{\beta^{1/2}}\frac{P_{\rm cr}}{P_{\rm th}}\frac{H_{\rm th}}{H_{\rm cr}}\frac{\tau_{\rm cool}}{\tau_{\rm ff}} \sim O(1), \tag{3.249}$$

where $\beta = P_{\rm th}/P_B$ is the plasma β parameter, i.e., the ratio of thermal to magnetic pressure, $H_{\rm th}$ and $H_{\rm cr}$ are the scale heights of thermal and cosmic ray pressures, respectively, $\tau_{\rm cool}$ is the radiative cooling timescale of the ICM, and $\tau_{\rm ff} = H_{\rm gas}/c_{\rm sd}$ is the free-fall timescale (assuming approximate hydrostatic equilibrium and $c_{\rm sd}$ is the sound speed). Given that (i) typically $\beta \sim 10^2$, (ii) the cosmic ray pressure profile can be locally much steeper than the thermal gas profile, (we assume $H_{\rm th}/H_{\rm cr} \sim 10$ for the sake of the argument), (iii) $\tau_{\rm cool}/\tau_{\rm ff} \sim 10$ when the gas is locally thermally unstable as is often the case in cool cores, and (iv) the ratio of cosmic ray-to-gas pressure is small (here assumed to be 10% for the sake of the argument), this estimate shows that cosmic ray heating can be competitive with radiative cooling for reasonable choices of model parameters.

- To understand whether cosmic ray streaming heating can also provide a stable heating at constant (local) pressure, we assume magnetic flux freezing for isotropic volume changes so that $B = \sqrt{\boldsymbol{B}^2} \propto n^{2/3}$ (Section 3.3.2.5), adiabatic cosmic rays so that $P_{\rm cr} \propto n^{4/3}$, and $r^2 \propto n^{-2/3}$ to obtain:

$$\mathcal{H}_{\rm cr} \propto r^3\frac{B}{n^{1/2}}\frac{P_{\rm cr}}{r} \propto n^{-2/3+2/3-1/2+4/3} \propto n^{5/6} \propto \left(\frac{P}{T}\right)^{5/6}. \tag{3.250}$$

  If we compare this heating rate to the radiative cooling rate in Fig. 3.6 (red vs. blue curves), we find a stable equilibrium above ~ 2 keV because $\alpha_{\rm heat} = -5/6 < \alpha_{\rm cool} = -1/2$.

- Of course, the final state of the system depends on boundary conditions and conservation laws as discussed in the following.
  - **Thermal conduction.** Here the free energy is borrowed from the temperature gradient $\boldsymbol{\nabla}T$ and heating comes to an end once a constant temperature profile has been reached.
  - **Turbulent dissipation** causes the temperature to increase until $3k_{\rm B}T \sim mv^2$ or if all turbulent kinetic energy has been dissipated.
  - **Cosmic ray streaming** requires a source of cosmic rays close to the center and a sufficiently steep pressure gradient to provide a large enough cosmic ray flux. This can be achieved by AGNs, but only during periods of active jet launching.

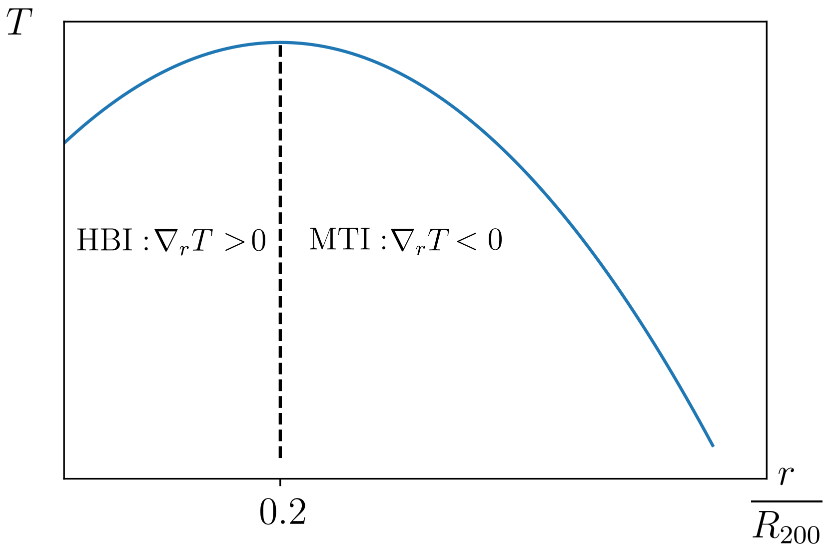


Figure 3.7: Typical temperature profile of a cool core cluster. Shown are the regions in which the magneto-thermal instability (MTI) and the heat-flux driven buoyancy instability (HBI) can be excited.

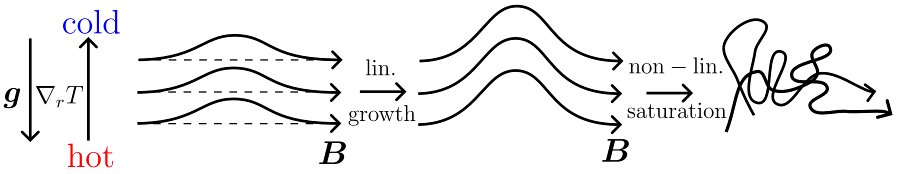


Figure 3.8: Sketch of the evolution of magnetic field lines of the MTI.

3.2.6.3 Thermal Stability with Magnetic Fields

- In a weakly collisional magnetized plasma which we encounter in a galaxy cluster, electrons cannot move "freely" but are bound to follow and gyrate around magnetic field lines. This modifies the convective stability criterion and the type of instability depends on the sign of $\boldsymbol{\nabla} T$.

- The typical temperature profile $T(r)$ of a cool core cluster is depicted in Fig. 3.7. In the center $T(r)$ increases to reach a maximum at a radius of around $0.2\,R_{200}$ and decreases again towards larger radii. As will explain below, the magneto-thermal instability (MTI, Balbus 2001) can only be excited in the outer cluster regions where $\boldsymbol{e}_r \cdot \boldsymbol{\nabla} T < 0$. Conversely, the heat-flux driven buoyancy instability (HBI, Quataert 2008) is excited in the cooling core region where $\boldsymbol{e}_r \cdot \boldsymbol{\nabla} T > 0$. Note that both types of buoyancy instabilities would be absent without magnetic fields because the ICM is stably stratified according to the Schwarzschild criterion of convective stability because in clusters, the entropy increases with radius.

1. **Magneto-thermal instability.** First, we consider the conceptually simpler MTI for which the temperature gradient points in the direction of gravity. We assume that magnetic fields are initially aligned horizontally (or only consider the horizontal magnetic field component). Displacing a volume element upwards in the gravitational potential would cause it to adiabatically expand and to cool in the absence of conduction. Instead, it is conductively heated from the hotter heat bath below to which it is connected by the magnetic field. This causes further expansion and dilution so that the volume element continues to rise as it remains lighter than the surrounding ICM. Hence, this dynamics reinforces the cause of the evolution, giving rise to instability (see Fig. 3.8).

   Non-linear simulations of the MTI show that it does not quiescently saturate with a radial field (as expected from the linear stability analysis) but in a turbulent state. The reason for this is that the radial field configuration is overstable, i.e., the magnetic field always overshoots this radial configuration while its amplitude continues to grow.

2. **Heat-flux driven buoyancy instability.** If the temperature gradient is antiparallel to the direction of gravity, the ICM is susceptible to exciting the HBI. To understand this instability, we introduce the displacement field $\boldsymbol{\xi} \equiv \mathrm{i}\delta\boldsymbol{v}/\omega \propto \delta\boldsymbol{v}$. We assume incompressible gas and consider Fourier transformed background and perturbed quantities. As always, we consider a background at rest, so that $\boldsymbol{v} = \boldsymbol{0}$ and find

   $$\boldsymbol{\nabla}\cdot\delta\boldsymbol{v} = 0 \quad \Rightarrow \quad \boldsymbol{k}\cdot\delta\boldsymbol{v} = 0 \quad \Rightarrow \quad \boldsymbol{k}\cdot\boldsymbol{\xi} = 0, \qquad (3.251)$$

   so that $\boldsymbol{k}$ is perpendicular to $\boldsymbol{\xi}$. If the temperature gradient increases outwards, there will be a background heat flux $\boldsymbol{Q}_0$ pointing inwards along the unperturbed radial magnetic field lines (our assumed initial state). Oblique perturbations with a wave vector $\boldsymbol{k}$ at some angle with the magnetic field perturbs $\boldsymbol{B}$ perpendicular to $\boldsymbol{k}$ (because of the incompressibility condition, $\boldsymbol{\nabla}\cdot\boldsymbol{v} = 0$). In Fig. 3.9, we work out the perturbations to the heat flux along the x direction, δQ_x, in response to the displaced component of the magnetic field, B_x, that is flux-frozen into the cluster plasma (as we will show in Section 3.3.2.5). The figure demonstrates that (i) regions with a positive displacement field, $\xi_z > 0$, experience a converging perturbation of the heat flux, δQ_x which implies heating. This causes the upwards displaced fluid elements to rise further, which reinforces the perturbation and causes an instability. (ii) Equivalently, regions with $\xi_z < 0$ experience cooling. They become denser, heavier and con-

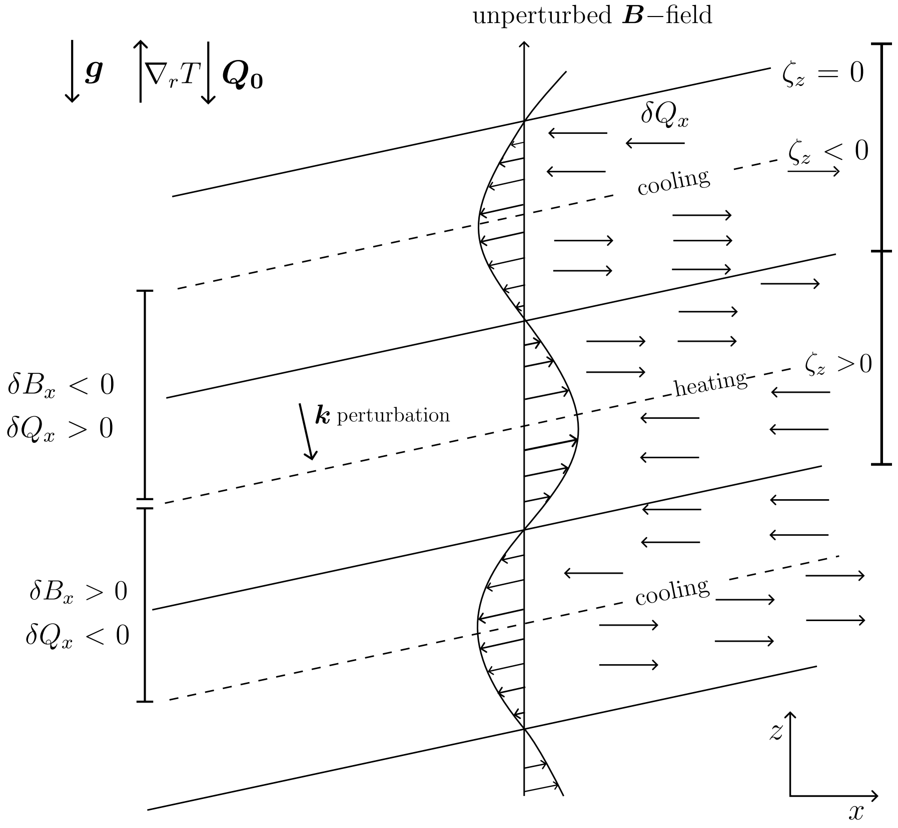


Figure 3.9: Sketch of the evolution of the HBI. The initial background heat flux $\boldsymbol{Q}_0$ experiences an oblique perturbation, which causes initially upwards perturbed regions ($\xi_z > 0$) to be furthermore heated. The fluid parcel continues to rise which reinforces the perturbation and causes instability.

tinue to sink in the gravitational potential which reinforces the perturbation and causes an instability.

Simulations of the non-linear stage of the instability demonstrate that the instability saturates quiescently with the magnetic field lines aligning horizontally (as suggested by the discussion of the linear regime of the instability), i.e., within the gravitational equipotential surfaces (shells of constant radius for a spherically symmetric cluster). This suppresses the inward heat flux by a large factor and thermally insulates the cooling core, which should reinforce the cooling catastrophe. Clearly, thermal conduction is not the solution to the cooling flow problem.

The HBI is very vulnerable to external turbulence: only 1% turbulent pressure support in comparison to the thermal pressure is sufficient to isotropize the magnetic field and to quench the instability.

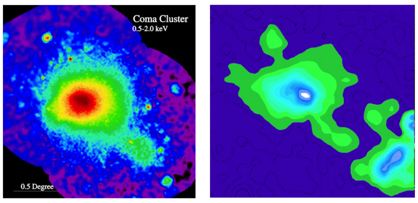
3.3 Non-thermal Processes

3.3.1 Non-thermal Radio Emission

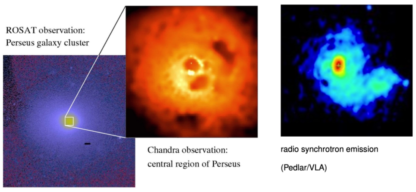
- The observation of diffuse radio synchrotron emission in form of *radio halos* and *radio relics* proves the existence of cluster-filling magnetic fields and relativistic electrons with a particle energy $E_\mathrm{e} \sim (1-30)$ GeV. This can be seen from the formula of the synchrotron frequency in the monochromatic approximation,

$$\nu_\mathrm{synch} = \frac{3eB}{2\pi\, m_\mathrm{e} c}\,\gamma^2 \simeq 1\,\mathrm{GHz}\,\frac{B}{\mu\mathrm{G}}\left(\frac{\gamma}{10^4}\right)^2, \qquad (3.252)$$

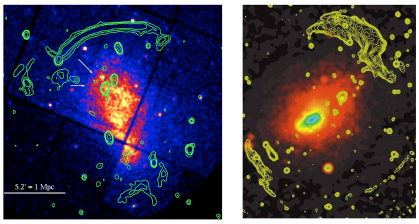
  where e denotes the elementary charge, m_e the electron mass, c is the speed of light, and $B = \sqrt{\langle \boldsymbol{B}^2\rangle}$ is the root-mean square of the magnetic vector field $\boldsymbol{B}$. The particle kinetic energy $E/(m_\mathrm{e}c^2) = \gamma - 1$ is defined in terms of the Lorentz factor γ.



The Coma galaxy cluster emits thermal X-rays (left) and has a giant radio halo (right).



The Perseus galaxy cluster emits thermal X-rays (left two panels) and has a radio mini-halo (right).



Two galaxy clusters caught after a major merger (left: CIZA J2242, right: Abell 3667). Shown are X-rays (colour) and radio emission of double relics (contours).

- *Giant radio halos* are centered on the cluster and similar in morphology to the thermal X-ray emission albeit with a patchier appearance. They occur in X-ray luminous clusters that have recently experienced a merger with another cluster. This correlation with X-ray luminous objects may either point to a selection effect or reveal intrinsic physics of their formation process. Although radio synchrotron emission is intrinsically polarized, the observed radio halo emission is unpolarized (within observational limits). This indicates that magnetic fields are volume filling or are threading a large fraction of the cluster volume and the observed emission is a superposition of causally uncorrelated radio-emitting patches.

- *Giant radio relics* are situated at cluster outskirts, correlate spatially with merger or accretion shocks, and have a high degree of polarization. This immediately points to a small emission volume and strengthens the hypothesis that radio relics are tracing freshly injected population of relativistic electrons, which has been accelerated at a shock and successively cooled via synchrotron emission in the post-shock region. To date, we know of more than 80 giant radio halos and relics.

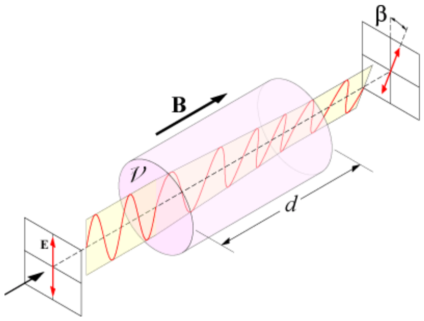
- *Faraday rotation measure.* When polarized radio emission propagates through a magnetized medium, its plane of polarization rotates for a non-zero line-of-sight component of the magnetic field $\boldsymbol{B}$ due to the birefringent property of the plasma – the so-call Faraday rotation effect. The Faraday rotation angle ϕ_obs of a linearly polarized electromagnetic wave (typically in the radio) is modified from its intrinsic position angle, $\phi_\mathrm{intrinsic}$. The rate of rotation scales with the wavelength squared and is given by

$$\phi_{\mathrm{obs}}(\boldsymbol{x}_\perp) = \lambda^2 \mathrm{RM}(\boldsymbol{x}_\perp) + \phi_{\mathrm{intrinsic}}(\boldsymbol{x}_\perp) \tag{3.253}$$

$$\mathrm{RM}(\boldsymbol{x}_\perp) = \frac{e^3}{2\pi\, m_{\mathrm{e}}^2 c^4} \int_0^L n_{\mathrm{e}}(\boldsymbol{x}_\perp, l)\, \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} \tag{3.254}$$

$$= 812\, \frac{\mathrm{rad}}{\mathrm{m}^2}\, \frac{B}{\mu\mathrm{G}}\, \frac{n_{\mathrm{e}}}{10^{-3}\mathrm{cm}^{-3}}\, \frac{L}{\mathrm{Mpc}}, \tag{3.255}$$

where n_{e} is the free electron number density and the integral extends along the line-of-sight from us to the emission site. Note that ϕ_{obs} is degenerate because of the $n\pi$ ambiguity of the observable polarization angle (i.e., rotations of this quantity by $n\pi$ are indistinguishable) so that we require a narrow frequency sampling of the radio observations.



Polarization rotation due to the Faraday effect.

- Faraday rotation measure (FRM) observations suggest central magnetic field strengths of 1–20 μG and values at the upper end of this range are indicative of cool core clusters. If the polarized radio wave passes through a turbulent magnetic field that has regions that point away and towards us, some of the FRM signal cancels out and only the only the signal of the "last turbulent cell" survives. In this case, it is advantageous to study the dispersion of the FRM signal that is sensitive to the root-mean square magnetic field strength even for cases with a vanishing mean FRM signal.

- Relativistic electrons (also called cosmic ray electrons) are plausibly accelerated at structure formation shocks. These shocks should be equally good (if not better) in accelerating cosmic ray protons that have much longer cooling times (of order the Hubble time) in comparison to the cosmic ray electrons. As a result, we expect cosmic ray protons to accumulate within galaxy clusters.

- The next two sections are meant to introduce the basic concepts of magnetic fields (their origin and transport) and cosmic rays (acceleration and transport) in a pedagogical manner and focuses on aspects important for galaxy clusters.

3.3.2 Magnetic Fields

3.3.2.1 Generating Magnetic Fields: Biermann Battery

- There are two different ways to generate cosmic magnetic fields. Either by means of a phase transition in the early universe that induces electric currents or astrophysically through the Biermann battery process. In the following, I will explain the second possibility that turns uncorrelated inhomogeneities in the electron density and temperature into magnetic fields. To see this, we start with Faraday's law:

$$\frac{\partial \boldsymbol{B}}{\partial t} = -c\boldsymbol{\nabla} \times \boldsymbol{E}. \tag{3.256}$$

- The momenta of electrons and protons change due to the Lorentz force (assuming for simplicity the non-relativistic approximation), the pressure and viscous forces:

$$m_\mathrm{e}\frac{\mathrm{d}\boldsymbol{v}_\mathrm{e}}{\mathrm{d}t} = -e\left(\boldsymbol{E} + \frac{\boldsymbol{v}_\mathrm{e}}{c} \times \boldsymbol{B} + \frac{1}{en_\mathrm{e}}\boldsymbol{\nabla} P_\mathrm{e}\right) - \frac{\nu m_\mathrm{e}}{n_\mathrm{e}}(\boldsymbol{v}_\mathrm{e} - \boldsymbol{v}_\mathrm{p}), \tag{3.257}$$

$$m_\mathrm{p}\frac{\mathrm{d}\boldsymbol{v}_\mathrm{p}}{\mathrm{d}t} = e\left(\boldsymbol{E} + \frac{\boldsymbol{v}_\mathrm{p}}{c} \times \boldsymbol{B} + \frac{1}{en_\mathrm{p}}\boldsymbol{\nabla} P_\mathrm{p}\right). \tag{3.258}$$

  If protons and electrons share the same temperature, we can neglect the second equation in comparison to the first equation because protons move on average slower than electrons by a factor $\sqrt{m_\mathrm{p}/m_\mathrm{e}}$ and can be considered to be almost at rest for the electrons. Moreover, viscous forces (i.e., the last term in Eq. 3.257) are very small on large scales, i.e., scales larger than the Kolmogorov length. On scales of the particle mean free path, we must consider plasma effects, which complicates the picture considerably.

- Because we are interested in growing a magnetic field on large (fluid) scales, we can assume a steady state on small (plasma) time and length scales. This steady state averages over the small-scale plasma motions and is valid on time scales $\tau > \omega_\mathrm{pl}^{-1}$, where $\omega_\mathrm{pl}^2 = 4\pi n_\mathrm{e} e^2/m_\mathrm{e}$ is the plasma frequency (i.e., the fundamental eigenmode of a plasma). By no means does this imply a steady state on (large) fluid scales and on the contrary, it is the aim of this derivation to arrive at an equation that described the change of the magnetic field on fluid scales.

- Solving Eq. (3.257) for the electric field, we obtain

$$\boldsymbol{E} = -\frac{\boldsymbol{v}_\mathrm{e} \times \boldsymbol{B}}{c} - \frac{\boldsymbol{\nabla} P_\mathrm{e}}{en_\mathrm{e}}. \tag{3.259}$$

  Multiplying this equation by $-c$, taking the curl of it and using Faraday's law (Eq. 3.256), we obtain

$$\frac{\partial \boldsymbol{B}}{\partial t} = \boldsymbol{\nabla} \times (\boldsymbol{v}_\mathrm{e} \times \boldsymbol{B}) + \frac{c}{e}\boldsymbol{\nabla} \times \left(\frac{\boldsymbol{\nabla} P_\mathrm{e}}{n_\mathrm{e}}\right). \tag{3.260}$$

  Note that the time derivative considers changes of the magnetic field on larger (fluid) time and length scales.

- Using $P_\mathrm{e} = n_\mathrm{e} k_\mathrm{B} T_\mathrm{e}$ and the identity $\boldsymbol{\nabla} \times (f \boldsymbol{\nabla} g) \equiv \boldsymbol{\nabla} f \times \boldsymbol{\nabla} g$, we can rewrite the second term of Eq. (3.260),

$$\begin{aligned}\frac{1}{k_\mathrm{B}} \boldsymbol{\nabla} \times \left(\frac{\boldsymbol{\nabla} P_\mathrm{e}}{n_\mathrm{e}}\right) &= \boldsymbol{\nabla} \times \left[\frac{1}{n_\mathrm{e}} \boldsymbol{\nabla}(n_\mathrm{e} T_\mathrm{e})\right] = \boldsymbol{\nabla} \times (\boldsymbol{\nabla} T_\mathrm{e}) + \boldsymbol{\nabla} \times \left(\frac{T_\mathrm{e}}{n_\mathrm{e}} \boldsymbol{\nabla} n_\mathrm{e}\right) \\ &= \boldsymbol{\nabla}\left(\frac{T_\mathrm{e}}{n_\mathrm{e}}\right) \times \boldsymbol{\nabla} n_\mathrm{e} = \frac{1}{n_\mathrm{e}} \boldsymbol{\nabla} T_\mathrm{e} \times \boldsymbol{\nabla} n_\mathrm{e} - \frac{T_\mathrm{e}}{n_\mathrm{e}^2} \boldsymbol{\nabla} n_\mathrm{e} \times \boldsymbol{\nabla} n_\mathrm{e} \\ &= \frac{1}{n_\mathrm{e}} \boldsymbol{\nabla} T_\mathrm{e} \times \boldsymbol{\nabla} n_\mathrm{e}. \end{aligned} \tag{3.261}$$

- Hence, we obtain the Biermann battery equation,

$$\frac{\partial \boldsymbol{B}}{\partial t} = \boldsymbol{\nabla} \times (\boldsymbol{v}_\mathrm{e} \times \boldsymbol{B}) - \frac{c k_\mathrm{B}}{e n_\mathrm{e}} \boldsymbol{\nabla} n_\mathrm{e} \times \boldsymbol{\nabla} T_\mathrm{e}. \tag{3.262}$$

  This equations shows that if there is no magnetic field to start with (i.e., a vanishing first term on the right-hand side), then the magnetic field can be generated by a baroclinic flow with $\boldsymbol{\nabla} n_\mathrm{e} \times \boldsymbol{\nabla} T_\mathrm{e} \neq \boldsymbol{0}$. This could be achieved in shocks of the interstellar medium, in ionization fronts, or similar astrophysical sites. In general, baroclinic flows are sourced by rotational motions at shocks of finite extent such as the chaotic collapse of a proto-galaxy.

- Typically, magnetic fields generated through this process have very small field strengths as we can easily verify to order of magnitude. Adopting a characteristic density and temperature gradient length of L of a proto-galaxy and assuming gravitational collapse on the free-fall time, $\tau \sim 1/\sqrt{G\rho}$, we obtain

$$\begin{aligned} B &\sim \frac{c k_\mathrm{B} T_\mathrm{e}}{e} \frac{\tau}{L^2} \sim \frac{c k_\mathrm{B} T_\mathrm{e}}{e} \frac{1}{\sqrt{G\rho} L^2} \\ &\sim 10^{-20}\mathrm{G} \left(\frac{T_\mathrm{e}}{10^4\,\mathrm{K}}\right) \left(\frac{n}{1\,\mathrm{cm}^{-3}}\right)^{-1/2} \left(\frac{L}{\mathrm{kpc}}\right)^{-2}. \end{aligned} \tag{3.263}$$

- Naively, going to smaller length scales L should increase B. But in order to explain the coherence on scales of several kpcs, we would have to evoke a process such as a small-scale wind that moves the magnetic fields back to kpc scales and in that process we would have to account for adiabatic losses that necessarily accompany the expansion from small to large scales (see Section 3.3.2.5) – in the end we would gain nothing from running a Biermann battery on smaller scales.

- Hence, the big challenge consists in growing coherent large-scale magnetic fields from the small-scale field of small amplitude. This is a major challenge of dynamo theory. Suggestions include (i) magnetic diffusion of small-scale field that fills the volume of the interstellar medium, (ii) shearing and stretching of magnetic field in shocks, (iii) an expulsion of magnetic fields by galactic winds and further dynamo amplification in the ICM.

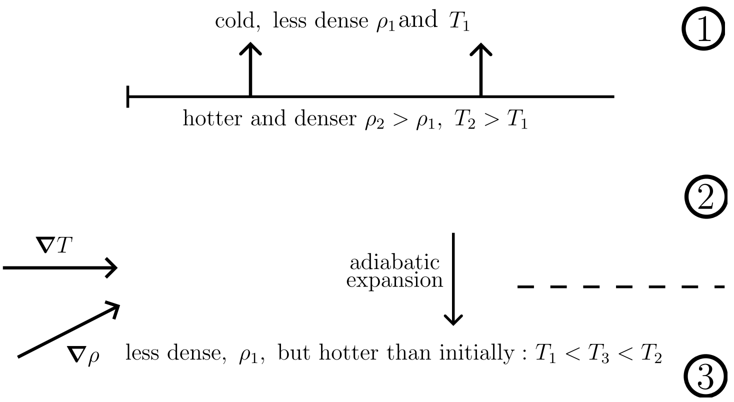


Figure 3.10: A shock of finite extent propagates into zero pressure. Shocks of limited spatial extent break the barotropic relation $P_\mathrm{e} = P_\mathrm{e}(n_\mathrm{e})$ which couples P_e to n_e.

3.3.2.2 Evolution of Magnetic Fields: the Induction Equation

- We would like to understand how magnetic fields modify the dynamics of a fluid, i.e. we derive the equations of magnetohydrodynamics. To this end, we need (i) an equation for the evolution of magnetic fields, i.e., the induction equation, and (ii) work out the magnetic forces and stress. Clearly, we need to make a number of simplifying assumptions. The momentum of a plasma, i.e. an ionized but overall charge-neutral fluid, is dominated by the ions. Hence, we assume a negligible electron mass so that they can move sufficiently fast and short-circuiting any electric fields on time scales much shorter than those of relevance for the fluid dynamics. Hence, we can neglect the displacement current, $\partial \boldsymbol{E}/\partial t = \boldsymbol{0}$.

- We are considering the comoving frame, where all electric fields can be expressed in terms of currents and the (magnetic) Lorentz force. We start with Ohm's Law:

$$\boldsymbol{E} = \eta \boldsymbol{j} - \frac{\boldsymbol{v}}{c} \times \boldsymbol{B}, \tag{3.264}$$

  where η is the resistivity (which is measured in units of time) and $\boldsymbol{v}$ is the fluid velocity.

- Using Faraday's Law, $\dfrac{\partial \boldsymbol{B}}{\partial t} = -c\, \boldsymbol{\nabla} \times \boldsymbol{E}$, we get

$$\frac{\partial \boldsymbol{B}}{\partial t} = \boldsymbol{\nabla} \times (\boldsymbol{v} \times \boldsymbol{B}) - \boldsymbol{\nabla} \times (c\, \eta \boldsymbol{j}). \tag{3.265}$$

We substitute the electric current $\boldsymbol{j}$ with Ampère's Law (where we neglect the displacement current, see above), $\boldsymbol{\nabla}\times\boldsymbol{B} = 4\pi\boldsymbol{j}/c$ and get

$$\frac{\partial\boldsymbol{B}}{\partial t} = \boldsymbol{\nabla}\times(\boldsymbol{v}\times\boldsymbol{B}) - \frac{c^2\eta}{4\pi}\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{B}), \tag{3.266}$$

where we assumed $\eta = \text{const}$.

- To proceed, we prove the following identity:

$$\begin{aligned}(\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{B}))|_i &= \varepsilon_{ijk}\partial_j\varepsilon_{klm}\partial_l B_m = (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl})\partial_j\partial_l B_m \\ &= \partial_i\partial_m B_m - \partial_l\partial_l B_i = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{B}) - \boldsymbol{\nabla}^2\boldsymbol{B}\big|_i, \end{aligned} \tag{3.267}$$

  where we have used Einstein's sum convention.

- Using the solenoidal condition, $\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$, we arrive at the *induction equation*:

$$\frac{\partial\boldsymbol{B}}{\partial t} = \boldsymbol{\nabla}\times(\boldsymbol{v}\times\boldsymbol{B}) + D\boldsymbol{\nabla}^2\boldsymbol{B}, \quad \text{where} \quad D = \frac{c^2\eta}{4\pi}. \tag{3.268}$$

  The first term is the "convective term", which states that the field is frozen into the flow (as we will see momentarily). This is an important property for astrophysical plasmas. The second term is the "diffusive term" and represents the diffusive leakage of magnetic field lines across the conducting field, which is important for changing the magnetic topology, e.g. in magnetic reconnection. The "diffusive term" can be neglected for infinite conductivity $\sigma = \eta^{-1}$ or for large magnetic Reynolds numbers $\mathrm{Re_m} \to \infty$:

$$\mathrm{Re_m} = \frac{|\text{convective term}|}{|\text{diffusive term}|} = \frac{L^{-1}v\,B}{D\,L^{-2}B} = \frac{L\,v}{D} \tag{3.269}$$

- It is no coincidence that the convective term of the induction equation (3.268) resembles the convective term of the Biermann equation (3.262). In fact, this is a necessary consequence of our assumption that electrons can be described by a nearly massless, charge-neutralizing fluid that moves alongside the ions. However, the logic here is different because the induction equation describes the evolution of the magnetic field under the action of the ion inertia of the fluid while Biermann's equation describes the generation and evolution of the magnetic field that is supported by the electron-proton fluid, and specifically relies on the faster electron fluid dynamics.

3.3.2.3 Magnetic Pressure and Curvature Force

- The Lorentz force is directed perpendicular to $\boldsymbol{B}$. Along the magnetic field, the fluid motion is therefore only subject to the hydrodynamic forces, which makes the dynamics of a magnetized fluid extremely anisotropic. To understand how exactly the Lorentz force acts in a hydrodynamic fluid, we write the force in terms of $\boldsymbol{B}$. Using Ampère's law at low frequencies, $\nabla \times \boldsymbol{B} = 4\pi\boldsymbol{j}/c$, we can show that the Lorentz force density can be written as follows:

$$\boldsymbol{f}_{\mathrm{L}} = \frac{1}{c}\boldsymbol{j} \times \boldsymbol{B} = \frac{1}{4\pi}(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} = \frac{1}{4\pi}(\boldsymbol{B}\cdot\nabla)\boldsymbol{B} - \frac{1}{8\pi}\nabla B^2, \tag{3.270}$$

where the two terms on the right-hand side are often (erroneously) attributed to the magnetic curvature and pressure forces, respectively. The third equality derives from the following identity:

$$\begin{aligned}(\nabla \times \boldsymbol{B}) \times \boldsymbol{B}|_i &= \varepsilon_{ijk}\varepsilon_{jlm}(\partial_l B_m)B_k = \varepsilon_{kij}\varepsilon_{jlm}(\partial_l B_m)B_k \\ &= (\delta_{kl}\delta_{im} - \delta_{km}\delta_{il})(\partial_l B_m)B_k = (\partial_k B_i)B_k - (\partial_i B_k)B_k \\ &= (\boldsymbol{B}\cdot\nabla)\boldsymbol{B} - \frac{1}{2}\nabla\boldsymbol{B}^2\bigg|_i .\end{aligned} \tag{3.271}$$

- In order to fully separate the effects of magnetic curvature and pressure we write $\boldsymbol{B} = B\boldsymbol{b}$, where $\boldsymbol{b}$ is the unit vector in the direction of $\boldsymbol{B}$ and obtain

$$\begin{aligned}\boldsymbol{f}_{\mathrm{L}} &= \frac{B^2}{4\pi}(\boldsymbol{b}\cdot\nabla)\boldsymbol{b} + \frac{1}{8\pi}\boldsymbol{b}(\boldsymbol{b}\cdot\nabla)B^2 - \frac{1}{8\pi}\nabla B^2 \\ &= \frac{B^2}{4\pi}(\boldsymbol{b}\cdot\nabla)\boldsymbol{b} - \frac{1}{8\pi}\nabla_\perp B^2 \equiv \boldsymbol{f}_{\mathrm{c}} + \boldsymbol{f}_{\mathrm{p}},\end{aligned} \tag{3.272}$$

where we define the gradient perpendicular to the magnetic field lines, $\nabla_\perp = (\mathbf{1} - \boldsymbol{bb})\cdot\nabla$. The second term, $\boldsymbol{f}_{\mathrm{p}}$, acts like a *pressure force perpendicular to the magnetic field lines* and the first term, $\boldsymbol{f}_{\mathrm{c}}$, is the *magnetic curvature force* that also acts in a plane orthogonal to the field line.

- To see this, we locally identify a curved field line with its curvature circle so that we can locally define an azimuthally directed field $\boldsymbol{B} = B\boldsymbol{e}_\varphi$ in cylindrical coordinates (R, φ, z). Hence, in this case we obtain

$$(\boldsymbol{b}\cdot\nabla)\boldsymbol{b} = (\boldsymbol{e}_\varphi\cdot\nabla)\boldsymbol{e}_\varphi = -\frac{\boldsymbol{e}_R}{R} \tag{3.273}$$

so that the curvature force always points towards the center of the curvature circle and aims to reduce the curvature by pulling the field line straight with a force that is the greater the smaller the curvature radius is.

- Hence, it is advisable to define a magnetic curvature via

$$\boldsymbol{\kappa} \equiv (\boldsymbol{b}\cdot\nabla)\boldsymbol{b} = \frac{(\mathbf{1} - \boldsymbol{bb})\cdot(\boldsymbol{B}\cdot\nabla)\boldsymbol{B}}{B^2} = \frac{4\pi\boldsymbol{f}_{\mathrm{c}}}{B^2}, \tag{3.274}$$

which immediately defines the curvature radius via

$$R_{\mathrm{c}} \equiv \frac{1}{\kappa} = \frac{1}{|\boldsymbol{\kappa}|} = \frac{1}{|(\boldsymbol{b}\cdot\nabla)\boldsymbol{b}|}. \tag{3.275}$$

- Equation (3.274) may seem to suggest that large curvature forces $f_\mathrm{c} = |\boldsymbol{f}_\mathrm{c}|$ and small magnetic field strengths imply a large magnetic curvature. However, idealized simulations of incompressible, driven turbulence show that in the regime of high curvature, κ is not strongly correlated with a large curvature force but instead with a small value of the magnetic field strength while small magnetic curvature is correlated with a low level of curvature force. Hence, the anticorrelation between B^2 and κ is a consequence of the curvature force normal to the magnetic field line. A large curvature force rapidly straightens out any curved field line, which apparently precludes the possibility of a joint presence of a high curvature and a large magnetic field. Heuristically, this means that a strong magnetic field resists bending.
- Most importantly, a magnetized plasma that does not experience any driving will evolve into a state that minimizes magnetic tension and curvature. Hence, the continued presence of magnetic curvature requires an active process such as a small-scale dynamo to build up and maintain a high level of magnetic curvature.

3.3.2.4 Magnetic Stress Tensor

- As shown in Eq. (3.270), the Lorentz force density can be written as follows:

$$\boldsymbol{f}_\mathrm{L} = \frac{1}{c}\boldsymbol{j}\times\boldsymbol{B} = \frac{1}{4\pi}(\boldsymbol{B}\cdot\boldsymbol{\nabla})\boldsymbol{B} - \frac{1}{8\pi}\boldsymbol{\nabla}B^2 = -\boldsymbol{\nabla}\cdot\mathsf{M}, \qquad (3.276)$$

where

$$\mathsf{M}_{ij} = -\frac{1}{4\pi}B_iB_j + \frac{1}{8\pi}B^2\delta_{ij} \qquad (3.277)$$

is the magnetic stress tensor: it plays a role analogous to the fluid pressure in ordinary fluid mechanics (explaining the minus sign introduced in its definition). The last equality makes use of the solenoidal condition, $\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$, so that we obtain:

$$\frac{1}{4\pi}\left[(\boldsymbol{B}\cdot\boldsymbol{\nabla})\boldsymbol{B} - \frac{1}{2}\boldsymbol{\nabla}\boldsymbol{B}^2\right]_i = \frac{1}{4\pi}\partial_k\left(B_iB_k - \frac{1}{2}B^2\delta_{ik}\right) = -\partial_k\mathsf{M}_{ik}. \qquad (3.278)$$

- To get a better understanding of the meaning of these terms, we work out the surface force (per unit area) exerted by a bounded volume V *on its surroundings*. The net Lorentz force acting *on a volume* V of fluid can be written as an integral of a magnetic stress vector acting on its surface,

$$\begin{aligned}\int_V \boldsymbol{f}_\mathrm{L}\mathrm{d}V &= \int_V \frac{1}{4\pi}(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}\,\mathrm{d}V \\ &= -\int_V \boldsymbol{\nabla}\cdot\mathsf{M}\,\mathrm{d}V = -\oint_S \boldsymbol{n}\cdot\mathsf{M}\,\mathrm{d}S. \qquad (3.279)\end{aligned}$$

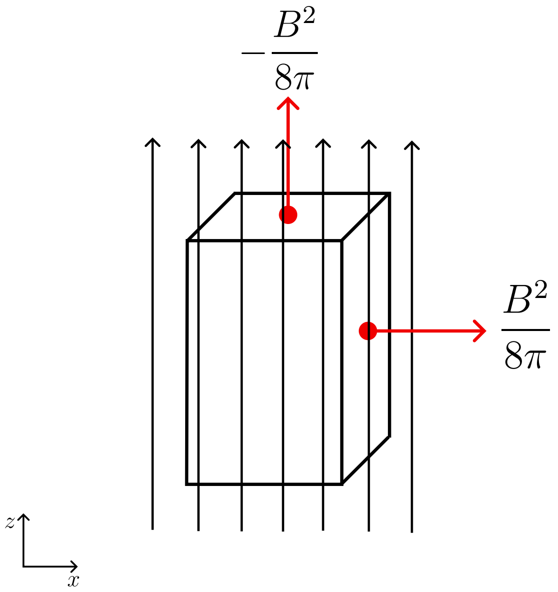


Figure 3.11: Sketch that visualizes magnetic pressure force and magnetic stress.

- To get the force density $\boldsymbol{f}_S$ exerted by the volume *on its surroundings*, we need to add a minus sign to the last term,

$$\boldsymbol{f}_S = \boldsymbol{n} \cdot \mathsf{M} = -\frac{1}{4\pi}\boldsymbol{B}B_n + \frac{1}{8\pi}B^2\boldsymbol{n}, \tag{3.280}$$

  where $B_n = \boldsymbol{B} \cdot \boldsymbol{n}$ is the component of $\boldsymbol{B}$ along the outward normal $\boldsymbol{n}$ to the surface of the volume.

- To visualize the meaning of magnetic pressure and tension, we take a uniform magnetic field along the z-direction ($\boldsymbol{B} = B\boldsymbol{e}_z$) and compute the surface forces $\boldsymbol{f}_S$ exerted by a rectangular volume that is aligned with the magnetic field. There are 6 different surface elements but symmetry limits the surface forces to two different types: 4 with a normal perpendicular to the magnetic field and 2 with a normal that is (anti-)parallel to the magnetic field. In particular, we ask which magnetic forces (pressure or tension) are contributing to these surface forces.

- Take the *surface* perpendicular to the x axis on the right-hand side of the box, so that we have $\boldsymbol{n} = \boldsymbol{e}_x$:

$$\boldsymbol{f}_{\text{right}} = \boldsymbol{e}_x \cdot \mathsf{M}, \tag{3.281}$$

$$f_{\text{right},x} = -\frac{1}{4\pi}B_xB_z + \frac{1}{8\pi}B^2 = \frac{1}{8\pi}B^2, \quad f_{\text{right},y} = f_{\text{right},z} = 0. \tag{3.282}$$

  Hence, a magnetic field in a conducting fluid is always under compression in two directions (perpendicular to the field) and causes the fluid volume to expand in the perpendicular directions to the magnetic field if there is not (other) hydrodynamic force balancing it.

- The stress exerted by the magnetic field at the top of the surface element (with $\boldsymbol{n} = \boldsymbol{e}_z$) is

$$\boldsymbol{f}_{\text{top}} = \boldsymbol{e}_z \cdot \mathsf{M}, \qquad \text{note that we have here: } B^2 = B_z^2, \tag{3.283}$$

$$f_{\text{top},z} = -\frac{1}{4\pi} B_z B_z + \frac{1}{8\pi} B^2 = -\frac{1}{8\pi} B^2, \quad f_{\text{top},x} = f_{\text{top},y} = 0. \tag{3.284}$$

  The stress is also perpendicular to the surface and of equal magnitude to that of the magnetic pressure exerted at the vertical surfaces, but of opposite sign.

- Because there are no magnetic monopoles (i.e., $\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$), magnetic field lines have no ‘ends’ and cannot be separated. Hence, the contraction along the magnetic field under magnetic stress does not happen in practice because the tension at its top and bottom surfaces is exactly balanced by the tension in the magnetic lines continuing above and below the box. As we discussed in Eq. (3.273), the effects of tension in a magnetic field manifest themselves more indirectly, through the curvature of field lines.

3.3.2.5 Magneto-hydrodynamics and Flux-freezing

- For a collisional fluid on scales larger than the particle mean-free path and on time scales longer than the inverse plasma frequency, $\tau > \omega_{\text{pl}}^{-1}$, the evolution of the hydrodynamic quanities (density, pressure, velocity, and magnetic vector field) is given by magneto-hydrodynamics (MHD). Ideal MHD assumes an inviscid (i.e., no viscosity), ideally conducting fluid (i.e., zero resistivity $\eta = 0$). To derive the MHD equations, we add the Lorentz force density to the momentum evolution equation (the Euler equation) and supplement the system of conservation equations of mass, momentum and entropy by the equation for magnetic induction, Eq. (3.268) without the diffusion term.

- This yields the equations of ideal MHD:

$$\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\boldsymbol{v}) = 0, \tag{3.285}$$

$$\begin{aligned}\rho\left(\frac{\partial\boldsymbol{v}}{\partial t} + \boldsymbol{v}\cdot\boldsymbol{\nabla}\boldsymbol{v}\right) &= -\boldsymbol{\nabla}P + \boldsymbol{j}\times\boldsymbol{B} \\ &= -\boldsymbol{\nabla}\cdot\left[\left(P + \frac{\boldsymbol{B}^2}{8\pi}\right)\bar{\mathbf{1}} - \frac{1}{4\pi}\boldsymbol{B}\boldsymbol{B}\right],\end{aligned} \tag{3.286}$$

$$\frac{\partial s}{\partial t} + \boldsymbol{v}\cdot\boldsymbol{\nabla}s = 0, \tag{3.287}$$

$$\frac{\partial\boldsymbol{B}}{\partial t} - \boldsymbol{\nabla}\times(\boldsymbol{v}\times\boldsymbol{B}) = \mathbf{0}, \quad \text{subject to the constraint } \boldsymbol{\nabla}\cdot\boldsymbol{B} = 0, \tag{3.288}$$

where $\rho = \rho(\boldsymbol{x})$, $P = P(\boldsymbol{x})$, $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x})$, $\boldsymbol{j} = \boldsymbol{j}(\boldsymbol{x})$, $s = s(\boldsymbol{x})$, and $\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{x})$ are the density, pressure, velocity, electric current, entropy, and magnetic fields. In the momentum equation (3.286), we expanded the Lorentz force in terms of magnetic pressure and tension forces, i.e. using Eqs. (3.276) and (3.277).

- This set of equations is non-linear in $\boldsymbol{B}$, which can be seen by combining Eqs. (3.286) and (3.288), and in general has to be solved numerically. Nevertheless, we will explore one very important property of magnetic fields and show that their flux is “frozen” into the plasma. Starting with the induction equation (3.288)

$$\frac{\partial\boldsymbol{B}}{\partial t} = \boldsymbol{\nabla}\times(\boldsymbol{v}\times\boldsymbol{B}), \tag{3.289}$$

and employing the property, that magnetic fields do not have sources nor sinks, $\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$, we obtain

$$\frac{\partial\boldsymbol{B}}{\partial t} = (\boldsymbol{B}\cdot\boldsymbol{\nabla})\boldsymbol{v} - (\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{B} - (\boldsymbol{\nabla}\cdot\boldsymbol{v})\boldsymbol{B}, \tag{3.290}$$

which can be rearranged to yield

$$\frac{\mathrm{d}\boldsymbol{B}}{\mathrm{d}t} \equiv \frac{\partial\boldsymbol{B}}{\partial t} + (\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{B} = (\boldsymbol{B}\cdot\boldsymbol{\nabla})\boldsymbol{v} - (\boldsymbol{\nabla}\cdot\boldsymbol{v})\boldsymbol{B}. \tag{3.291}$$

- With the continuity equation,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{\partial\rho}{\partial t} + (\boldsymbol{v}\cdot\boldsymbol{\nabla})\rho = -(\boldsymbol{\nabla}\cdot\boldsymbol{v})\rho, \tag{3.292}$$

we obtain the following equation,

$$\frac{\mathrm{d}\boldsymbol{B}}{\mathrm{d}t} \equiv \frac{\partial\boldsymbol{B}}{\partial t} + (\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{B} = (\boldsymbol{B}\cdot\boldsymbol{\nabla})\boldsymbol{v} - \frac{\boldsymbol{B}}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t}. \tag{3.293}$$

Multiplying Eq. (3.293) by ρ^{-1} and rearranging terms yields

$$\frac{1}{\rho}\frac{\mathrm{d}\boldsymbol{B}}{\mathrm{d}t} - \frac{\boldsymbol{B}}{\rho^2}\frac{\mathrm{d}\rho}{\mathrm{d}t} = \left(\frac{\boldsymbol{B}}{\rho}\cdot\boldsymbol{\nabla}\right)\boldsymbol{v}, \tag{3.294}$$

or after combining the terms on the left-hand side, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\boldsymbol{B}}{\rho}\right) = \left(\frac{\boldsymbol{B}}{\rho}\cdot\boldsymbol{\nabla}\right)\boldsymbol{v}. \tag{3.295}$$

This is the flux-freezing equation of magnetic fields.

- To see this explicitly, consider the evolution of two neighboring points in the fluid that are connected by the vector $\delta\boldsymbol{x}$:

$$\Delta\boldsymbol{x}(t) = \delta\boldsymbol{x}, \tag{3.296}$$

$$\Delta\boldsymbol{x}(t+\Delta t) = \delta\boldsymbol{x} + (\delta\boldsymbol{x}\cdot\boldsymbol{\nabla})\boldsymbol{v}\Delta t + \mathcal{O}(t^2), \tag{3.297}$$

$$\frac{\mathrm{d}\delta\boldsymbol{x}}{\mathrm{d}t} \equiv \frac{\Delta\boldsymbol{x}(t+\Delta t) - \Delta\boldsymbol{x}(t)}{\Delta t} = (\delta\boldsymbol{x}\cdot\boldsymbol{\nabla})\boldsymbol{v}. \tag{3.298}$$

We see that $\boldsymbol{B}/\rho$ and $\delta\boldsymbol{x}$ satisfy the same ordinary differential equation. Hence, if initially $\delta\boldsymbol{x} = \epsilon\boldsymbol{B}/\rho$ ($\epsilon \ll 1$), the same relation will hold for all times. If $\delta\boldsymbol{x}$ connects two particles on the same field line, they remain on the same field line for all times (neglecting magnetic diffusion effects that would constitute non-ideal MHD).

- We can work out the consequences of this flux-freezing condition by considering a uniform contraction of the plasma towards the origin. The plasma resides in a sphere of radius r and conserves mass and magnetic flux $\mathrm{d}\Phi = \boldsymbol{B}\cdot\mathrm{d}\boldsymbol{A}$ (where $\mathrm{d}\boldsymbol{A}$ is the surface element on the sphere). Thus, both ρr^3 and Br^2 are constant and we obtain

$$B \equiv \sqrt{\langle\boldsymbol{B}^2\rangle} \propto \rho^{\alpha_B}, \quad \alpha_B = \frac{2}{3}, \tag{3.299}$$

for isotropic contraction or expansion, independent of the magnetic topology. Thus, flux freezing alone predicts a tight relation between B and ρ. Note that the scaling exponent α_B depends on the type of symmetry invoked during collapse (whether it is isotropic or not) and can differ for contraction along a homogeneous magnetic field ($\alpha_B = 0$) or perpendicular to it ($\alpha_B = 1$).

3.3.2.6 Magneto-hydrodynamic Waves and Turbulence

- This topics is generically subject to plasma physics. Here, we will provide the basic picture while leaving the rigorous proofs for the *plasma physics* lectures.

- Linearizing the MHD equations (3.285) through (3.288) yields eight equations for eight unknowns: the perturbations in density, pressure, velocity, and magnetic field, $\delta\rho$, δP, $\delta\boldsymbol{v}$, and $\delta\boldsymbol{B}$. These are subject to the constraint $\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$, which reduces the dimensionality by one. To proceed, we carry out a Fourier analysis of the perturbed quantities, e.g.,

$$\delta P(\boldsymbol{x}, t) = \int \delta\tilde{P}(\boldsymbol{k}, t)\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x})\mathrm{d}^3k. \tag{3.300}$$

Hence, we have seven different wave modes in a magnetized plasma that propagate forward in time (two more than in an unmagnetized plasma):

1. An Alfvén wave is a type of plasma wave that arises due to oscillations of background ions in response to a restoring force provided by the tension of magnetic field lines. The ion mass density provides the inertia for the resulting electro-magnetic wave and causes it to oscillate much slower than a light wave, provided the plasma is non-relativistic. As such, Alfvén waves are low-frequency (compared to the ion gyrofrequency) oscillations and the perturbations of the magnetic field are transverse to the direction of propagation, which makes them incompressible. In astrophysical plasmas, Alfvén waves propagate alongside or opposite to the direction of the mean magnetic field at the group velocity which is called Alfvén velocity,

$$\boldsymbol{v}_{\mathrm{gr}} = \frac{\partial\omega}{\partial\boldsymbol{k}} = \frac{\boldsymbol{B}_0}{\sqrt{4\pi\rho}} \equiv \boldsymbol{v}_{\mathrm{A}}. \tag{3.301}$$

using the Gaussian system of units and ρ is the ion mass density of the plasma. There are two polarization states of *Alfvén modes*. In a cluster, we typically have

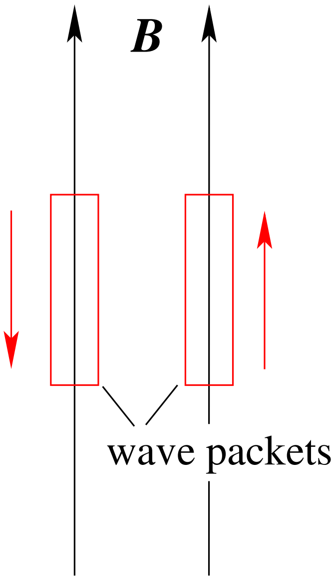
$$v_{\mathrm{A}} \sim 90\,\frac{\mathrm{km}}{\mathrm{s}} \times \left(\frac{B}{\mu\mathrm{G}}\right)\left(\frac{n}{10^{-3}\mathrm{cm}^{-3}}\right)^{-1/2}. \tag{3.302}$$

2. There are two polarization states of compressible *fast magnetosonic modes* which are almost equivalent to sound waves in the ICM with the plasma-β parameter

$$\beta = \frac{P}{P_B} = \frac{nk_{\mathrm{B}}T}{B^2/(8\pi)} = \frac{2nk_{\mathrm{B}}T}{B^2/(4\pi\rho)\,\mu m_{\mathrm{p}}n} = \frac{2}{\gamma}\frac{c_{\mathrm{s}}^2}{v_{\mathrm{A}}^2} \gtrsim 100, \tag{3.303}$$

where $\gamma = 5/3$ is the adiabatic index and $c_{\mathrm{s}} = \sqrt{\gamma P/\rho}$ is the adiabatic sound speed. In this case, the restoring force of those (mostly) longitudinal *fast magnetosonic waves* is the pressure gradient force. Fast magnetosonic modes do not interact with Alfvén modes.

3. There are also two polarization states of compressible *slow magnetosonic modes* which are nearly degenerate with Alfvén modes in a high-β plasma.
4. Finally, there is the *entropy mode*, which is a compressible zero-frequency mode whose density fluctuations are offset by temperature fluctuations.



Interacting Alfvén wave packets.

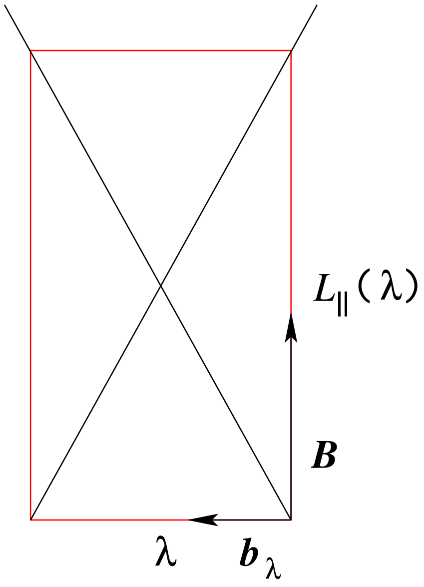
- We now consider (spatially localized) Alfvén wave packets that transport energy and momentum along $\boldsymbol{B}$ fields. MHD turbulence deals with the physics of interacting wave packets. Alfvénic turbulence is incompressible:

$$\frac{\delta v_{\mathrm{A}}}{v_{\mathrm{A}}} = \frac{\delta B}{B} \equiv \frac{b_\lambda}{B}, \tag{3.304}$$

where b_λ denotes the transverse magnetic perturbation of an Alfvén wave.

- What happens when two Alfvén wave packets are interacting? The down-going packet causes field line wandering such that the upward going packet is broken apart after a distance $L_\|(\lambda)$. This leads to the “critical balance condition” of Alfvénic turbulence (Goldreich & Shridhar 95, 97, see the drawing to the right)

$$L_\| = \frac{\lambda B}{b_\lambda}. \tag{3.305}$$



Geometrical interpretation of the “critical balance” condition.

- Perpendicular to the mean magnetic field, we have Kolmogorov turbulence. As we have seen in Section 3.1.5, the energy flux of the fluctuating field at scale λ is constant, $\dot{\epsilon} = b_\lambda^2/t_\lambda = \mathrm{const.}$ so that we get

$$t_\lambda = \frac{L_\|}{v_{\mathrm{A}}} = \frac{\lambda\, B}{v_{\mathrm{A}}\, b_\lambda} \propto b_\lambda^2, \quad \text{and} \quad B \propto v_{\mathrm{A}} = \mathrm{const.} \tag{3.306}$$

Hence, we obtain the scaling of Alfvénic turbulence:

$$b_\lambda \propto \lambda^{1/3}, \quad \text{or} \quad L_\| \propto \lambda^{2/3}\, L_{\mathrm{MHD}}^{1/3}. \tag{3.307}$$

This shows that the smaller the scale λ, the more anisotropic is the turbulent scaling and the more elongated are the eddies ($L_\|/\lambda \propto \lambda^{1/3}$). The long axis of the spaghetti-like eddies is aligned with the local mean field, $\langle \boldsymbol{B} \rangle$.

3.3.3 Cosmic Rays

3.3.3.1 General Properties and Cosmic Ray-Wave Interactions

- In our Galaxy there exist non-thermal ion and electron populations (called cosmic rays or CRs for short). The CR ions obey

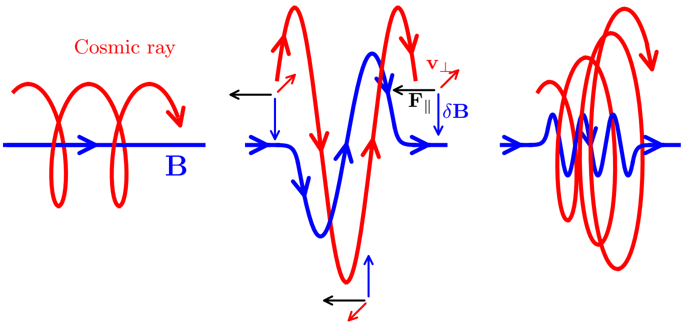


Figure 3.12: Illustration of a CR (red) that gyrates around an unperturbed magnetic field line (left) and interacts with a single transverse magnetic field fluctuation, i.e., an Alfvén wave (middle). Lorentz forces due to the fluctuation change the parallel velocity component of the particle. Because there are no electric fields in the Alfvén wave frame, energy is conserved and the perpendicular velocity has to increase, thereby increasing the CR's pitch angle (angle between CR momentum and $\boldsymbol{B}$). This happens as a result of collective effects of several perturbations (right) that eventually causes the CR to be reflected.

a power-law distribution with more than 33 decades in flux and 12 decades in particle energy ranging from the non-relativistic to the highly relativistic regime. Integrating the energy spectrum demonstrates that CRs, magnetic fields, thermal gas and turbulence are in pressure equilibrium in the midplane of the Milky Way, suggesting that CRs have an important dynamical role in maintaining the energy balance of the interstellar medium and likely play an active role during galaxy formation and evolution. In the Milky Way, the energy density of CR electrons is around one hundredth that of CR protons.

- In clusters, the large-scale radio emission in form of radio halos and relics proves the existence of CR electrons. We have not yet obtained a direct proof of the existence of CR protons in clusters. Because relativistic electrons and protons are indistinguishable in their wave-particle interactions (barring their sense of rotation around the magnetic field) and because relativistic electrons have shorter cooling times in comparison to CR protons, we also expect a significant CR proton population to be present, which could have significant thermodynamic implications.

- CRs are collisionless, i.e., they do not interact with the thermal plasma by scattering on individual particles but by interacting with electro-magnetic waves. As CRs gyrate around the mean magnetic field they can interact with Alfvén waves that propagate along the magnetic field and they experience a decelerating force as a result of the $\boldsymbol{v}_\perp \times \delta\boldsymbol{B}$ Lorentz force if the Doppler-shifted

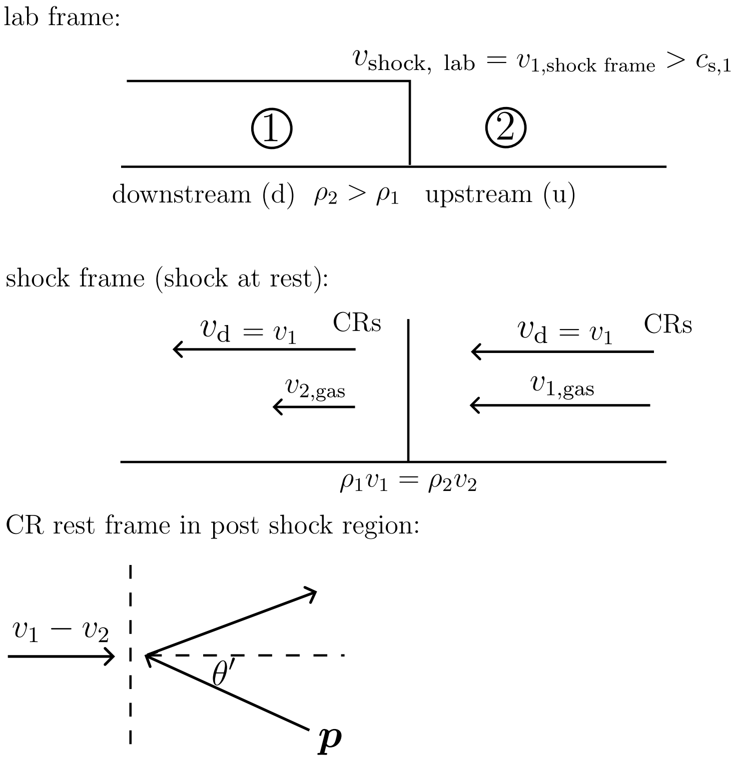


Figure 3.13: Definitions of the different frames at a shock.

MHD frequency ω equals the relativistic ion gyro frequency,

$$\Omega_i = \frac{ZeB}{\gamma m_i c} = \omega - k_\parallel v_\parallel. \tag{3.308}$$

Here, Ze, γ, m_i are the ion's charge, its Lorentz factor, and its mass, while $k_\parallel$ and $v_\parallel$ are the Alfvén wave number and particle velocity parallel to the mean magnetic field (see Fig. 3.12).

3.3.3.2 Diffusive Shock Acceleration

- At a collisionless shock, electrons and protons can be accelerated to highly relativistic energies, giving rise to a population of CRs. This mechanism only works in a collisionless plasma for particles with energies exceeding the thermal energy because frequent collisions with surrounding particles will cause severe energy loss and as a result no acceleration will occur. Hence, we only consider wave-particle interactions so that we can neglect the recoil of the scattering agents (the MHD wave modes) which are supported by the magnetic field and as such, acquire a huge inertia because magnetic fields are flux frozen into the thermal plasma.

- To explain how this process works, we consider a plane-parallel shock and show the definition of the different frames in Fig. 3.13. 1. In the lab frame, the shock moves with a velocity that exceeds the sounds speed, $\upsilon_s|_{\rm lab} = \upsilon_1|_{\rm shock\ frame} > c_{s,1}$ so that the gas upstream of the shock (i.e., in the pre-shock region) is causally unconnected and does not know that is approaches. 2. In the shock frame, the shock is at rest by definition and the upstream gas is decelerated at the shock so that $\upsilon_2 < \upsilon_1$. By contrast, the *collisionless* CR particle does not feel the shock and continues to move with its drift velocity $\upsilon_d = \upsilon_1$ in the downstream. 3. Finally, the rest frame of the CR particle in the downstream (i.e., the post-shock) region is obtained by boosting the shock frame by υ_1 towards the upstream (i.e., to the right in Fig. 3.13).

- We assume that there are resonant MHD wave modes in the upstream and downstream regimes that enable CRs to scatter. At the end of this subsection, we will come back to this assumption and show that a combination of plasma physical processes at collisionless shocks generates these MHD modes self-consistently. In the CR rest frame, the post-shock medium acts as an approaching "magnetic mirror". Scattering off of this magnetic mirror, i.e., scattering at macroscopic magnetic fluctuations or MHD waves, energizes the particle.

- We can work out a condition for a particle to scatter downstream and to reach the upstream again where it was initially advected with the flow. To this end, we perform Lorentz transformation from the lab to the post-shock frame and back to the lab frame and work out the energy and momentum balance during scattering events in these frames. Cluster accretion shocks are non-relativistic shocks with $\upsilon_s \ll c$ while shocks in AGN jets are (initially) relativistic.

- The Lorentz boost is taken along the shock normal with $\beta c = \upsilon_1 - \upsilon_2$. We define the parallel component of the particle momentum with respect to the shock normal in the lab frame, $p_\parallel = \cos\theta p \equiv \mu p$. In the post-shock frame (where all quantities are denoted by a prime), the CR energy and momentum are given by

$$\begin{pmatrix} E' \\ p'c \end{pmatrix} = \begin{pmatrix} \gamma & \pm\beta\gamma \\ \pm\beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} E \\ \mu pc \end{pmatrix} = \begin{pmatrix} \gamma E + \pm\beta\gamma\mu pc \\ \pm\beta\gamma E + \gamma\mu pc \end{pmatrix} \approx \begin{pmatrix} E + \beta\mu pc \\ \beta E + \mu pc \end{pmatrix}. \tag{3.309}$$

  In the last step, we assumed a non-relativistic shock and adopted the "+" sign for a Lorentz transformation in the direction of the CR's parallel momentum.

- After colliding with the "magnetic mirror" p' is reversed to $-p'$ and E' remains unchanged. Transforming back to the lab frame (where all quantities are denoted by a double prime), we find

$$\begin{aligned}\begin{pmatrix}E''\\p''c\end{pmatrix}&=\begin{pmatrix}\gamma & -\beta\gamma\\-\beta\gamma & \gamma\end{pmatrix}\begin{pmatrix}E'\\-p'c\end{pmatrix}\approx\begin{pmatrix}E+\beta\mu pc+\beta(\beta E+\mu pc)\\-\beta(E+\beta\mu pc)-\beta E-\mu pc\end{pmatrix}\\&\approx\begin{pmatrix}E+2\beta\mu pc\\-2\beta E-\mu pc\end{pmatrix}\equiv\begin{pmatrix}E+\delta E\\p+\delta p\end{pmatrix}.\end{aligned}\tag{3.310}$$

The two approximate signs indicate the limit of a non-relativistic shock.

- For a relativistically moving CR, we have a condition to be advected downstream, scatter there and reach the upstream again, provided

$$\delta E=\delta pc=2(\upsilon_1-\upsilon_2)\mu p\quad\text{and}\quad\mu=\cos\theta>0,\tag{3.311}$$

since $\beta c=\upsilon_1-\upsilon_2$. Hence the relative energy/momentum increase during an upstream-downstream-upstream ("udu") half cycle is

$$\frac{\delta p}{p}=\frac{2(\upsilon_1-\upsilon_2)\mu}{c}>0\quad(\mu>0).\tag{3.312}$$

- A similar calculation as above shows that energy/momentum is also increased during a "dud" half cycle:

$$\frac{\delta p}{p}=-\frac{2(\upsilon_1-\upsilon_2)\mu}{c}>0\quad(\mu<0).\tag{3.313}$$

Hence, there is a net energy increase of a particle that finishes a full cycle "udud".

- To obtain the energy increase of a CR population, we have to integrate these expressions in Eqs. (3.312) and (3.313) for individual particles that collectively make up a distribution function $f(\boldsymbol{x},\boldsymbol{p})$ over phase space. We conveniently employ spherical coordinates in momentum space (p, $\mu=\cos\theta$, ϕ_p) and cylindrical coordinates in configuration space (r, ϕ, z). The first two coordinates span an (arbitrary) cross section σ of the plane-parallel shock front and the z coordinate is transformed into a time coordinate via $\mathrm{d}z=\boldsymbol{n}_\mathrm{s}\cdot\boldsymbol{\upsilon}\,\mathrm{d}t$, where $\boldsymbol{n}_\mathrm{s}$ denotes the shock normal. Hence, we have the differential volume element $\mathrm{d}^3x=\mathrm{d}\sigma\,\boldsymbol{n}_\mathrm{s}\cdot\boldsymbol{\upsilon}\,\mathrm{d}t=\mathrm{d}\sigma\upsilon\mu\mathrm{d}t$. For a half cycle "udu", we get

$$\begin{aligned}\frac{\delta p}{p}&=2\frac{\upsilon_1-\upsilon_2}{c}\,\frac{1}{2}\,\frac{\int\mu f(\boldsymbol{x},\boldsymbol{p})\mathrm{d}^3x\mathrm{d}^3p}{\int f(\boldsymbol{x},\boldsymbol{p})\mathrm{d}^3x\mathrm{d}^3p}\\&=2\frac{\upsilon_1-\upsilon_2}{c}\,\frac{\int_0^1\mu f\sigma\upsilon\mu\mathrm{d}t2\pi p^2\mathrm{d}p\mathrm{d}\mu}{2\int_0^1 f\sigma\upsilon\mu\mathrm{d}t2\pi p^2\mathrm{d}p\mathrm{d}\mu}\\&=2\frac{\upsilon_1-\upsilon_2}{c}\,\frac{\mu^3/3\big|_0^1}{\mu^2\big|_0^1}=\frac{2}{3}\,\frac{\upsilon_1-\upsilon_2}{c}\end{aligned}\tag{3.314}$$

where the factor of 1/2 in the first line accounts for the half space in the downstream region and we assume isotropy in momentum space.

- An analogous calculation for the second half cycle yields the same answer so that we obtain a relative energy gain for a full cycle, averaged over all particles,

$$\varepsilon \equiv \frac{\delta E}{E} = \frac{\delta p}{p} = \frac{4}{3} \frac{\upsilon_1 - \upsilon_2}{c}. \tag{3.315}$$

  This is the central result of this calculation: after crossing the shock surface, the CR diffuses and scatters on the turbulent, magnetic field in the upstream and downstream regions. Because the plasma is moving towards the CR in both regions, the CR always sees an approaching "magnetic mirror" so that it gains the same amount of energy and momentum in the upstream and downstream regions. Because the shock represents a macroscopic velocity divergence it breaks the symmetry and ensures that the energy gain per shock crossing during the process of diffusive shock acceleration is "first order" in $\beta = (\upsilon_1 - \upsilon_2)/c$. Hence, this process is also called first-order Fermi acceleration.

- The escape probability is the ratio of particle flux carried by the downstream flow at υ_2 over the particle flux that crosses the shock front at speed c (assuming again relativistic particles for simplicity),

$$P_\mathrm{e} = \frac{f_2 4\pi p^2 \mathrm{d}p \sigma \upsilon_2 \mathrm{d}t}{\int_0^1 f_0 2\pi p^2 \mathrm{d}p \sigma c \mu \mathrm{d}t \mathrm{d}\mu} = \frac{4\upsilon_2}{c} \frac{f_2}{2 f_0 \, \mu^2/2\big|_0^1} \approx \frac{4\upsilon_2}{c}. \tag{3.316}$$

  Here, f_2 is the distribution function far downstream the shock, f_0 if the distribution function after the shock transition and in the last step, we assume that the CR distribution function is rapidly homogenized behind the shock, i.e., $f_2 = f_0$.

- After n cycles, the particle has the energy

$$E = E_0 \,(1 + \varepsilon)^n \quad \Rightarrow \quad n = \frac{\log(E/E_0)}{\log(1 + \varepsilon)}. \tag{3.317}$$

- At each acceleration cycle, there is the escape probability P_e. After n cycles, the particle has the probability $(1 - P_\mathrm{e})^n$ still to participate in the process. Hence, the number of particles with energy larger than E is given by

$$\begin{aligned} N(> E) &\propto \sum_{m=n}^{\infty} (1 - P_\mathrm{e})^m = (1 - P_\mathrm{e})^n \sum_{m=0}^{\infty} (1 - P_\mathrm{e})^m \\ &= (1 - P_\mathrm{e})^n \frac{1}{1 - (1 - P_\mathrm{e})} = \frac{(1 - P_\mathrm{e})^n}{P_\mathrm{e}}, \end{aligned} \tag{3.318}$$

where we have used the geometric series in the second to last step. Using the identity $a^{\ln b} = \left(\mathrm{e}^{\ln a}\right)^{\ln b} = b^{\ln a}$, we can rewrite Eq. (3.318) and obtain

$$N(>E) \propto \frac{1}{P_\mathrm{e}}\left(\frac{E}{E_0}\right)^{-\tilde{\alpha}}, \quad \tilde{\alpha} = -\frac{\log(1-P_\mathrm{e})}{\log(1+\varepsilon)}. \tag{3.319}$$

- Using Eqs. (3.315) and (3.316) and Taylor expanding $\log(1+x) = x - x^2/2 + O(x^3)$, we get

$$\tilde{\alpha} = -\frac{\log\left(1-\dfrac{4v_2}{c}\right)}{\log\left(1+\dfrac{4}{3}\dfrac{v_1-v_2}{c}\right)} \approx \frac{4v_2/c}{\dfrac{4}{3}\dfrac{v_1-v_2}{c}} = \frac{3v_2}{v_1-v_2} = \frac{3}{r-1}, \tag{3.320}$$

  where we introduced the density compression factor $r = \rho_2/\rho_1 = v_1/v_2$ in the last step.

- The cumulative particle energy spectrum due to diffusive shock acceleration is therefore given by

$$N(>E) = \int N(E)\mathrm{d}E \propto \frac{c}{4v_2}\left(\frac{E}{E_0}\right)^{-3/(r-1)}. \tag{3.321}$$

- As one can easily verify, the differential CR energy spectrum $N(E) \propto E^{-\alpha}$ has a spectral index α that connects to $\tilde{\alpha}$ via

$$-\tilde{\alpha} = 1 - \alpha = -\frac{3}{r-1}, \quad \text{so that} \tag{3.322}$$

$$\alpha = 1 + \frac{3}{r-1} = \frac{r+2}{r-1} \to 2 \tag{3.323}$$

  for a strong shock with $r = 4$.

- Realizing that the density compression factor at the shock $r = \rho_2/\rho_1$ is a function of Mach number only (Eq. 3.118) we can rewrite the CR spectral index so that it becomes a function of Mach number only,

$$\alpha = \frac{r+2}{r-1} = \frac{\frac{1}{2}(3\gamma-1)\mathcal{M}_1^2+2}{\mathcal{M}_1^2-1} = \frac{2\mathcal{M}_1^2+2}{\mathcal{M}_1^2-1}, \tag{3.324}$$

  where we have adopted $\gamma = 5/3$ in the last step.

- In deriving this result, we have adopted a one-dimensional geometry, i.e., effectively an infinitely extended shock surface and assumed that there is no additional physics associated with the two dimensions perpendicular to the shock normal. In particular, we have assumed that particles cannot escape from the shock in the perpendicular directions. To account for the finite extend of realistic shocks, we can estimate the maximum energy of a particle that are still confined by the shock. Integrating the Lorentz force yields the work necessary to overcome confinement, which is also known as the Hillas criterion for the maximum particle energy of an accelerator,

$$E_{\max} = \int \boldsymbol{F}_{\mathrm{L}} \cdot \mathrm{d}\boldsymbol{r} = ZeBR \sim 10^{20} \left(\frac{B}{0.1\,\mu\mathrm{G}}\right)\left(\frac{R}{1\,\mathrm{Mpc}}\right)\,\mathrm{eV}, \tag{3.325}$$

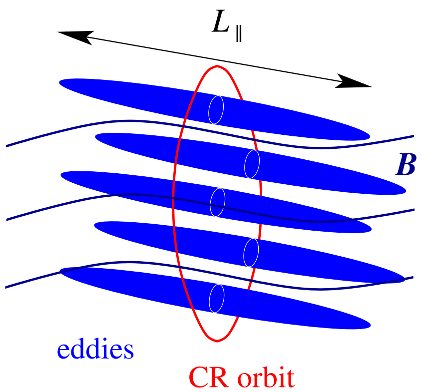
where Ze is the ion charge, B is the root-mean-square magnetic field and R is the characteristic shock radius. This would imply that galaxy cluster accretion shocks can accelerate ultra-high energy cosmic rays, which are the highest energy elementary particles ever detected. However, if CRs are accelerated via diffusive shock acceleration, $E_{\max}$ is reduced by the efficiency factor $\beta_{\mathrm{s}} = \upsilon_{\mathrm{s}}/c$, so that the resulting energy is $E_{\max} \sim 10^{18}$ eV, assuming a shock velocity $\upsilon_{\mathrm{s}} \sim 3,000$ km s^{-1}, which is still very high.

- We have seen that astrophysical collisionless, non-relativistic shocks in the regime of high Mach numbers (equivalent to high shock compression ratios) can accelerate ions and electrons. However, this result relies on the existence of resonant wave modes up- and downstream of the shock that efficiently scatter CRs. Thanks to the growing computational power, which enabled detailed numerical simulations of the kinetic physics at collisionless shocks, we have now learned that diffusive shock acceleration is a highly non-linear process that efficiently accelerates ions and electrons, amplifies magnetic fields (or generates them from scratch), and exchanges energy between electrons and ions. Removing one piece of the physics (collisionless shocks, energetic particles, or electro-magnetic waves) would collapse the entire process.

- The plasma simulations show that once ions are advected towards a shock and are impinging on the shock transition layer, they can be specularly reflected at the time-dependent electromagnetic shock potential provided it has favorable conditions. As the reflected ions propagate upstream, they non-resonantly excite a spectrum of Alfvén waves (Bell 2004). Efficient resonant scatterings of lower energy ions at these waves cause the ions to be reflected, to be advected back to the shock and to cross it. We have seen that each shock crossing and the successive recoil-free scattering results in an energy gain that is proportional to $(\upsilon_1 - \upsilon_2)/c$ and that the particles have a probability of leaving the Fermi cycle by being swept downstream. Overall, this yields a power-law momentum spectrum with a maximum energy that increases with time.

3.3.3.3 Second-order Fermi Acceleration

- In addition to the very efficient first-order Fermi acceleration mechanism, there also exists second-order Fermi acceleration. Unlike diffusive shock acceleration, where the shock breaks the symmetry and ensures that the CRs always see an approaching scattering "magnetic mirror" as they transition the shock discontinuity, second-order Fermi acceleration relates to the amount of energy gained during the motion of a charged particle in the presence of randomly moving magnetic mirrors. If the magnetic mirror is moving towards the particle, the particle will gain energy upon reflection while the opposite holds if the mirror is receding.
- This property was used by Fermi (1949) to explain the generation of CRs. In this case the magnetic mirror is a moving interstellar magnetized cloud. Fermi argued that the probability of a head-on collision is greater than a head-tail collision for randomly moving clouds, so particles would, on average, be accelerated. This random process is now called second-order Fermi (Fermi-II) acceleration, because the mean energy gain per bounce depends on the mirror velocity squared, $\beta_{\rm m}^2 \sim (\upsilon_{\rm m}/c)^2$ and because in the rest system of the galaxy, the mean motion of the clouds vanishes.
- The resulting energy spectrum anticipated from this physical setup, however, is not universal as in the case of diffusive shock acceleration. As in that case, Fermi-II acceleration only works in a collisionless plasma and requires particles with energies exceeding the thermal energy, because otherwise, frequent collisions with surrounding particles will share their energy among all scattering partners and as a result no acceleration will occur.
- In practice, CRs gain energy through Fermi-II acceleration by interacting with wave modes of MHD turbulence. But how exactly does this work? In Section 3.3.2.6 we encountered Alfvénic MHD turbulence which however turns out to have a negligibly small contribution to particle acceleration provided the turbulence has been injected on a scale large in comparison to the CR's gyro orbit. Consider the schematic picture of turbulent eddies in the framework of Goldreich-Shridhar Alfvénic turbulence in this case (see the sketch to the right.)



Sketch of anisotropic, turbulent eddies of Alfvénic turbulence at the gyroresonant scale of a CR.

- The gyro-radius of a CR encloses many eddies that are not coherently aligned because of the anisotropic property of Alfvénic turbulence:

$$\lambda \equiv L_\perp \ll L_\parallel \sim r_\mathrm{L} = \frac{p_\perp c}{ZeB}. \tag{3.326}$$

Here, we assumed a gyroresonant Alfvén wave-CR interaction (see Fig. 3.12), which implies $L_\parallel \sim r_\mathrm{L}$. This incoherence of the electro-magnetic eddies implies incoherent Lorentz forces during a gyro orbit and causes a random walk of the CRs. This broadens the gyro resonance and reduces the CR scattering efficiency.

- Equivalently, we can make the same argument in k-space. From Eq. (3.307), we get

$$k_\perp \propto L_\mathrm{MHD}^{1/2} k_\parallel^{3/2}, \tag{3.327}$$

$$\mathrm{d}k_\perp \propto \frac{3}{2} L_\mathrm{MHD}^{1/2} k_\parallel^{1/2} \mathrm{d}k_\parallel. \tag{3.328}$$

Perpendicular to $\langle \boldsymbol{B} \rangle$, we have Kolmogorov turbulence (see Eq. 3.92) with

$$E(k_\perp)\mathrm{d}k_\perp \propto k_\perp^{-5/3} \mathrm{d}k_\perp \tag{3.329}$$

but the energy spectrum parallel to $\langle \boldsymbol{B} \rangle$ is much steeper than in the Kolmogorov case,

$$\begin{aligned} E(k_\parallel)\mathrm{d}k_\parallel &= E\left[k_\perp(k_\parallel)\right] \frac{\mathrm{d}k_\perp}{\mathrm{d}k_\parallel} \mathrm{d}k_\parallel \\ &\propto k_\parallel^{-5/2}\, k_\parallel^{1/2} \mathrm{d}k_\parallel = k_\parallel^{-2} \mathrm{d}k_\parallel. \end{aligned} \tag{3.330}$$

This steeper spectrum in comparison to Kolmogorov makes parallel modes decay faster and implies that there is less wave energy available for CR scattering on the resonant scale. Hence Alfvénic turbulence is not efficient in scattering CRs.

- In addition to Alfvén modes there are compressible fast modes that dominate the CR scattering in spite of their fast damping. Let's return to the resonance condition for CR-wave interactions in Eq. (3.308) and look at the full resonance condition:

$$\omega - k_\parallel v_\parallel = n\Omega_\mathrm{i} = n \frac{ZeB}{\gamma m_\mathrm{i} c}, \quad n \in \{0, \pm 1, \pm 2, \ldots\}, \tag{3.331}$$

which states that the Doppler-shifted MHD wave frequency ω is a multiple of the relativistic ion gyro frequency, Ω_i. The cases $n \in \{\pm 1, \pm 2, \ldots\}$ refer to resonant interactions of the CR and the fundamental mode and its harmonics of the corresponding Alfvén waves, which are not efficient for CR scattering if the turbulence is not injected close to the CR's gyro scale.

- The case $n = 0$ (the "Landau resonance") refers to a non-resonant interaction named "transit time damping" with $\omega = k_{\|} v_{\|}$. This implies that the particle transit time $t = \lambda_{\|}/v_{\|} = 2\pi/(k_{\|} v_{\|})$ across the confining region matches the wave period, $T = 2\pi/\omega$. Physically, this means that the electron or proton is trapped by a magnetic mirror force. The particle surfs the wave and gains energy because head-on collisions are more frequent than head-tail collisions. This leads to stochastic acceleration with an efficiency that is proportional to the square of the velocity of the fast magnetosonic waves divided by the CR velocity (which is close to the light speed), $\beta_{\mathrm{w}}^2 \sim (v_{\mathrm{w}}/c)^2$, in the high-$\beta$ plasma of galaxy clusters.

- To conclude, let's compare first-order and second-order Fermi acceleration:

  - First-order Fermi acceleration works at shocks and is very efficient. The basic aspects of it are well understood. But shocks, in particularly strong (high Mach number) shocks are rare, especially within the cluster volume.
  - By contrast, second-order Fermi acceleration is less efficient. However, subsonic compressible turbulence should be ubiquitous in the ICM, which is permanently shaken by minor and (less frequently) major cluster mergers. The process itself involves difficult plasma physics and it less well understood. There has not yet been a first-principle theory of second-order Fermi acceleration suggested.

3.3.3.4 Cosmic Ray Transport

- Consider a spatial random coordinate $x(t)$ of a CR particle diffusing in a fluid of bulk velocity v. For simplicity, we restrict ourselves to the one-dimensional case. During a time interval which is much shorter compared to the diffusion time, the particle's position varies by $\Delta x = v\Delta t + \delta x$. The first contribution is due to the bulk motion of the scattering medium and the second term is due to the random walk diffusion with vanishing mean and the variance $\langle \delta x^2 \rangle = 2D(x, p)\Delta t$, where $D(x, p)$ denotes the diffusion coefficient.

- The distribution of CRs is governed by a competition between injection, escape, energy gain (acceleration), and energy loss (catastrophic and continuous) processes. The transport equation which describes the balance of these processes is a Fokker-Planck type equation that includes the description of fluid motions, radiative losses, and phase space diffusion. It can be obtained by considering the collisionless Boltzmann equation and working out the magneto-hydrodynamic forces acting on a CR particle including the Lorentz force as well as pitch angle scattering on hydromagnetic waves (details can be found in my *High Energy Astrophysics* lectures; in the following I only sketch the picture).

- The transport equation governs the evolution of the isotropic part $f(x, p)$ of the CR distribution function in phase space, assuming weak anisotropy of the CR momentum distribution function:

$$\begin{aligned}\frac{\partial}{\partial t}f + \frac{\partial}{\partial x}v(x,p)f &= -\frac{1}{p^2}\frac{\partial}{\partial p}p^2 A(x,p)f + \frac{1}{p^2}\frac{\partial}{\partial p}p^2\Gamma(x,p)\frac{\partial}{\partial p}f \\ &\quad + \frac{\partial}{\partial x}D(x,p)\frac{\partial}{\partial x}f + s(x,p). \qquad (3.332)\end{aligned}$$

The distribution is normalized such that the number density of CRs $n_{\mathrm{CR}} = 4\pi \int f p^2 \mathrm{d}p$. The 'friction' term A describes not only various kinds of energy losses but also the energy gain by first order processes in $\beta \equiv v/c$ (adopting relativistic particles), the second contribution on the right-hand side describes the energy gain through the second-order Fermi process, the third term on the right-hand side describes spatial diffusion, and the last term accounts for sources such as freshly injected CR particles at shocks whose origin can be understood by means of plasma physical calculations. The physical meaning of these processes will be sketched in the following:

  - **Synchrotron and inverse Compton losses.** A relativistic charged particle of a Lorentz factor $\gamma = (1-\beta^2)^{-1/2}$ experiences Compton scattering with either real or virtual photons (which represent the magnetic field in the case of synchrotron radiation). This causes the particle to emit photons in the forward direction into a narrow cone of half-angle γ^{-1} with respect to its momentum leading to an energy loss which can effectively be described by a friction force in opposite direction to its momentum:

$$A_{\mathrm{rad,e/i}} \equiv \left.\frac{\langle \Delta p \rangle}{\Delta t}\right|_{\mathrm{rad}} = -\frac{4}{3}\sigma_{\mathrm{T}}\left(\frac{m_{\mathrm{e}}}{m_{\mathrm{e/i}}}\right)^2 (\varepsilon_B + \varepsilon_{\mathrm{ph}})\gamma^2, \qquad (3.333)$$

    where σ_{T} denotes the Thompson cross section, $\varepsilon_B = B^2/(8\pi)$ and $\varepsilon_{\mathrm{ph}}$ are the energy densities of the magnetic field (responsible for synchrotron losses) and the low energy photon field (causing the Compton effect in the Thompson regime). The radiative losses of ions are suppressed by $(m_{\mathrm{e}}/m_{\mathrm{i}})^2$ such that they can be neglected unless they are ultra-high energetic CRs with energies $\gtrsim 10^{18}$ eV.

  - **First-order Fermi process.** The contribution of the first order Fermi process can be described by a non inertial entrainment due to the deceleration of the scattering medium:

a compressed flow ($\boldsymbol{\nabla} \cdot \boldsymbol{v} < 0$) produces first order acceleration of charged particles. In this situation, the inertial force is $F_j = -p_i(\partial v_j/\partial x_i)$ that gives rise to an accelerating power

$$P_{\rm acc} = -\langle v_j p_i \rangle \frac{\partial v_j}{\partial x_i} = -\frac{pv}{3}\boldsymbol{\nabla} \cdot \boldsymbol{v} \quad \to \quad A_{\rm acc} = -\frac{p}{3}\frac{\partial v}{\partial x}. \tag{3.334}$$

- **Second-order Fermi process.** Charged particles gyrate around, and travel slowly along magnetic field lines. Occasionally, they get scattered on magnetic irregularities and plasma waves. This scattering process can be described by a random walk of the particle's pitch angle with the magnetic field direction, θ, yielding the characteristic variance $\langle \delta\mu^2 \rangle \propto \nu_{\rm s}\Delta t$ where $\nu_{\rm s}$ denotes the average scattering frequency and $\mu = \cos\theta$. Because the particle scatters off of moving magnetic mirrors, the particle systematically gains energy through random variations of the CR momentum $\delta p = \pm\beta_{\rm m} p \delta\mu$ where $\beta_{\rm m} = v_{\rm m}/c$ is the dimensionless velocity of the magnetic mirror. The second-order Fermi process is thus described by a diffusion process in momentum space with the momentum diffusion coefficient

$$\Gamma \equiv \frac{\langle \delta p^2 \rangle}{2\Delta t} \sim \beta_{\rm A}^2 \nu_{\rm s} p^2. \tag{3.335}$$

- **Diffusive losses from the central cluster region.** CRs experience momentum dependent diffusion in a turbulent magnetic field with a Kolmogorov-type spectrum on small scales. This process leads to a loss time scale which is proportional to $p^{-1/3}$. In an equilibrium situation, this results in a steepening of the observed spectrum within the cluster core by $p^{-1/3}$ relative to the injected spectrum.
- **Coulomb losses** are strongest for protons or heavier nuclei, but also relevant for electrons. Energetic CRs experience energy losses within an ionized medium through Coulomb interactions. Coulomb losses efficiently remove the low-energetic part of the injected CR spectrum on a short timescale and redistribute these particles and their energy into the thermal pool.
- **Catastrophic losses.** Another loss process is the inelastic reaction of CR nuclei with ions, atoms and molecules of the ICM. The CR protons interact hadronically with the ambient thermal gas and produce mainly neutral and charged pions, provided their momentum exceeds the kinematic threshold of 0.78 GeV for the reaction. The neutral pions successively decay into γ-rays while the charged pions decay into secondary electrons and neutrinos.

- **Spallation.** Spallation describes the destruction of atomic nuclei in a collision with a CR particle that is in most cases a proton or an alpha particle. In this destruction process, many pieces of debris are formed where commonly a single nucleon gets stripped and a distribution of lighter nuclei is obtained. Since the abundances of the elements Lithium, Beryllium, and Boron are much larger in CRs than in the interstellar medium, spallation processes are assumed to account for the origin of these elements. For any specific isotope, spallation can again occur as a loss or a gain process in the equation of balance.

Chapter 4

Cluster Astrophysics and Cosmology across Wavelengths

4.1 Optical: Galaxy Interactions and Virial Theorem

4.1.1 Observational Facts

- **Morphology-density relation.** Observations of galaxy types reveal that denser environments which are found in galaxy groups and clusters host larger fractions of galaxies that are morphologically classified as early types (ellipticals) and lenticular galaxies of type S0, i.e., an intermediate between an elliptical and a spiral galaxy in galaxy morphological classification schemes which contains a large-scale disc and a central bulge. This is very different from low-density environments in the "field" where late-type galaxies of spiral morphology dominate the numbers. In addition, galaxies in denser environments are on average redder, are more massive, more concentrated, less gas rich and have lower specific star formation rates.

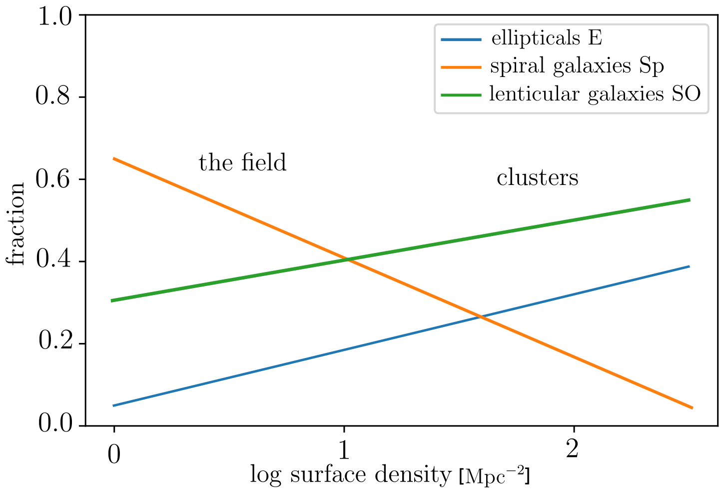


Figure 4.1: Fraction of galaxy types versus environmental density. The decreasing fraction of spirals and the increasing fraction of ellipticals and galaxies with bulges at higher densities is called the morphology-density relation.

- **Butcher-Oemler effect.** Galaxy populations in clusters at intermediate redshifts ($0.3 \lesssim z \lesssim 0.5$) have a dramatically increased fraction of blue galaxies in comparison to present-day clusters. In addition, morphological studies have shown that this effect, which was discovered by Butcher and Oemler, is associated with an increasing spiral and star forming fraction with increasing redshift. Many of these star forming spirals show disturbed morphologies.
- Although far from unambiguous, these strong environmental and redshift dependencies are often interpreted as indications that

Figure 4.2: The mice galaxies, NGC 4676A+B, showing a spectacular tidal interaction.

galaxies undergo transformations from late-type (spirals) to early-type galaxies (ellipticals) or from star forming to passive galaxies once they become part of a denser environment. Clusters as the largest collapsed structures in the Universe are also the environments with the highest number densities of galaxies. Hence, galaxy interactions are frequent, making clusters the ideal environments to look for possible transformational processes of their morphologies.

- We will now discuss various processes that operate in clusters and may be responsible for transforming star forming disks into passive spheroids:

  1. **Tidal interactions** with other cluster galaxies or with the cluster potential can torque the stellar orbits and the gas distribution and cause a disturbed non-equilibrium morphology (see Fig. 4.2) or even transform its type.
  2. **Dynamical friction** can cause the galaxy to slowly migrate to the cluster center and eventually to merge with the *cD galaxy* that resides in the cluster center.
  3. **Ram-pressure interactions** of a galaxy with the hot, X-ray emitting ICM severely impacts the interstellar medium of a galaxy.

- A *cD galaxy* is a galaxy morphology classification, a subtype of a type-D giant elliptical galaxy that can be found near the centres of rich galaxy clusters. They are characterized by a large halo of stars, which is believed to have formed through tidally stripped stars of galaxies that have migrated to the center via dynamical friction and have then merged with the central elliptical galaxy to form such a supergiant elliptical or central dominant galaxy. Some cDs even have multiple galactic nuclei which is strong evidence for their ongoing merging activity.

- cD galaxies are one of the types frequently found to be the *brightest cluster galaxy (BCG)* of a cluster, which is defined as the brightest galaxy in a cluster of galaxies. BCGs include the most massive galaxies in the universe. They are generally elliptical cD galaxies which lie close to the geometric and kinematical center of their host galaxy cluster, at the bottom of the cluster potential well. Hence, the terms cD galaxy and BCG are often used synonymously.
- The modern picture of galaxy formation has refined the original thinking: transformational processes of galaxies are not only important in clusters today but especially in proto-cluster environments or groups before those merged into clusters.

4.1.2 Tidal Interactions of Galaxies

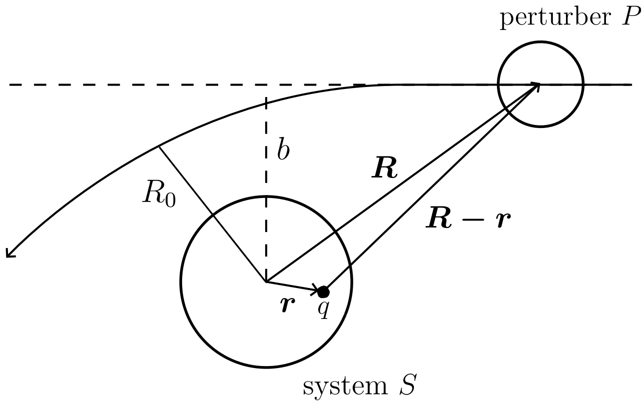


Figure 4.3: Geometry of the interaction of two systems. A perturber P modifies the orbit of a particle q in system S as a result of the gravitational interaction.

- Consider a body S which has an encounter with a perturber P, which approaches S with an impact parameter b (see Fig. 4.3). Let q be a particle (e.g., a star) in S at a distance $\boldsymbol{r}$ from the center. Because the gravitational force due to P is not uniform over the body S, the particle q experiences a tidal force per unit mass of

$$\tilde{\boldsymbol{F}}_{\text{tide}}(\boldsymbol{r}) = -\boldsymbol{\nabla}\phi_p(|\boldsymbol{R}-\boldsymbol{r}|) + \boldsymbol{\nabla}\phi_p(\boldsymbol{R}) \tag{4.1}$$

where ϕ_p is the gravitational potential of P. As a result of the encounter, the particle q gains energy at a rate per unit mass

$$\frac{\mathrm{d}\epsilon_q}{\mathrm{d}t} = \boldsymbol{v}\cdot\tilde{\boldsymbol{F}}_{\text{tide}}(\boldsymbol{r}), \tag{4.2}$$

where $\boldsymbol{v}$ is the velocity of q with respect to the center of S.

- Similar to the way the Moon induces oceanic tides on Earth, the gravitational interaction between S and P enhances the gravitational multipole moments of both bodies which may cause a backreaction on their orbit. Let t_{tide} be the time for the tide to rise and $t_{\text{enc}} \approx R_{\max}/\Delta v_{SP}$ is the time of the encounter where Δv_{SP} is the relative velocity between S and P and $R_{\max} = \max(R_0, R_S, R_P)$, where R_S and R_P are the characteristic radii of S and P, and R_0 is the minimum distance of the encounter.

- If $t_{\text{enc}} \gg t_{\text{tide}}$ then the internal structure of deformable bodies has enough time to adiabatically adjust to the perturbation in form of tides (due to the change of relative position and orientation of S and P). Hence, the effects of the encounter during approach and departure cancel each other (the deformations are adiabatic). As a result, there is no transfer of energy.

- If $t_{\text{enc}} < t_{\text{tide}}$ then the response of the bodies lags behind the instantaneous magnitude and direction of the force. This causes a backreaction on the orbit. The net effect in this case is a transfer of orbital energy to internal energy of the two bodies, which increases the mutual binding energy of them.

4.1.2.1 Tidal Stripping

- **Tidal radius – the Roche problem.** First, we consider a slow encounter and work out the "tidal radius" outside of which material can get stripped. For simplification, let us image a galaxy on a circular orbit in a cluster. This can be generalized in a straight forward manner. What is the fate of the stars inside this galaxy?

- The differential tidal force per unit mass between a star and the galaxy center is (by magnitude)

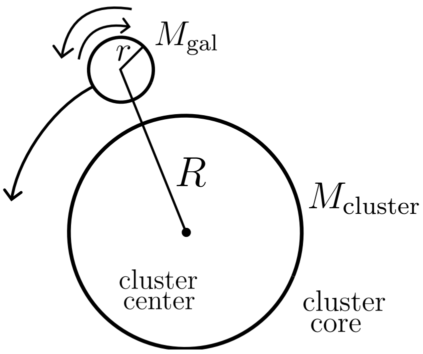
$$\tilde{F}_{\text{tide}}(R) = r\frac{\mathrm{d}}{\mathrm{d}R}\left(\frac{GM_{\text{cluster}}(R)}{R^2}\right) \tag{4.3}$$

  and the restoring force from the galaxy is

$$\tilde{F}_{\text{gal}}(r) = -\frac{GM_{\text{gal}}(r)}{r^2}. \tag{4.4}$$

- The tidal radius in the galaxy is defined by equating both forces, $\tilde{F}_{\text{tide}} = \tilde{F}_{\text{gal}}$, which yields

$$\frac{M_{\text{gal}}(r_{\text{tide}})}{r_{\text{tide}}^3} = -\frac{\mathrm{d}}{\mathrm{d}R}\left(\frac{M_{\text{cluster}}(R)}{R^2}\right) = \left(2 - \frac{\mathrm{d}\ln M_{\text{cluster}}}{\mathrm{d}\ln R}\right)\frac{M_{\text{cluster}}}{R^3}, \tag{4.5}$$



Setup for finding the tidal radius of a galaxy orbiting in a cluster.

which can be solved for the tidal radius,

$$r_{\rm tide} = \left[\frac{M_{\rm gal}}{M_{\rm cluster}(R)} \frac{1}{2 - \dfrac{\mathrm{d}\ln M_{\rm cluster}}{\mathrm{d}\ln R}} \right]^{1/3} R. \tag{4.6}$$

Equivalently, Eq. (4.5) can be reformulated in terms of average densities contained within the corresponding radii $r_{\rm tide}$ and R, which yields the "Roche criterion" for slow encounters,

$$\bar{\rho}_{\rm gal}(r_{\rm tide}) = \left(2 - \frac{\mathrm{d}\ln M_{\rm cluster}}{\mathrm{d}\ln R} \right) \bar{\rho}_{\rm cluster}(R), \tag{4.7}$$

so that the internal (stellar) distribution can adjust adiabatically to the perturbation provided $\rho > \bar{\rho}_{\rm gal}$ or, $r < r_{\rm tide}$.

- Outside $r_{\rm tide}$, stars are only loosely bound so that they can be "stripped off" by tidal forces exerted by the cluster core on the stars in the galaxy. This implies a resonance condition where the period of the stellar orbit in the galaxy at $r_{\rm tide}$ matches the period of the galaxy orbit inside the clusters,

$$P_\star(r_{\rm tide}) \approx P_{\rm gal}(R). \tag{4.8}$$

- There are obviously limitations to this simplified picture:
    1. Galaxy orbits are not circular. Hence, the formula for the tidal radius of Eq. (4.5) only applies to the pericenter of the galaxy orbit.
    2. Tidal effects depend on phase and inclination of stellar orbits within the galaxy. This complicates the picture to the point that in general, the problem has to be solved numerically.

4.1.2.2 Tidal Shocks: Galaxy Harassment

- Now, we treat a rapid encounter in the impulse approximation and assume that the time of passing $t_{\rm enc}$ is much smaller than the internal dynamical time $t_{\rm tide}$, as visualized in Fig. 4.4.

- Because the tidal force per unit mass exerted by P on a star in S is (by magnitude) $\tilde{F}_{\rm tide} = \mathrm{d}v_\star/\mathrm{d}t$, we obtain an impulsive change in star velocity relative to the center of mass in S over the interaction time, $2b/v_P$,

$$\Delta \boldsymbol{v}_\star \approx \frac{GM_P}{b^3} \frac{2b}{v_P} \begin{pmatrix} -x \\ -y \\ 0 \end{pmatrix} = \frac{2GM_P r}{b^2 v_P} \begin{pmatrix} -x \\ -y \\ 0 \end{pmatrix}. \tag{4.9}$$

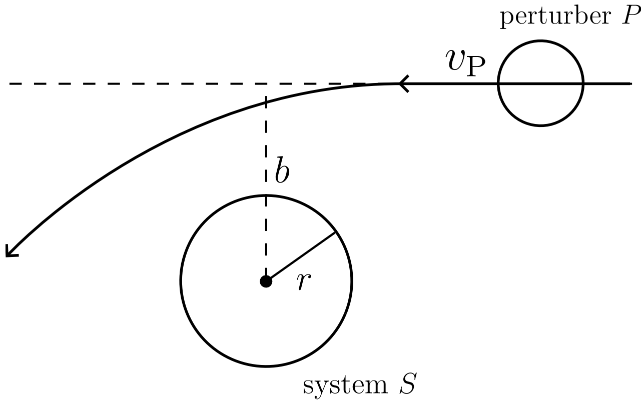


Figure 4.4: Setup for exerting a tidal shock in system S by the passage of a perturber P.

- The change in the total energy per unit mass of a star in S is given by

$$\Delta\epsilon = \frac{1}{2}(\boldsymbol{v}_\star + \Delta\boldsymbol{v}_\star)^2 - \frac{1}{2}\boldsymbol{v}_\star^2 = \boldsymbol{v}_\star \cdot \Delta\boldsymbol{v}_\star + \frac{1}{2}|\Delta\boldsymbol{v}_\star|^2 . \tag{4.10}$$

- We want to compute the change in total energy per unit mass $\Delta\epsilon_S$ by integrating $\Delta\epsilon$ over the entire system S. Because $\Delta\boldsymbol{v}_\star$ always points to the center of S, the problem is symmetric with respect to the point where $\boldsymbol{v}_\star$ is perpendicular to $\Delta\boldsymbol{v}_\star$ (assuming the Born approximation and integrating along unperturbed orbits). As a result, $\Delta v_{\star,\|}$ changes sign as the perturber passes through this point and obviously $\boldsymbol{v}_\star \cdot \Delta\boldsymbol{v}_{\star,\perp} = 0$. Hence, the integral of the first term on the right-hand side of Eq. (4.10) vanishes identically in the limit that the relative velocity between the two systems is much larger than the internal velocity dispersion of system S so that

$$\Delta\epsilon_S = \frac{1}{2M_S}\int |\Delta\boldsymbol{v}_\star|^2 \rho(\boldsymbol{r})\mathrm{d}^3 r. \tag{4.11}$$

  Substituting Eq. (4.9) into Eq. (4.11), we get a change in energy per unit mass (Spitzer 1958),

$$\Delta\epsilon_S = \frac{4}{3}\frac{G^2 M_P^2 \langle r^2\rangle}{v_P^2 b^4} \tag{4.12}$$

  where the assumption of spherical symmetry in S yields $\langle x\rangle = \langle y\rangle = \langle r\rangle/3$ and we adopted mass-weighted averages over all stars in a shell at radius

$$\langle r^2\rangle = \frac{1}{M_S}\int r^2 \rho(\boldsymbol{r})\mathrm{d}^3 r. \tag{4.13}$$

- What is the relative change in energy due to an impulsive encounter? Using the approximation $\epsilon_S(r) \approx -GM_S(r)/r$, we obtain

$$\frac{\Delta\epsilon_S}{\epsilon_S} \approx -\frac{4}{3}\frac{GM_P^2 r^3}{M_S(r) v_P^2 b^4} = -\frac{4}{3}\frac{\bar\rho_P(b)}{\bar\rho_S(r)}\frac{v_\mathrm{parabolic}^2}{v_P^2}, \tag{4.14}$$

where we used

$$\bar{\rho}_P(b) = \frac{M_P(b)}{\frac{4}{3}\pi b^3}, \quad \bar{\rho}_S(r) = \frac{M_S(r)}{\frac{4}{3}\pi r^3}, \quad v^2_{\text{parabolic}} = \frac{GM_P}{b}. \tag{4.15}$$

- There are a few notes in order:
  1. If $v_P \sim v_{\text{parabolic}}$, we obtain $\Delta\epsilon \sim \epsilon$ for $\rho_P \sim \rho_S$, i.e., we recover the Roche criterion.
  2. The formalism shown only applies to fast encounters, i.e., down to r_{tide}, below which the adiabatic case applies.
  3. The cumulative effect of many encounters can strip stars to radii $r \ll r_{\text{tide}}$.

4.1.3 Dynamical Friction

- When a heavy object of mass m_2 (e.g., a galaxy) moves through a large collisionless system that consists of particles of mass $m_1 \ll m_2$ (e.g., dark matter particles), it experiences a drag force that is called dynamical friction, which transfers energy and momentum from the galaxy to the dark matter particles. The reason for this is that a system evolves towards thermodynamic equilibrium through energy exchange by means of two-body encounters. If particles have different masses, thermodynamical equilibrium implies $m_1\langle v_1^2\rangle = m_2\langle v_2^2\rangle$. Because $m_2 \gg m_1$ and dark matter particles at the same radius have similar orbital velocities (which are both sourced by the underlying dark matter potential), the galaxy has usually larger kinetic energy than the dark matter particles it encounters. As a result this causes the galaxy to loose net energy and momentum.

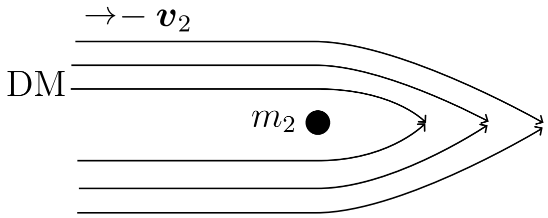


Figure 4.5: Sketch to visualize the physics of dynamical friction. In its system of reference, the galaxy experiences a wind of light particles that blows from the left to the right. The wind particles are gravitationally deflected into the galaxy's wake to produce an overdensity.

- An alternative way to understand dynamical friction is to move in the rest system of the galaxy. In this frame, dark matter particle are deflected through the gravity of the galaxy to form an

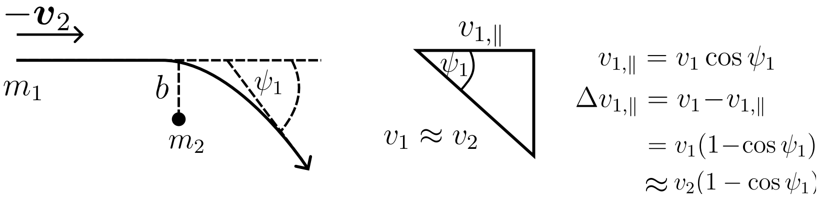


Figure 4.6: Defining the angles during a deflection event.

overdensity of dark matter in the wake (a so-called "trailing enhancement"). The gravitational pull of this wake on the galaxy slows it down.

- In real systems with a gravitational potential, dynamical friction produces mass segregation and not equipartition. Consider a massive galaxy on a circular orbit within a cluster that experiences dynamical friction. As a result of energy loss it transitions inward to a tighter bound orbit and moves faster.

4.1.3.1 The Chandrasekhar Formula

- In a single encounter, the deflection angle due to gravity is (in the rest system of the galaxy with mass m_2)

$$\psi_1 = \frac{\Delta v_\perp}{v_2} \approx \underbrace{\frac{Gm_2}{b^2}}_{|\tilde{\boldsymbol{F}}_\perp|} \underbrace{\frac{2b}{v_2}}_{t} \frac{1}{v_2} = \frac{2Gm_2}{bv_2^2}. \tag{4.16}$$

- Momentum balance of the deflection yields (with $m_1 \ll m_2$, see Fig. 4.6):

$$\Delta v_{2,\|} = -\frac{m_1}{m_2} \Delta v_{1,\|} = -\frac{m_1}{m_2} v_2 (1 - \cos\psi_1) \approx -\frac{2G^2 m_1 m_2}{b^2 v_2^3}, \tag{4.17}$$

using the Taylor expansion of $\cos\psi_1 = 1 - \psi_1^2/2 + O(\psi_1^4)$ and substituting Eq. (4.16). Note that we assume small-angle scatterings that dominate the total scattering rate (see Section 3.2.5.2).

- Only particles with speed $v_1 < v_2$ contribute to dynamical friction and are deflected by the galaxy with a differential rate that is given by

$$\mathrm{d}\Gamma = \int n(< v_2) v_2 \mathrm{d}\sigma, \tag{4.18}$$

where $\mathrm{d}\sigma$ is the differential scattering cross section and the number density of light particles is

$$n(< v_2) = 4\pi \int_0^{v_2} f(v) v^2 \mathrm{d}v, \tag{4.19}$$

where we assume that the distribution function $f(v)$ is isotropic. Hence, we obtain for the rate of change of the galaxy velocity

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{v}_2}{\mathrm{d}t} &= \int_{b_{\min}}^{b_{\max}} \Delta v_{2,\|} n(< v_2) v_2 2\pi b \mathrm{d}b \times \frac{\boldsymbol{v}_2}{|\boldsymbol{v}_2|} \\ &= -4\pi G^2 m_2 \rho(< v_2) \ln \Lambda \frac{\boldsymbol{v}_2}{|\boldsymbol{v}_2|^3}, \end{aligned} \tag{4.20}$$

where we used Eq. (4.17), $\rho(< v_2) = m_1 n(< v_2)$ is the mass density of light particles moving slower than v_2, and $\Lambda = b_{\max}/b_{\min}$. Clearly, the symmetry of the problem only causes the galaxy to slow down and not to change direction. This formula was first derived by Chandrasekhar in 1943.

- We can draw a few far-reaching conclusions from Chandrasekhar's formula of dynamical friction (identifying $\boldsymbol{v}_2 = \boldsymbol{v}_{\mathrm{gal}}$ for concreteness):

  1. The rate of change of the galaxy velocity, $\dot{v}_{\mathrm{gal}}$, is independent of the dark matter mass (i.e., whether it is composed of WIMPS, axions or primordial black holes) and only depends on the mass density ρ.
  2. We find $\dot{v}_{\mathrm{gal}} \propto m_{\mathrm{gal}}$, implying that heavier galaxies experience a larger drag which moves them faster to the bottom of the cluster potential and causes a segregation of galaxy masses in a cluster.
  3. We assume a Boltzmann distribution for dark matter (DM), $f(r, \boldsymbol{v}) = n(r)(2\pi\sigma_{\mathrm{DM}}^2)^{-3/2} \exp(-\boldsymbol{v}^2/2\sigma_{\mathrm{DM}}^2)$, and $v_{\mathrm{gal}} \ll \sigma_{\mathrm{DM}}$, where σ_{DM} is the DM velocity dispersion. In this case,

$$\begin{aligned} & f \propto \exp\left(-\frac{v^2}{2\sigma_{\mathrm{DM}}^2}\right) \to 1, \quad \text{for} \quad v_{\mathrm{gal}} \ll \sigma_{\mathrm{DM}} \\ \Longrightarrow \quad & \rho(< v_{\mathrm{gal}}) = 4\pi m_{\mathrm{DM}} \int_0^{v_{\mathrm{gal}}} f(r, v) v^2 \mathrm{d}v \propto \frac{v_{\mathrm{gal}}^3}{\sigma_{\mathrm{DM}}^3}, \end{aligned} \tag{4.21}$$

     which is a manifestation of Liouville's theorem. Substituting Eq. (4.21) into Eq. (4.20) yields

$$\dot{v}_{\mathrm{gal}} \propto v_{\mathrm{gal}} \quad \text{for} \quad v_{\mathrm{gal}} \ll \sigma_{\mathrm{DM}}, \tag{4.22}$$

     which is also called Stokes friction.
  4. In the opposite limit for $v_{\mathrm{gal}} \gg \sigma_{\mathrm{DM}}$, we have

$$\dot{v}_{\mathrm{gal}} \propto v_{\mathrm{gal}}^{-2} \quad \text{for} \quad v_{\mathrm{gal}} \gg \sigma_{\mathrm{DM}}. \tag{4.23}$$

4.1.3.2 Orbit Decay through Dynamical Friction

- We consider a simple singular isothermal sphere model for the dark matter (see Section 2.5.2),

$$\rho(r) = \frac{\sigma_1^2}{2\pi G r^2}, \quad M(r) = \frac{2\sigma_1^2}{G} r, \tag{4.24}$$

  and adopt a galaxy circular orbit in the cluster for simplicity,

$$\frac{v_c^2}{r} = \frac{GM(r)}{r^2} \quad \Rightarrow \quad v_c = \sqrt{2}\sigma_1 \tag{4.25}$$

  for all radii.

- If we assume that dark matter locally follows a Maxwellian velocity distribution, the phase space distribution is given by

$$f(r, v) = \frac{n(r)}{\left(2\pi\sigma_1^2\right)^{3/2}} \exp\left(-\frac{v^2}{2\sigma_1^2}\right) \tag{4.26}$$

  so that the integral in Eq. (4.19) simplifies for circular galaxy orbits as follows:

$$\begin{aligned} n(r, < v_2) &= 4\pi \int_0^{v_2} f(r, v) v^2 \mathrm{d}v \qquad (4.27) \\ &= n(r) \left[\mathrm{erf}\left(\frac{v_2}{v_c}\right) - \frac{2}{\sqrt{\pi}} \frac{v_2}{v_c} \exp\left(-\frac{v_2^2}{v_c^2}\right)\right] = 0.43\, n(r), \end{aligned}$$

  because on circular orbits in a singular isothermal sphere $v_2 = v_c$.

- The galaxy's binding energy per unit mass is

$$\begin{aligned} \frac{\Phi(r)}{M(r)} &= \frac{G m_{\text{shell}}(r)}{r} = 4\pi G \int_{r_0}^{r} \frac{r'^2 \rho(r') \mathrm{d}r'}{r'} \\ &= 2\sigma_1^2 \int_{r_0}^{r} \frac{\mathrm{d}r'}{r'} = 2\sigma_1^2 \ln \frac{r}{r_0} \qquad (4.28) \end{aligned}$$

  so that the galaxy loses energy per unit mass at a rate

$$\dot{\epsilon} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\Phi(r)}{M(r)}\right) = \frac{2\sigma_1^2}{r} \frac{\mathrm{d}r}{\mathrm{d}t}. \tag{4.29}$$

- Because the galaxy evolves according to dynamical friction, its energy loss per unit mass can also be calculated according to Chandrasekhar's formula,

$$\begin{aligned} \dot{\epsilon} &= \boldsymbol{v}_2 \cdot \tilde{\boldsymbol{F}}_2 = \boldsymbol{v}_2 \cdot \frac{\mathrm{d}\boldsymbol{v}_2}{\mathrm{d}t} = -4\pi G m_2 \ln \Lambda \frac{0.43\sigma_1^2}{2\pi G r^2} \frac{1}{\sqrt{2}\sigma_1} \\ &= -0.61 \frac{G m_2 \sigma_1 \ln \Lambda}{r^2}. \qquad (4.30) \end{aligned}$$

Equating Eqs. (4.29) and (4.30) yields

$$\frac{2\sigma_1^2}{r}\frac{\mathrm{d}r}{\mathrm{d}t} = -0.61\frac{Gm_2\sigma_1 \ln\Lambda}{r^2}, \quad \text{or} \tag{4.31}$$

$$\frac{\mathrm{d}r^2}{\mathrm{d}t} = 2r\frac{\mathrm{d}r}{\mathrm{d}t} = -0.61\frac{Gm_2 \ln\Lambda}{\sigma_1} = -\frac{Gm_2 \ln\Lambda}{1.64\sigma_1 r_0^2}r_0^2 \equiv -\frac{r_0^2}{t_\mathrm{dec}}.$$

Integration gives

$$r^2 = r_0^2\left(1 - \frac{t}{t_\mathrm{dec}}\right), \quad \text{where} \tag{4.32}$$

$$t_\mathrm{dec} = 1.64\,\frac{\sigma_1 r_0^2}{Gm_2 \ln\Lambda} = \frac{0.185}{\ln\Lambda}\frac{M(r_0)}{m_2}\,t_\mathrm{orbit,\,init}, \tag{4.33}$$

and $t_\mathrm{orbit,\,init} = 2\pi r_0/v_2 = \sqrt{2}\pi r_0/\sigma_1$.

- In this simple model, t_dec is the orbit decay time due to dynamical friction. As expected $t_\mathrm{dec} \propto m_2^{-1}$ and larger galaxies migrate faster to the center. To get a numerical value, we set the maximal impact parameter, b_max, equal to the size of the host system, r_{200}, and use that $\bar{v}_\mathrm{encl} \simeq v_\mathrm{c} = \sqrt{GM_{200}/r_{200}}$. The Coulomb logarithm becomes

$$\ln\Lambda \simeq \ln\frac{M_{200}}{m_\mathrm{gal}}, \quad \text{where} \quad m_\mathrm{gal} = m_2. \tag{4.34}$$

- The ratio of virial radius to circular velocity can be written as

$$\frac{r_{200}}{v_\mathrm{c}} = \left(\frac{r_{200}^3}{GM_{200}}\right)^{1/2} = \frac{1}{10H(z)}, \quad \text{where we used} \tag{4.35}$$

$$\frac{M_{200}}{\frac{4}{3}\pi r_{200}^3} = 200\,\frac{3H^2(z)}{8\pi G} \quad \Rightarrow \quad \frac{GM_{200}}{r_{200}^3} = 100H^2(z). \tag{4.36}$$

Hence, we can rewrite the orbit decay time in Eq. (4.33) as

$$t_\mathrm{dec} = \frac{1.17}{\ln(M_{200}/m_\mathrm{gal})}\,\frac{M_{200}}{m_\mathrm{gal}}\,\frac{1}{10H(z)}. \tag{4.37}$$

Thus, the dynamical friction decay time from the edge of a halo to the center is longer than the age of the universe for $M_{200}/m_\mathrm{gal} \gtrsim 15$. Only the most massive subhalos and satellite galaxies are expected to be substantially segregated by mass.

- If a galaxy sinks to the center due to dynamical friction (e.g., because of its large mass or highly eccentric orbit), it will merge with the *cD galaxy* that resides in the center, a processes called “galactic cannibalism”.

4.1.4 Ram Pressure Stripping

- Consider a disk of radius r_d moving through the ICM of mass density ρ_ICM with velocity $\boldsymbol{v}$. For simplicity, we consider a face-on moving disk (i.e., $\boldsymbol{v}$ is parallel to the disk normal). The amount of ICM material swept by the disk is

$$\dot{M}_\mathrm{ICM} = \sigma_\mathrm{d}\rho_\mathrm{ICM}v = \pi r_\mathrm{d}^2\rho_\mathrm{ICM}v, \tag{4.38}$$

  where σ_d is the disk cross section.

- If we assume that the interstellar medium (ISM) of the galaxy stops the hot intracluster wind, then the momentum transferred from the ICM wind to the disk per unit time is

$$\dot{p} = \dot{M}_\mathrm{ICM}v = \pi r_\mathrm{d}^2\rho_\mathrm{ICM}v^2 = \pi r_\mathrm{d}^2 P_\mathrm{ram}, \tag{4.39}$$

  where $P_\mathrm{ram} = \rho_\mathrm{ICM}v^2$ is the ram pressure.

- If this pressure exceeds the force per unit area that binds the interstellar medium to the disk, then the gas gets stripped from the disk. To estimate the binding force, we assume that the mean surface density of interstellar gas is $\Sigma_\mathrm{ISM} \approx M_\mathrm{ISM}/(\pi r_\mathrm{d}^2)$ and that the mean mass density (usually dominated by stars) is $\Sigma_\star \approx M_\star/(\pi r_\mathrm{d}^2)$. The gravitational field of the disk (in the disk) is

$$\Phi_\mathrm{grav} \approx 2\pi G\Sigma_\star. \tag{4.40}$$

  Hence, the gravitational force per unit area on the interstellar gas is

$$\frac{F_\mathrm{grav}}{A} \approx 2\pi G\Sigma_\star\Sigma_\mathrm{ISM}. \tag{4.41}$$

- Gas stripping occurs if $P_\mathrm{ram} > F_\mathrm{grav}/A$, or

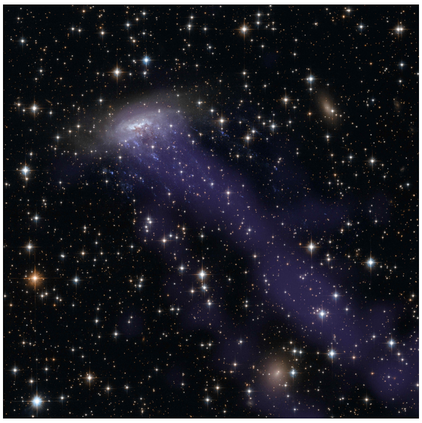
$$\rho_\mathrm{ICM} > \frac{2\pi G\Sigma_\star\Sigma_\mathrm{ISM}}{v^2}. \tag{4.42}$$

- Consider a Milky Way-like disk with

$$M_\star \sim 5\times10^{10}\,\mathrm{M}_\odot, \quad M_\mathrm{ISM} \sim 5\times10^{9}\,\mathrm{M}_\odot, \quad r = 10\ \mathrm{kpc}, \tag{4.43}$$

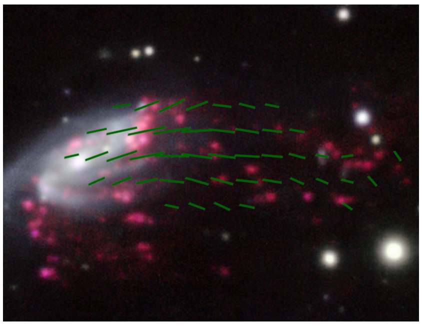
  which is moving with $v = 1000\ \mathrm{km\ s^{-1}}$. Equation (4.42) gives $\rho_\mathrm{ICM} > 4.6\times10^{-27}\ \mathrm{g\ cm^{-3}}$ for stripping to occur, or $n \gtrsim 4.6\times10^{-3}\ \mathrm{cm^{-3}}$ which compares well to ICM densities towards the cluster core.

Composite image showing the two ram pressure stripping galaxies NGC 4522 and NGC 4402 (optical B in blue, optical V in green, infrared I in red, NASA/ESA).



The jellyfish galaxy ESO 137-001 experiences ram pressure stripping and forms a multi-phase gaseous tail with the *Chandra* X-ray emission in blue (NASA, ESA, CXC).

- In general, $P_{\rm ram}$ will change as a function of time due to the eccentric orbit of a galaxy (causing v and $\rho_{\rm ICM}$ to change). Hence, Eq. (4.42) depends on time and eccentricity. Since $\Sigma_\star$ and $\Sigma_{\rm ISM}$ decrease to the disk periphery, there will be critical radius beyond which ram pressure stripping will be effective. The stripped gas can form spectacular tails (see image on the right) which gives rise to jellyfish galaxies. Draping of intracluster magnetic field forms a protective layer around the galaxy and its tail (see Section 4.3.5) that shields the cold stripped interstellar medium from the hostile hot ICM wind and enables ongoing star formation in the stripped tails.



The galaxy JO206 experiences ram pressure stripping. The ordered magnetic field (green lines) along its gas tail protects molecular clouds and enables star formation as can be inferred from the Hα emission (red clumps, Müller et al. 2021)

- When spirals loose their gas, the potential for future star formation is reduced. Ram-pressure stripping is considered important to explain the quenching of star formation in clusters.

- If only parts of the outer disk gets stripped, star formation can continue until it runs out of its fuel after about a few Gyrs. This is called "strangulation" of a galaxy as opposed to the abrupt quenching due to ram-pressure stripping.

4.1.5 Virial Theorem

4.1.5.1 Derivation

- To derive the virial theorem for orbiting galaxies in a cluster, we start with the Newtonian N-body equations,

$$m_i\ddot{\boldsymbol{r}}_i = -\boldsymbol{\nabla}\sum_{\substack{j=1\\ j\neq i}}^{N}\Phi_{ij}, \qquad i = 1,\dots,N \tag{4.44}$$

  and adopt the Newtonian potential,

$$\Phi_{ij} = -\frac{Gm_im_j}{r_{ij}}, \qquad \text{where} \qquad r_{ij} = |\boldsymbol{r}_j - \boldsymbol{r}_i| \tag{4.45}$$

  is the radius vector from galaxy i to galaxy j.

- We proceed by dotting the radius vector of particle i from the left and subsequently summing over all particles,

$$\sum_i \boldsymbol{r}_i\cdot\Big| \qquad m_i\ddot{\boldsymbol{r}}_i = -\boldsymbol{\nabla}\sum_{\substack{j=1\\ j\neq i}}^{N}\Phi_{ij}, \tag{4.46}$$

$$\sum_{i=1}^{N} m_i\boldsymbol{r}_i\cdot\ddot{\boldsymbol{r}}_i = \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{Gm_im_j\boldsymbol{r}_i\cdot(\boldsymbol{r}_j-\boldsymbol{r}_i)}{|\boldsymbol{r}_j-\boldsymbol{r}_i|^3}. \tag{4.47}$$

- Switching the subscripts i and j yields

$$\sum_{j=1}^{N} m_j \boldsymbol{r}_j \cdot \ddot{\boldsymbol{r}}_j = \sum_{i=1}^{N} \sum_{\substack{j=1\\ j\neq i}}^{N} \frac{G m_j m_i \boldsymbol{r}_j \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j)}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^3}. \tag{4.48}$$

Adding Eqs. (4.47) and (4.48), we get

$$2\sum_{i=1}^{N} m_i \boldsymbol{r}_i \cdot \ddot{\boldsymbol{r}}_i = \sum_{i=1}^{N} \sum_{\substack{j=1\\ j\neq i}}^{N} \frac{G m_i m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^3} \underbrace{\left[\boldsymbol{r}_i \cdot (\boldsymbol{r}_j - \boldsymbol{r}_i) + \boldsymbol{r}_j \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j)\right]}_{-(\boldsymbol{r}_i - \boldsymbol{r}_j)^2}. \tag{4.49}$$

Dividing by 2 yields the expression

$$\sum_{i=1}^{N} m_i \boldsymbol{r}_i \cdot \ddot{\boldsymbol{r}}_i = -\frac{1}{2} \sum_{\substack{i,j=1\\ j\neq i}}^{N} \frac{G m_i m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} = -\sum_{\substack{i,j=1\\ i>j}}^{N} \frac{G m_i m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}. \tag{4.50}$$

- After rewriting the left-hand side, we obtain the virial theorem,

$$\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\left(\sum_{i=1}^{N} m_i |\boldsymbol{r}_i|^2\right) - \sum_{i=1}^{N} m_i |\dot{\boldsymbol{r}}_i|^2 = -\sum_{\substack{i,j=1\\ i>j}}^{N} \frac{G m_i m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}, \tag{4.51}$$

$$\frac{1}{2}\ddot{I} = 2E_{\mathrm{kin}} + E_{\mathrm{pot}} = 0 \tag{4.52}$$

in equilibrium. Here, we introduced the moment of inertia, I, the total kinetic energy, E_{kin}, and the total gravitational energy, E_{pot}, via

$$I = \sum_{i=1}^{N} m_i \boldsymbol{r}_i^2, \tag{4.53}$$

$$E_{\mathrm{kin}} = \frac{1}{2}\sum_{i=1}^{N} m_i \dot{\boldsymbol{r}}_i^2, \tag{4.54}$$

$$E_{\mathrm{pot}} = -\sum_{\substack{i,j=1\\ i>j}}^{N} \frac{G m_i m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}. \tag{4.55}$$

- Defining the total mass $M = \sum_i m_i$, we can define the three-dimensional velocity dispersion V, the gravitational radius R_{g}, and the total energy E via

$$V^2 = \frac{2E_{\mathrm{kin}}}{M}, \tag{4.56}$$

$$R_{\mathrm{g}} = -\frac{GM^2}{2E_{\mathrm{pot}}}, \tag{4.57}$$

$$E = E_{\mathrm{kin}} + E_{\mathrm{pot}}. \tag{4.58}$$

This defines then a time scale $t_{\rm dyn} = t_{\rm cross} = 2R_{\rm g}/V$ which is the dynamical or crossing time of the system. A system in virial equilibrium is characterized by

$$2E_{\rm kin} + E_{\rm pot} = 0 \tag{4.59}$$

and determined by E and $E_{\rm pot}$ only.

4.1.5.2 Weighting a Cluster with Galaxies

- How can we weight a galaxy cluster through observations of galaxy spectra? Clearly, by means of the virial theorem, as already demonstrated in Section 1.2.1. But how reliable are those estimates? In the following, we scrutinize the underlying assumptions that (if not fully realized) will bias the mass estimate.

- **Assumption 1:** the system is close to equilibrium and not forming at the present time. In this case, it follows

$$\sum_{i=1}^{N} m_i v_i^2 - \sum_{\substack{i,j=1\\ i>j}}^{N} \frac{G m_i m_j}{r_{ij}} = 0. \tag{4.60}$$

  Usually, we only measure radial velocities along the line of sight, $v_{\rm los}$, and projected separations between galaxy i and j, $r_{\perp,ij} \simeq D\theta_{ij}$, where D is the angular diameter distance and we used the small-angle approximation.

- **Assumption 2:** the system is viewed from a random direction, i.e., the galaxies form a representative sample and are not biased by selection. To test this assumption, we have to derive unbiased estimators for the line-of-sight velocities and projected separation of galaxies. If we define the line-of-sight velocity component of galaxy i via $v_{{\rm los},\,i} = v_i \cos\theta_i$, we get for the kinetic energy that is associated with $v_{{\rm los},\,i}$,

$$\sum_{i=1}^{N} m_i v_i^2 \cos^2\theta_i = \sum_{i=1}^{N} m_i v_{{\rm los},\,i}^2. \tag{4.61}$$

  Taking averages with respect to the random sample on both sides yields

$$\left\langle \sum_{i=1}^{N} m_i v_i^2 \cos^2\theta_i \right\rangle = \left\langle \sum_{i=1}^{N} m_i v_{{\rm los},\,i}^2 \right\rangle. \tag{4.62}$$

Making use of the fact that m_i and v_i and not correlated with θ_i, we obtain

$$\begin{aligned}\left\langle v_{\mathrm{los},\,i}^2\right\rangle &= \left\langle v_i^2\right\rangle\left\langle \cos^2\theta_i\right\rangle && (4.63)\\ &= \left\langle v_i^2\right\rangle \frac{\displaystyle\int_0^{2\pi}\int_0^{\pi/2}\cos^2\theta\sin\theta\mathrm{d}\theta\mathrm{d}\phi}{\displaystyle\int_0^{2\pi}\int_0^{\pi/2}\sin\theta\mathrm{d}\theta\mathrm{d}\phi} && (4.64)\\ &= \left\langle v_i^2\right\rangle \left.\frac{x^3}{3}\right|_0^1 = \frac{\left\langle v_i^2\right\rangle}{3}, && (4.65)\end{aligned}$$

where we substituted $x = \cos\theta$. Hence, an unbiased velocity estimator yields

$$\left\langle v_i^2\right\rangle = 3\left\langle v_{\mathrm{los},\,i}^2\right\rangle \qquad \text{with} \qquad 2E_{\mathrm{kin}} = 3\left\langle \sum_{i=1}^{N} m_i v_{\mathrm{los},\,i}^2\right\rangle. \qquad (4.66)$$

- To derive an unbiased estimator for the angular separation of galaxies, we define the projected physical separation of two galaxies i and j on the sky, $r_{\perp,\,ij} = v_i\cos\theta_i$.

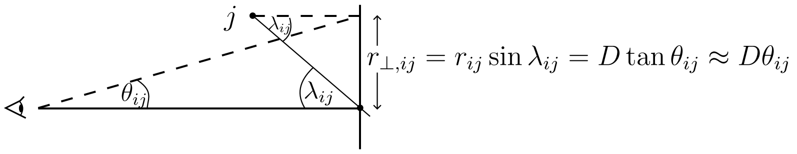


Figure 4.7: Defining angles for projected projected separation of two galaxies i and j on the sky.

- We can write down the following identity

$$\sum_{\substack{i,j=1\\ i>j}}^{N}\frac{Gm_im_j}{r_{ij}\sin\lambda_{ij}} \equiv \sum_{\substack{i,j=1\\ i>j}}^{N}\frac{Gm_im_j}{r_{\perp,\,ij}}. \qquad (4.67)$$

Taking again the averages with respect to the random sample yields

$$-E_{\mathrm{pot}}\left\langle\frac{1}{\sin\lambda_{ij}}\right\rangle = \left\langle\sum_{\substack{i,j=1\\ i>j}}^{N}\frac{Gm_im_j}{r_{\perp,\,ij}}\right\rangle, \qquad (4.68)$$

$$\left\langle\frac{1}{\sin\lambda_{ij}}\right\rangle = \frac{\displaystyle\int_0^{2\pi}\int_0^{\pi/2}\sin^{-1}\lambda_{ij}\sin\lambda_{ij}\mathrm{d}\lambda_{ij}\mathrm{d}\phi_{ij}}{\displaystyle\int_0^{2\pi}\int_0^{\pi/2}\sin\lambda_{ij}\mathrm{d}\lambda_{ij}\mathrm{d}\phi_{ij}} = \frac{\pi}{2} \qquad (4.69)$$

$$\Longrightarrow \quad -E_{\mathrm{pot}} = \frac{2}{\pi}\left\langle\sum_{\substack{i,j=1\\ i>j}}^{N}\frac{Gm_im_j}{r_{\perp,\,ij}}\right\rangle \simeq \frac{2}{\pi}\left\langle\sum_{\substack{i,j=1\\ i>j}}^{N}\frac{Gm_im_j}{D\theta_{ij}}\right\rangle. \qquad (4.70)$$

Hence, this demonstrates that an unbiased estimator for the angular separation of two galaxies i and j is given by

$$\frac{1}{r_{ij}} = \left\langle \frac{2}{\pi} \frac{1}{D\theta_{ij}} \right\rangle . \tag{4.71}$$

- What about fluctuations around this mean separation, i.e., $\langle \theta_{ij}^{-2} \rangle$?

$$\left\langle \frac{1}{\sin^2 \lambda_{ij}} \right\rangle = \frac{\int_0^{\pi/2} \sin^{-2} \lambda_{ij} \sin \lambda_{ij} \mathrm{d}\lambda_{ij}}{\int_0^{\pi/2} \sin \lambda_{ij} \mathrm{d}\lambda_{ij}} = \log \tan \frac{\lambda_{ij}}{2} \Big|_0^{\pi/2} \to \infty . \tag{4.72}$$

This demonstrates that while we have an unbiased estimator for r_{ij}, the uncertainty of this estimator is substantial. This shows that mass estimates based on the virial theorem are subject to large errors.

- **Assumption 3:** all particles are counted and all mass is in observed objects with mass $m_i = (M/L)L_i$, with the average mass-to-light ratio M/L and the individual galaxy luminosities L_i. This implies

  1. no objects are excluded by selection, and
  2. all mass is in galaxies.

For a system of equal mass particles in virial equilibrium, we obtain by combining Eqs. (4.51), (4.66), and (4.71),

$$M = \frac{3 \sum_{i=1}^{N} v_{\mathrm{los},i}^2}{\frac{2}{\pi} \frac{G}{D} \sum_{\substack{i,j=1 \\ i>j}}^{N} \theta_{ij}^{-1}} . \tag{4.73}$$

Hence, the mass-to-light ratio is given by

$$\frac{M}{L} = \frac{3\pi D \sum_{i=1}^{N} L_i v_{\mathrm{los},i}^2}{2G \sum_{\substack{i,j=1 \\ i>j}}^{N} L_i L_j \theta_{ij}^{-1}} . \tag{4.74}$$

Clearly, there are several problems associated with this mass estimation method: 1. luminosity weighting is inefficient, 2. M/L may not be constant, and 3. not all mass may be attached to galaxies.

- **Assumption 4:** positions and velocities of objects are representative:

$$M = \frac{3\pi D \sum\limits_{i=1}^{N} v_{\mathrm{los},\,i}^2}{2G \sum\limits_{\substack{i,j=1\\ i>j}}^{N} \theta_{ij}^{-1}}. \tag{4.75}$$

Unfortunately, the bad statistical behaviour of θ_{ij}^{-1} (with an infinite variance) precludes a reliable mass determination. A solution that circumvents this problem is to obtain an angular size through model fitting, where a specific profile is assumed and the galaxies are takes and tracers that probe the underlying gravitational potential. However, this makes our mass determinations depending on the correctness of the model.

- With the caveats following from these considerations we can weigh a cluster with galaxies to order of magnitude. Adopting a cluster radius $R_{\mathrm{cl}} \sim 1$ Mpc, a galaxy velocity dispersion in a cluster, $v_{\mathrm{cl}} \sim 1000$ km s^{-1}, $N_{\mathrm{gal}} \sim 200$ cluster galaxies with radius $R_{\mathrm{gal}} \sim 3$ kpc and rotation velocities $v_{\mathrm{gal}} \sim 150$ km s^{-1}, we estimate the ratio of the mass in visible galaxies to the total cluster mass,

$$\frac{\sum_i m_{\mathrm{gal},\,i}}{M_{\mathrm{cl}}} \sim \frac{N_{\mathrm{gal}} m_{\mathrm{gal}}}{M_{\mathrm{cl}}} \sim \frac{N_{\mathrm{gal}} R_{\mathrm{gal}} v_{\mathrm{gal}}^2}{R_{\mathrm{cl}} v_{\mathrm{cl}}^2} \sim N_{\mathrm{gal}} \frac{R_{\mathrm{gal}}}{R_{\mathrm{cl}}} \left(\frac{v_{\mathrm{gal}}}{v_{\mathrm{cl}}}\right)^2 \sim 0.0135 \tag{4.76}$$

Hence, to get $\sum_i m_{\mathrm{gal}} \sim M_{\mathrm{cl}}$ you would need $R_{\mathrm{gal}} \sim 0.2 R_{\mathrm{cl}}$. There are several theoretical solutions to this problem:

1. galaxies are much more extended than light,
2. most mass is not attached to galaxies,
3. clusters are very far from equilibrium, or
4. gravity is not Newtonian.

To decide which solution is the right one, we need to turn our attention to other methods for weighting galaxy clusters.

4.2 Gravitational Lensing

The theory presented in this section is based on two main assumptions: (i) the Newtonian limit of a slowly varying gravitational field is taken from Einstein's field equations, namely $|\Phi| \ll c^2$ and $|v_{\rm lens}| \ll c$, in order to characterize the properties of lenses, and (ii) the lensing objects are considered to be thin, i.e. the deflecting mass is isolated and concentrated within a region L much smaller than the distances between source and deflector and deflector and observer, $L \ll cH_0^{-1}$. This approximation holds remarkably well in the astrophysical cases of galaxies or clusters of galaxies.

4.2.1 Deflection Angle

Linearizing the gravitational field equations and taking non–relativistic sources results in the "post–Minkowskian" metric to first order, neglecting the gravitational vector potential,

$$\mathrm{d}s^2 = \left(1 + \frac{2\Phi}{c^2}\right) c^2 \mathrm{d}t^2 - \left(1 - \frac{2\Phi}{c^2}\right) \mathrm{d}\boldsymbol{r}^2 \,, \tag{4.77}$$

where Φ represents the Newtonian potential and $\mathrm{d}\boldsymbol{r}$ characterizes the spatial part of the Minkowski metric. Using the fact that light propagates on null geodesics, namely $\mathrm{d}s^2 = 0$, yields an effective velocity of light c' in the presence of a weak gravitational field,

$$c' = \frac{|\mathrm{d}\boldsymbol{r}|}{\mathrm{d}t} \simeq c\left(1 + \frac{2\Phi}{c^2}\right) \equiv \frac{c}{n} \,, \tag{4.78}$$

$$n = \left(1 - \frac{2\Phi}{c^2}\right) \geq 1 \,, \tag{4.79}$$

which defines an *effective index of refraction n* of the gravitational field in analogy to geometrical optics in dense media. Note that the gravitational potential Φ is by definition negative as it represents an attractive gravitational force. Applying Fermat's principle leads to an equation for the spatial light paths by using the Euler–Lagrange equations for carrying out the variation

$$\delta \int_A^B n \,\mathrm{d}l = \delta \int_A^B n(\boldsymbol{r}) \sqrt{|\dot{\boldsymbol{r}}|^2} \,\mathrm{d}\lambda \overset{!}{=} 0 \tag{4.80}$$

or

$$\ddot{\boldsymbol{r}} = -\frac{2}{c^2} \boldsymbol{\nabla}_{\perp} \Phi(\boldsymbol{r}), \tag{4.81}$$

where the different curves are parametrized by the affine curve parameter λ, the dot denotes a derivative with respect to λ and $\boldsymbol{\nabla}_{\perp}\Phi(\boldsymbol{r})$ is the gradient of the potential perpendicular to the perturbed light ray. The total deflection is therefore the integral along the light path of the differential displacements,

$$\hat{\alpha}(\boldsymbol{r}) = -\int \boldsymbol{\nabla}_\perp n(\boldsymbol{r})\,\mathrm{d}l = \frac{2}{c^2}\int \boldsymbol{\nabla}_\perp \Phi(\boldsymbol{r})\,\mathrm{d}l\,. \tag{4.82}$$

Because in nearly all cases of astrophysical interest the deflection angle is small, $\hat{\alpha} \ll 1$, one usually applies the “Born approximation” and evaluates the integral along the unperturbed ray, i.e. along a straight line. Since the non–relativistic matter is characterized by its density perturbations only, the gravitational potential which gives rise to light deflections in Eq. (4.82) neither depends on the actual nature of matter nor its composition or physical state. Therefore gravitational light deflection probes the total matter density of gravitationally interacting particles irrespective of baryonic and dark matter.

4.2.2 Lens Equation

The lensing equation relates the intrinsic angular source position of an astrophysical object to its observable image position on the sky which was possibly changed in the presence of gravitational light deflection along the line of sight. In order to derive this equation in the thin screen approximation, it is useful first to consider lensing by a point mass. The Newtonian potential as well as its perpendicular gradient can be written as

$$\Phi(\boldsymbol{\xi}, z) = -\frac{GM}{\sqrt{\xi^2 + z^2}} \tag{4.83}$$

$$\text{and} \qquad \boldsymbol{\nabla}_\perp \Phi(\boldsymbol{\xi}, z) = \frac{GM\boldsymbol{\xi}}{(\xi^2 + z^2)^{3/2}}\,, \tag{4.84}$$

where the tree dimensional vector $\boldsymbol{r}$ is decomposed into the z–coordinate along the unperturbed ray and the two dimensional impact parameter $\boldsymbol{\xi}$ orthogonal to the unperturbed ray pointing towards the point mass. Equation (4.82) leads to the deflection angle

$$\hat{\alpha}(\boldsymbol{\xi}) = \frac{2}{c^2}\int_{-\infty}^{\infty} \frac{GM\boldsymbol{\xi}}{(\xi^2 + z^2)^{3/2}}\,\mathrm{d}z = \frac{4GM}{c^2\xi}\frac{\boldsymbol{\xi}}{|\boldsymbol{\xi}|} = \frac{2R_S}{\xi}\frac{\boldsymbol{\xi}}{|\boldsymbol{\xi}|}\,, \tag{4.85}$$

with R_S being the Schwarzschild radius of the point mass. The Born approximation in this context makes sure that the integral is evaluated along the straight coordinate line z.

If we now consider extended objects acting as lenses, but still located within a small region compared to the total distance between lens and observer, the mass distribution of the lensing object can be projected along the line of sight. The smooth three–dimensional distribution can then be replaced by a mass layer perpendicular to the line of sight, which is called *lens plane*. The surface mass density on the lens plane is given by

$$\Sigma(\boldsymbol{\xi}) = \int \rho(\boldsymbol{\xi}, z)\,\mathrm{d}z\,, \tag{4.86}$$

and the deflection angle at position $\boldsymbol{\xi}$ is the overall deflection effect due to a superposition of "point–mass" elements in the plane because of linearity of the system:

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2}\int \frac{\Sigma(\boldsymbol{\xi}')(\boldsymbol{\xi}-\boldsymbol{\xi}')}{|\boldsymbol{\xi}-\boldsymbol{\xi}'|^2}\,\mathrm{d}^2\xi'\,. \tag{4.87}$$

This equation holds in the lens plane with the impact parameter measured in physical units. Assuming the small angle approximation, the *lens equation* relates the position of the source to the observable image position on the sky. The geometry of a typical gravitational lens system is shown in Fig. 4.8.

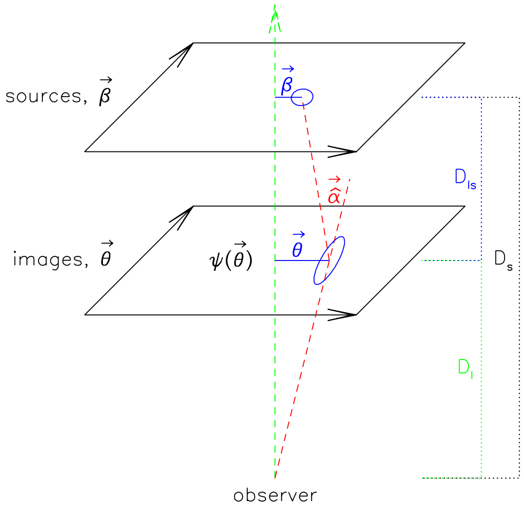


Figure 4.8: Illustration of a typical gravitational lens system. The angles are exaggerated for visualization purposes.

The true position of the source with respect to some arbitrarily chosen optical axis is denoted by $\boldsymbol{\beta}$ and the angular image position on the sky as viewed by an observer is given by $\boldsymbol{\theta}$. All distances along the line of sight are angular diameter distances, where D_{ls} denotes the distance between lens and source, D_l the distance between lens and observer and D_s the distance between source and observer. Using the relation $\boldsymbol{\xi} \simeq D_l\boldsymbol{\theta}$ and introducing the reduced deflection angle,

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{D_{ls}}{D_s}\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}), \tag{4.88}$$

Eq. (4.87) can be written as

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{4G}{c^2}\frac{D_l D_{ls}}{D_s}\int \frac{\Sigma(\boldsymbol{\theta}')(\boldsymbol{\theta}-\boldsymbol{\theta}')}{|\boldsymbol{\theta}-\boldsymbol{\theta}'|^2}\,\mathrm{d}^2\theta'\,. \tag{4.89}$$

The critical surface mass density Σ_{cr} and the convergence κ are defined by

$$\Sigma_{\mathrm{cr}} \equiv \left(\frac{4\pi G}{c^2}\frac{D_l D_{ls}}{D_s}\right)^{-1} \qquad \text{and} \qquad \kappa \equiv \frac{\Sigma}{\Sigma_{\mathrm{cr}}}\,. \tag{4.90}$$

It is important to note that the distance combination appearing in Eq. (4.89), $D_l D_{ls}/D_s$, acts as a lensing efficiency function. It approaches zero at both the source and the observer and has a maximum in between. Using definitions in Eq. (4.90), the deflection angle as a function of the image position $\boldsymbol{\theta}$ reduces to

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi}\int \kappa(\boldsymbol{\theta}')\frac{\boldsymbol{\theta}-\boldsymbol{\theta}'}{|\boldsymbol{\theta}-\boldsymbol{\theta}'|^2}\,\mathrm{d}^2\theta'\,. \tag{4.91}$$

This equation shows that only the ration of Σ and Σ_{cr} is measurable, or in other words, using gravitational lensing on its own, one is not able to determine both the mass of a lensing object and the involved distances independently. From Fig. 4.8 we can read off $\boldsymbol{\theta} D_s - \hat{\boldsymbol{\alpha}} D_{ls} = \boldsymbol{\beta} D_s$, assuming the small angle approximation and using the theorem on intersecting lines. Using the expression for the reduced deflection angle, this establishes the *lens equation* in its simplest form

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta})\,. \tag{4.92}$$

In general, this equation is nonlinear and can thus yield multiple images on the sky for a single source position $\boldsymbol{\beta}$. Moreover, the shape and the size of the images will differ from the original source because light bundles are deflected differentially.

4.2.3 Circular Symmetric Lenses – Einstein Radius

Consider a circularly symmetric lens with an arbitrary mass profile. Due to the high degree of symmetry, we can place the coordinate origin at the center of symmetry and reduce light deflection to a one-dimensional problem. The deflection angle always points toward the center of symmetry with a modulus

$$\hat{\alpha}(\xi) = \frac{4GM(\xi)}{c^2\xi}, \tag{4.93}$$

where $\xi = D_l\theta$ is the distance from the lens center and $M(\xi)$ is the enclosed mass within ξ,

$$M(\xi) = 2\pi\int_0^{\xi}\Sigma(\xi')\xi'\,\mathrm{d}\xi'. \tag{4.94}$$

Combining Eqs. (4.88) and (4.93) enables us to rewrite the lensing equation (4.92), yielding

$$\beta(\theta) = \theta - \frac{D_{ls}}{D_l D_s}\frac{4GM(\theta)}{c^2\theta} \tag{4.95}$$

Owing to the rotational symmetry of the lens, a source, which lies exactly on the optical axis ($\beta = 0$) is imaged as a ring if the lens is supercritical ($\Sigma > \Sigma_{\mathrm{cr}}$). Setting $\beta = 0$ in Eq. (4.95) gives the radius of the ring, the so-called *Einstein radius*,

$$\theta_{\mathrm{E}} = \sqrt{\frac{4GM(\theta)}{c^2} \frac{D_{ls}}{D_l D_s}}. \tag{4.96}$$

It is not only a property of the lens, but also depends on the distance efficiency function. It provides the natural angular scale to describe the lensing geometry for the following reasons: (i) in the case of multiple imaging, the angular separation of images is of order $2\theta_{\mathrm{E}}$, (ii) sources that lie closer than approximately θ_{E} to the optical axis experience strong lensing yielding to strong magnification and sheared images whereas sources at much larger distances are only very little magnified, and (iii) in many lens models the Einstein radius roughly represents the boundary of source positions that are either multiply-imaged if they lie inside θ_{E} or singly-imaged. Comparing Eqs. (4.90) and (4.96) reveals that the surface mass density inside θ_{E} is exactly the critical surface density Σ_{cr}. For a point mass, we can obtain the Einstein radius

$$\theta_{\mathrm{E}} = \sqrt{\frac{4GM}{c^2} \frac{D_{ls}}{D_l D_s}} \approx 30'' \left(\frac{M}{10^{14} M_\odot}\right)^{1/2} \left(\frac{D}{\mathrm{Gpc}}\right)^{-1/2}, \tag{4.97}$$

where we defined the lensing efficiency distance $D = D_l D_s / D_{ls}$ and inserted typical values for clusters to highlight the relevant angular scales for giant (tangential) arcs in clusters. In the case of clusters, detailed mass modeling is necessary since only a fraction of the cluster mass resides within the Einstein radius.

For a point mass lens, we can use the Einstein radius of Eq. (4.96) to rewrite the lens equation in the form

$$\beta = \theta - \frac{\theta_{\mathrm{E}}^2}{\theta}. \tag{4.98}$$

This equation has two solutions

$$\theta_\pm = \frac{1}{2}\left(\beta \pm \sqrt{\beta^2 + 4\theta_{\mathrm{E}}^2}\right). \tag{4.99}$$

Any source inside θ_{E} is imaged twice by a point mass lens. The two images are on either side of the source with one image inside the Einstein ring and the other one outside. As the source moves away from the center of the lens (i.e., with increasing β), one of the images approaches the lens and becomes very faint, while the other image approaches the true position of the source and tends toward a magnification of order unity.

4.2.4 The Lensing Potential and Local Lens Properties

It is convenient to define the lensing potential $\Psi(\boldsymbol{\theta})$ which is the scaled and projected Newtonian potential of the lens,

$$\Psi(\boldsymbol{\theta}) = \frac{D_{ls}}{D_l D_s} \frac{2}{c^2} \int \Phi(D_l \boldsymbol{\theta}, z) \mathrm{d}z \,. \tag{4.100}$$

The lensing potential has the nice property that its gradient with respect to $\boldsymbol{\theta}$ is the deflection angle

$$\boldsymbol{\nabla}_\theta \Psi(\boldsymbol{\theta}) = \frac{D_{ls}}{D_s} \frac{2}{c^2} \int \boldsymbol{\nabla}_\perp \Phi(\boldsymbol{\xi}, z) \mathrm{d}z \; = \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{4.101}$$

where the perpendicular gradient is now acting on the physical impact parameter having used the small angle approximation $\boldsymbol{\xi} \simeq D_l \boldsymbol{\theta}$. Assuming further that the changes of the Newtonian potential along the line of sight average out, which is true for instance, as long as the lensing object is only slowly varying and does not undergo a rapid collapse. More precisely, the time–scale on which light travels across the lensing object, has to be much smaller than the collapse time–scale of the light deflecting object. Then the two–dimensional Laplacian can be replaced by its three–dimensional analogue,

$$\Delta^{(2)} \Phi(\boldsymbol{r}) = \sum_{i=1}^{2} \frac{\partial^2 \Phi(\boldsymbol{r})}{\partial \xi_i^2} \simeq \sum_{i=1}^{3} \frac{\partial^2 \Phi(\boldsymbol{r})}{\partial r_i^2} = \Delta^{(3)} \Phi(\boldsymbol{r}) \,. \tag{4.102}$$

Therefore, the Laplacian of the lensing potential acting on its angular coordinate $\boldsymbol{\theta}$ equals twice the surface mass density scaled with its critical value, i.e. the convergence κ,

$$\Delta_\theta^{(2)} \Psi(\boldsymbol{\theta}) = \frac{2}{c^2} \frac{D_l D_{ls}}{D_s} \int \Delta^{(3)} \Phi(\boldsymbol{\xi}, z) \mathrm{d}z = 2 \frac{4\pi G}{c^2} \frac{D_l D_{ls}}{D_s} \int \rho(\boldsymbol{\xi}, z) \mathrm{d}z = 2\kappa(\boldsymbol{\theta}) \,, \tag{4.103}$$

where Poisson's equation has been used in the second step. Since Ψ satisfies the two–dimensional Poisson's equation, its Green's function has to be considered, namely

$$\Delta^{(2)} G(\boldsymbol{\theta}, \boldsymbol{\theta}') = 2\pi \delta_D(\boldsymbol{\theta}, \boldsymbol{\theta}') \qquad \Longrightarrow \qquad G(\boldsymbol{\theta}, \boldsymbol{\theta}') = \ln |\boldsymbol{\theta} - \boldsymbol{\theta}'| \,. \tag{4.104}$$

Therefore the lensing potential $\Psi(\boldsymbol{\theta})$ is given by the convolution integral of the source function $\kappa(\boldsymbol{\theta})$ and the Green's function in two dimensions,

$$\Psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int \kappa(\boldsymbol{\theta}') \ln |\boldsymbol{\theta} - \boldsymbol{\theta}'| \mathrm{d}^2 \theta' \,. \tag{4.105}$$

Liouville's theorem and the conservation of the physical number density of photons during the process of gravitational light bending imply that lensing conserves surface brightness or specific intensity. Assuming that the angular scale on which the lens properties change is much larger than the extent of the source, the lens equation can locally be linearized yielding

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}) \simeq \boldsymbol{\beta}_0 + \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) . \tag{4.106}$$

The local lens properties of the lens mapping are described by its Jacobian matrix $\mathcal{A}$

$$\mathcal{A} \equiv \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \left(\delta_{ij} - \frac{\partial \alpha_i(\boldsymbol{\theta})}{\partial \theta_j} \right) = \left(\delta_{ij} - \frac{\partial^2 \Psi(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right) \equiv \left(\delta_{ij} - \Psi,_{ij}(\boldsymbol{\theta}) \right) = \mathcal{M}^{-1} , \tag{4.107}$$

where an abbreviation for partial derivatives has been introduced and $\mathcal{A}$ is the inverse of the magnification tensor $\mathcal{M}$. This is justified, because a solid–angle element $\delta\beta^2$ of the source is mapped onto the solid–angle element $\delta\theta^2$ on the image, and thus the magnification due to the mapping is given by

$$\frac{\delta\theta^2}{\delta\beta^2} = \det \mathcal{M} = \frac{1}{\det \mathcal{A}} . \tag{4.108}$$

The trace of the Jacobian $\mathcal{A}$ describes the isotropic magnification of the source,

$$\mathrm{tr}(\mathcal{A}) = (1 - \Psi,_{11}) + (1 - \Psi,_{22}) = 2(1 - \kappa) . \tag{4.109}$$

This also intuitively explains the meaning of the convergence κ, which is a measure for how much the lens focuses light rays isotropically. Subtracting the trace from $\mathcal{A}$ leads to an expression for anisotropic distortion (astigmatism) of the image,

$$\mathcal{A}_{ij} - \frac{1}{2}\delta_{ij}\,\mathrm{tr}(\mathcal{A}) = \delta_{ij} - \Psi,_{ij} - \delta_{ij}(1 - \kappa) = -\Psi,_{ij} + \kappa\delta_{ij} \equiv \Gamma , \tag{4.110}$$

where the *shear tensor* Γ has been defined in the last step. This distortion is due to the tidal gravitational field. Particularly, it decomposes in

$$\Gamma = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} \tag{4.111}$$

and

$$\gamma_1 = \frac{1}{2}(\Psi,_{11} - \Psi,_{22}) \equiv \gamma(\boldsymbol{\theta}) \cos(2\phi(\boldsymbol{\theta})) \tag{4.112}$$

$$\gamma_2 = \Psi,_{12} = \Psi,_{21} \equiv \gamma(\boldsymbol{\theta}) \sin(2\phi(\boldsymbol{\theta})) . \tag{4.113}$$

Here $\gamma = \sqrt{\gamma_1^2 + \gamma_2^2}$ describes the magnitude of the shear and ϕ its orientation, whereas the factor 2 shows that γ is not a vector, but a 2 × 2–tensor.

4.2.5 Strong and Weak Cluster Lensing

The name of the game in cluster lensing research consists in reconstructing the mass distribution. Depending on the type of lensing – strong or weak lensing – there are different algorithms used.

4.2.5.1 Strong Cluster Lensing

Using our simplified mass density model of a singular isothermal sphere (SIS, see Section 2.4.1), we can readily work out the relevant strong lensing properties for this model. Recall that the mass density and rotational velocity in this model was given by

$$\rho(r) = \frac{\sigma_v^2}{2\pi G}\frac{1}{r^2} \quad \text{and} \quad v_{\rm rot}^2 = \frac{GM(r)}{r} = 2\sigma_v^2 = \text{const.} \tag{4.114}$$

Upon projection the density along the line-of-sight, we obtain the surface mass density

$$\Sigma(\xi) = \int_0^s \rho(\boldsymbol{\xi}, z)\mathrm{d}z = 2\int_\xi^\infty \frac{\rho(r)r\mathrm{d}r}{\sqrt{r^2 - \xi^2}} = \frac{\sigma_v^2}{2G}\frac{1}{\xi}, \tag{4.115}$$

where ξ is the distance from the center of the two-dimensional profile. Using Eq. (4.93), we obtain the deflection angle

$$\hat{\alpha} = 4\pi\frac{\sigma_v^2}{c^2} \tag{4.116}$$

which is independent of ξ and points to the center of the lens. The Einstein radius of the SIS is given by Eq. (4.96),

$$\theta_{\rm E} = 4\pi\frac{\sigma_v^2}{c^2}\frac{D_{ls}}{D_s} = \hat{\alpha}\frac{D_{ls}}{D_s} = \alpha. \tag{4.117}$$

The symmetry of the problem reduces the dimensionality of the problem to become one-dimensional. Multiple images are only obtained if the source lies inside the Einstein ring, i.e., if $\beta < \theta_{\rm E}$. If this condition is satisfied, we obtain the following two solutions,

$$\theta_\pm = \beta \pm \theta_{\rm E}. \tag{4.118}$$

The images at $\theta_\pm$, the source, and the lens all lie on a straight line. (The third image with zero flux lies at $\theta = 0$ and only acquires a non-zero flux if the singularity of the lens is replaced by a core of finite density).

Rich concentrated clusters can produce giant arcs when a background galaxy is aligned with one of the cluster caustics. Typically, a parametrized lens model (such as the SIS above or a more complicated functional form) is optimized so as to obtain a good fit to the observed image. If there are many constraints from a number of strongly lensed galaxies such as their position and detailed properties of their distortion (magnitude distribution across their arcs), ray tracing through an adaptive grid is possible. This can even constrain the detailed mass distribution within the cluster including their substructure mass distribution.

4.2.5.2 Weak Cluster Lensing – The Kaiser & Squires Algorithm

Every cluster weakly distorts images of a large number of background galaxies giving rise to so-called arclets – this phenomenon is referred to as weak lensing. With the development of the Kaiser & Squires (1993) algorithm, weak lensing is being used increasingly to derive parameter-free two-dimensional mass maps of galaxy clusters.

4.3 X-ray Cluster Astrophysics

4.3.1 Hydrostatic Equilibrium Masses and Biases

- Unless the intracluster gas gets continuously disturbed by (major) mergers, we expect the ICM to relax on a few sound crossing time scales, t_s, of the clusters, where

  $$t_\mathrm{s} \equiv \frac{D}{c_\mathrm{s}} \approx 7 \times 10^8 \left(\frac{T}{10^8\,\mathrm{K}}\right)^{-1/2} \left(\frac{D}{1\,\mathrm{Mpc}}\right) \mathrm{yr}. \tag{4.119}$$

  Because t_s is shorter than the age of a typical cluster, which is a fraction of the Hubble time, the gas should be close to hydrostatic equilibrium. However, because clusters are forming today, there will always be a fraction of clusters that have experienced a recent merger which causes them to be out of hydrostatic equilibrium. Moreover, if AGN jets have recently injected energy into the central regions, those should also be out of equilibrium. Hence, the hydrostatic equilibrium assumption has to be used with great care.

- Assuming hydrostatic equilibrium, the pressure force of the gas balances gravity,

  $$\boldsymbol{\nabla} P = -\rho_\mathrm{gas} \boldsymbol{\nabla} \Phi \qquad \text{so that} \tag{4.120}$$

  $$\frac{1}{\rho_\mathrm{gas}} \frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{GM(r)}{r^2} \qquad \text{for spherical symmetry,} \tag{4.121}$$

  where $M(r)$ is the total enclosed mass at radius r. Because this is a single equation for ρ_gas and P, we must specify an equation of state to close this equation. To this end, we assume an ideal gas with $P = \rho_\mathrm{gas} k_\mathrm{B} T / \bar{m}$, where $\bar{m}$ is the mean particle mass (i.e., $\bar{m} = \mu m_\mathrm{p}$ with $\mu = 0.588$ for the intracluster plasma) and obtain

  $$\frac{k_\mathrm{B} T}{\bar{m}} \frac{\mathrm{d}\rho_\mathrm{gas}}{\mathrm{d}r} + \frac{\rho_\mathrm{gas} k_\mathrm{B}}{\bar{m}} \frac{\mathrm{d}T}{\mathrm{d}r} = -\frac{GM\rho_\mathrm{gas}}{r^2}. \tag{4.122}$$

- The gas temperature defines the gas velocity dispersion along one coordinate direction (e.g., the line of sight), $\sigma_{v,\mathrm{gas}}$, via

  $$\frac{3\bar{m}}{2} \sigma_{v,\mathrm{gas}}^2 = \frac{3}{2} k_\mathrm{B} T \quad \Longrightarrow \quad \bar{m}\sigma_{v,\mathrm{gas}}^2 = k_\mathrm{B} T. \tag{4.123}$$

  Hence, Eq. (4.122) becomes

  $$M(r) = -\frac{r k_\mathrm{B} T}{G\bar{m}} \left(\frac{\mathrm{d}\ln\rho_\mathrm{gas}}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln T}{\mathrm{d}\ln r}\right) \tag{4.124}$$

  $$= -\frac{r\sigma_{v,\mathrm{gas}}^2}{G} \left(\frac{\mathrm{d}\ln\rho_\mathrm{gas}}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln \sigma_{v,\mathrm{gas}}^2}{\mathrm{d}\ln r}\right). \tag{4.125}$$

- Collisions between gas particles lead to thermal equilibrium and to an isotropization of their velocity distribution. Galaxies represent a collisionless population and hence, do not equilibrate via collisions. Instead, collisionless processes during the formation of a galaxy cluster such as violent relaxation (Lynden-Bell 1967) cause a redistribution of the kinetic energy of galaxies, which is "forced" by temporal changes in the gravitational potential. It is believed that this leads to an isotropic velocity distribution of galaxies that we assume to be realized here.

- Galaxies with identical velocities share the same orbits as gas particles. Hence, galaxies should obey the same equation (4.125) for the cluster mass as gas particles do,

$$M(r) = -\frac{r\sigma^2_{v,\mathrm{gal}}}{G}\left(\frac{\mathrm{d}\ln\rho_{\mathrm{gal}}}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln\sigma^2_{v,\mathrm{gal}}}{\mathrm{d}\ln r}\right). \tag{4.126}$$

- Combining both mass estimates of Eqs. (4.124) and (4.126) yields

$$\sigma^2_{v,\mathrm{gal}}\left(\frac{\mathrm{d}\ln\rho_{\mathrm{gal}}}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln\sigma^2_{v,\mathrm{gal}}}{\mathrm{d}\ln r}\right) = \frac{k_{\mathrm{B}}T}{\bar{m}}\left(\frac{\mathrm{d}\ln\rho_{\mathrm{gas}}}{\mathrm{d}\ln r} + \frac{\mathrm{d}\ln T}{\mathrm{d}\ln r}\right). \tag{4.127}$$

- In the following, we assume that galaxies have a comparable velocity dispersion as the gas,

$$\sigma^2_{v,\mathrm{gal}} = \sigma^2_{v,\mathrm{gas}}\beta, \tag{4.128}$$

  where $\beta \sim 1$ because they probe the same external gravitational potential provided by dark matter.

- Introducing the ratio of specific energies

$$\beta \equiv \frac{\bar{m}\sigma^2_{v,\mathrm{gal}}}{k_{\mathrm{B}}T} \quad \Longrightarrow \quad \mathrm{d}\ln\beta = \mathrm{d}\ln\sigma^2_{v,\mathrm{gal}} - \mathrm{d}\ln T \tag{4.129}$$

  enables us to rewrite Eq. (4.127) and leads to the expression

$$\mathrm{d}\ln\rho_{\mathrm{gas}} = \beta\left(\mathrm{d}\ln\rho_{\mathrm{gal}} + \mathrm{d}\ln\sigma^2_{v,\mathrm{gal}}\right) - \mathrm{d}\ln T. \tag{4.130}$$

- Combining Eqs. (4.129) and (4.130) to eliminate σ_v^2 yields

$$\mathrm{d}\ln\rho_{\mathrm{gas}} = \beta\mathrm{d}\ln\rho_{\mathrm{gal}} + (\beta-1)\mathrm{d}\ln T + \mathrm{d}\beta \tag{4.131}$$

  or equivalently

$$\rho_{\mathrm{gas}} \propto \rho^{\beta}_{\mathrm{gal}}T^{\beta-1} \quad \text{for} \quad \beta = \text{const.} \tag{4.132}$$

- Because the galaxy distribution follows a King profile,

$$\rho_{\mathrm{gal}}(r) = \rho_0 \left[1 + \left(\frac{r}{r_{\mathrm{c}}}\right)^2\right]^{-3/2}, \tag{4.133}$$

  we obtain a beta profile for an isothermal gas according to Eq. (4.132),

$$\rho_{\mathrm{gas}}(r) = \rho_0 \left[1 + \left(\frac{r}{r_{\mathrm{c}}}\right)^2\right]^{-3\beta/2}. \tag{4.134}$$

- Because the X-ray emissivity $j_{\mathrm{X}} \propto \rho_{\mathrm{gas}}^2$, we obtain

$$j_{\mathrm{X}}(r) = j_0 \left[1 + \left(\frac{r}{r_{\mathrm{c}}}\right)^2\right]^{-3\beta}, \tag{4.135}$$

  which yields the X-ray surface brightness profile upon a line-of-sight integration

$$S_{\mathrm{X}}(r_\perp) = \int_{-\infty}^{\infty} j_{\mathrm{X}}[r(z)]\mathrm{d}z \tag{4.136}$$

$$= 2\int_{r_\perp}^{\infty} \frac{j_{\mathrm{X}}(r) r\,\mathrm{d}r}{\sqrt{r^2 - r_\perp^2}} = S_0 \left[1 + \left(\frac{r_\perp}{r_{\mathrm{c}}}\right)^2\right]^{-3\beta+1/2}, \tag{4.137}$$

  where $S_0 \propto j_0$ is a constant. This functional form provides excellent fits to the X-ray surface brightness maps of observed clusters with $\beta \approx 2/3 \ldots 1$, where the larger values correspond to deeper X-ray observations, which reach out to larger cluster-centric radii.

- To obtain the mass profile $M(r)$ we calculate

$$\frac{\mathrm{d}\ln\rho_{\mathrm{gas}}}{\mathrm{d}\ln r} = -\frac{3\beta r}{2}\frac{\mathrm{d}}{\mathrm{d}r}\ln\left[\rho_0\left(1 + \frac{r^2}{r_{\mathrm{c}}^2}\right)\right] \tag{4.138}$$

$$= -3\beta \frac{r^2/r_{\mathrm{c}}^2}{1 + r^2/r_{\mathrm{c}}^2}. \tag{4.139}$$

  For an isothermal cluster, we obtain from Eq. (4.124)

$$M(r) = \frac{3\beta r k_{\mathrm{B}} T}{G\bar{m}} \frac{r^2/r_{\mathrm{c}}^2}{1 + r^2/r_{\mathrm{c}}^2} \xrightarrow{r \gg r_{\mathrm{c}}} \frac{3\beta k_{\mathrm{B}} T}{G\bar{m}} r. \tag{4.140}$$

  Assuming isothermal gas in hydrostatic equilibrium and spherical symmetry yields X-ray surface brightness profiles that are excellently fit with a beta profile. However, the resulting mass profiles are wrong at large radii because the mass of an NFW dark matter profile scales as $M(r) \propto \ln r$ for large radii. The reasons for this apparent inconsistency are the following:

  - Many of the simplifying assumptions (isothermal gas, hydrostatic equilibrium and spherical symmetry) break down in the cluster outskirts.

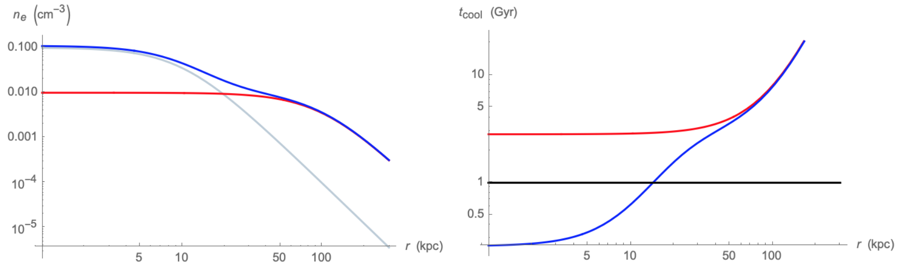


Figure 4.9: Left: the density distribution of a *cool core cluster* (blue, double beta profile) in comparison to a *non-cool-core cluster* (red, single beta profile). Right: the corresponding cooling time profiles. Note that cool core clusters are characterized by central cooling times with $t_\mathrm{cool} < 1$ Gyr.

  - Shallow X-ray observations are mostly sensitive to regions around the cluster core radius, which is often close to the cluster scale radius r_s of the dark matter density profile. The dark matter imprints its structure onto the gas and scales at this radius as $\rho_\mathrm{DM} \propto r^{-2}$. Coincidentally, this is the asymptotic scaling of the gas density $\rho_\mathrm{gas} \propto r^{-3\beta} \sim r^{-2}$ for $\beta \sim 2/3$.
  - Deeper X-ray observations enable probing larger radii and find a transition to larger values of $\beta \sim 1$. However, if they reach regions around the virial radius, they also see an increased level of clumping and asphericity. Masking all these clumps yields an even steeper density profile.

- The kinetic-to-thermal pressure contribution substantially increases towards and beyond the virial radius so that hydrostatic cluster masses (that neglect the unobservable kinetic pressure contribution) are biased low.

4.3.2 Cluster Population and Evolution

- The electron density distribution of a *cool-core cluster* is characterized by a double beta profile,

$$n_\mathrm{e}(r) = \sum_{i=1,2} n_i \left[1 + \left(\frac{r}{r_{\mathrm{c},i}}\right)^2\right]^{-3\beta_i/2}. \tag{4.141}$$

  Typical parameters of the density profile are $\beta_{1,2} = 1$, $(n_1, n_2) = (10^{-1}, 10^{-2})\,\mathrm{cm}^{-3}$, and $r_{\mathrm{c},1}, r_{\mathrm{c},2} = (10, 100)$ kpc (blue curve in Fig. 4.9). The central dense core (light blue) is accompanied by a central dip in temperature, that is lower by factor of about three in comparison to the temperature maximum at $0.2R_{200}$.

- A *non cool-core cluster* shows a constant central temperature plateau and no central density enhancement. Its density distribution follows a single beta profile with parameters that are typically given by the outer (red) beta profile of Fig. 4.9.

- These different density profiles imply a qualitatively different cooling time distribution. Recall the definition of the cooling time of Eq. (3.180), which compares the thermal energy content to the total (frequency-integrated) X-ray emissivity,

$$t_{\mathrm{cool}} = \frac{\varepsilon_{\mathrm{th}}}{\dot{\varepsilon}_{\mathrm{brems}}} \approx 2.8 \left(\frac{k_{\mathrm{B}}T}{2\ \mathrm{keV}}\right)^{1/2} \left(\frac{n_{\mathrm{e}}}{10^{-2}\ \mathrm{cm}^{-3}}\right)^{-1} \mathrm{Gyr}. \tag{4.142}$$

  The cooling time profile is obtained by replacing n_{e} with the (double) beta profiles for the gas density. Because t_{cool} depends only weakly on temperature, which only varies slowly with radius, we assume a constant temperature of 2 keV and show the cooling time profiles in Fig. 4.9.

- While $t_{\mathrm{cool}} < 1$ Gyr in the center of *cool core clusters*, the central cooling time of *non-cool-core clusters* typically never reaches values below 3 Gyr. These are typical time scales for mergers or accretion of smaller groups to occur so that the injected turbulence and shocks mix the cooling central gas with the hotter, more dilute gas from the outskirts and increase thereby the central cooling times, implying that non-cool core clusters do not suffer from the cooling flow problem. This consideration is formalized by introducing the cooling radius, which is the radius where the cooling time $t_{\mathrm{cool}} = 1$ Gyr. For the cool-core cluster in Fig. 4.9 we get $r_{\mathrm{cool}} = 14.4$ kpc, while there is no cooling radius in the non-cool core cluster.

- What is the relative fraction of cool core and non-cool core clusters and how does their abundance evolve with redshift? So far, there has not been a volume-limited (all-sky) survey with the necessary resolution to definitively answer these questions. However, this will soon change with the eROSITA survey already taking data. Instead, Cavagnolo et al. (2009) have analyzed 239 clusters of the *Chandra* data archive[1] and find that most ICM entropy profiles are well fitted by a model which is a power law at large radii and approaches a constant value at small radii:

$$K_{\mathrm{e}}(r) = K_0 + K_{100} \left(\frac{r}{100\ \mathrm{kpc}}\right)^{\alpha}, \tag{4.143}$$

  where K_0 quantifies the typical excess of core entropy above the best-fitting power law found at larger radii, see Fig. 4.10.

---

[1] Note that this is by no means a uniform sample but suffers from a selection bias driven by the scientific interest of proposing researchers and the eventually selected cluster targets.

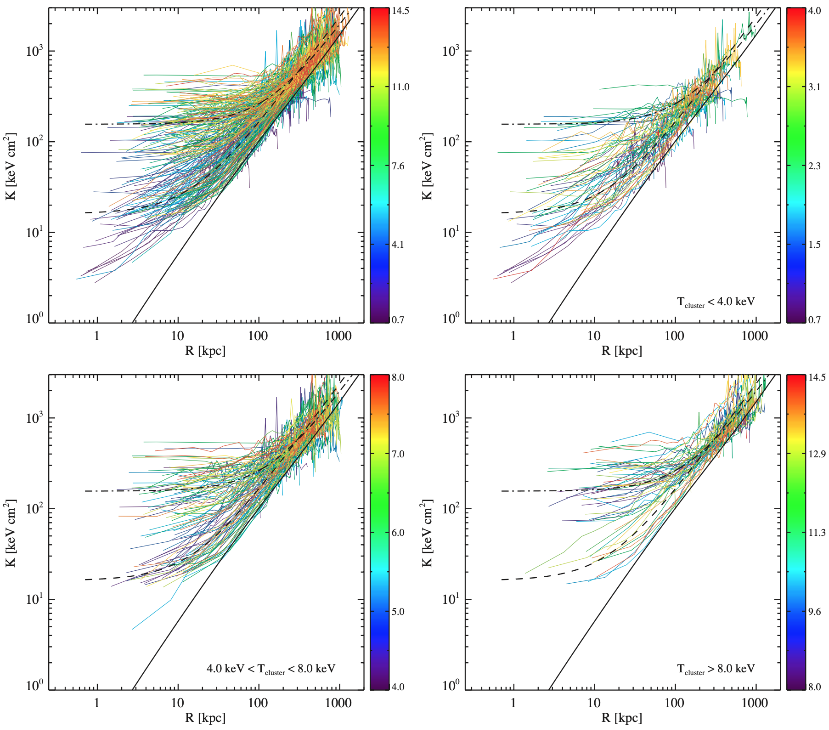


Figure 4.10: Composite plots of entropy profiles for varying cluster temperature ranges. Profiles are color-coded based on average cluster temperature. The solid line is the pure-cooling model of Voit et al. (2002), the dashed line is the mean profile for clusters with $K_0 \leq 50$ keV cm^2, and the dashed-dotted line is the mean profile for clusters with $K_0 > 50$ keV cm^2. Panels from top left to bottom right show all the entropy profiles, and entropy profiles selected by average cluster temperature (indicated in the legends). Note that while the dispersion of core entropy for each temperature range is large, as the $k_B T_X$ range increases so does the mean core entropy (Cavagnolo et al. 2009).)

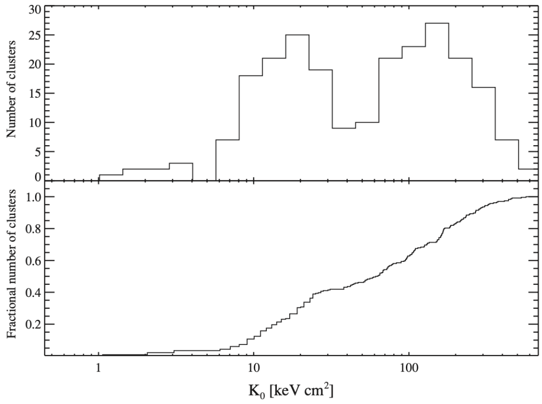


Figure 4.11: Top panel: histogram of the best-fit K_0 for all the clusters in the *Chandra* data base. Bottom panel: cumulative distribution of K_0 values for the full sample. The distinct bimodality in K_0 is present in both distributions, which would not be seen if it were an artifact of the histogram binning. A KMM test finds the K_0 distribution cannot arise from a simple unimodal Gaussian (Cavagnolo et al. 2009).

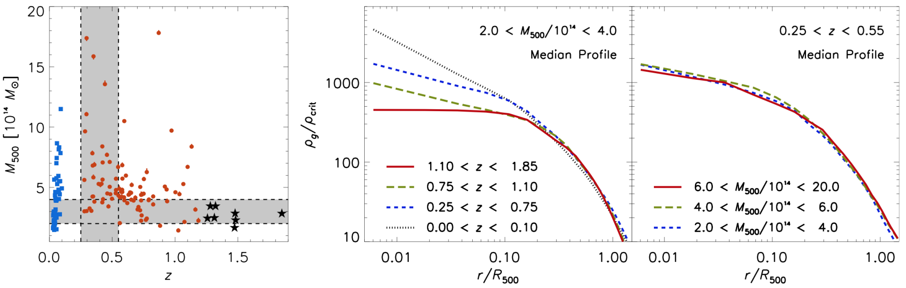


Figure 4.12: Left panel: distribution of galaxy cluster masses and redshifts used in the study of McDonald et al. (2017). Gray shaded regions represent cuts for two subsamples: a large mass range at nearly fixed redshift and a large redshift range at nearly fixed mass. Center panel: median density profiles for clusters over a broad redshift range and narrow mass range showing a pseudo evolution of the scaled density profile. Right panel: median density profiles for clusters over a broad mass range and narrow redshift range. These median profiles are indistinguishable, suggesting that the median cluster profile evolves self-similarly.

- Interestingly, the K_0 distribution of the full archival sample is bimodal with a distinct gap between $K_0 \approx 30$–50 keV cm^2 and the populations peak at $K_0 \sim 15$ keV cm^2 and $K_0 \sim 150$ keV cm^2, see Fig. 4.11. This result is robust to effects of point-spread function smearing and angular resolution. The central entropy and cooling time are connected via Eq. (3.182),

  $$t_{\rm cool} = 1\,{\rm Gyr}\left(\frac{K_{\rm e}}{K_0}\right)^{3/2}\frac{k_{\rm B}T_0}{k_{\rm B}T}, \tag{4.144}$$

  where $K_0 = 21.5$ keV cm^2 and $k_{\rm B}T_0 = 2$ keV. Hence, for a given cluster mass (or average temperature), the core entropy K_0 is a measure of cooling time in the core. The cluster population peaking at $K_0 \sim 15$ keV cm^2 corresponds to cool core clusters and the high-entropy systems are represented by non-cool core clusters.

- Combining *Chandra* X-ray observations of clusters at low redshift $0 < z < 0.1$ with X-ray observations of clusters that are selected from Sunyaev-Zel'dovich observations at higher redshifts $0.25 < z < 1.85$ enables to constrain the evolution of the ICM over the past 10 Gyr. McDonald et al. (2017) find that the *bulk of the ICM* has evolved self-similarly over the full redshift range probed, with the ICM density at $r > 0.2R_{500}$ scaling like $E(z)^2$, which is the same redshift scaling as the critical density of the universe, $\rho_{\rm crit} = 3H(z)^2/(8\pi G) \propto E(z)^2$. Contrarily, the density in the centers of clusters ($r \lesssim 0.01R_{500}$) significantly deviates from self-similarity ($n_{\rm e} \propto E(z)^{0.2\pm0.5}$), consistent with no redshift dependence. This can be shown by isolating clusters with overdense cores (i.e., cool cores). Their average overdensity profile has not evolved with redshift, that is, cool cores have not changed in size,

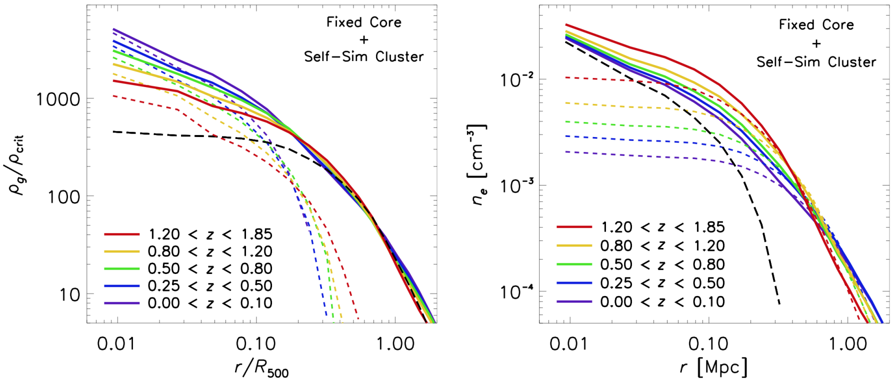


Figure 4.13: Left: expected density profiles (solid lines) for a self-similarly evolving, non-cool core cluster (dashed black line) combined with a non-evolving cool core (dotted colored lines). Because of the choice of scaling, the non-evolving cool core term appears to be evolving. Right: same as left panel but for the profiles in absolute physical units. Without any cosmological scaling, the cool core now appears nearly static, while the bulk of the cluster shows the expected self-similar evolution (McDonald et al. 2017).

density, or total mass over the past ~ 9–10 Gyr. This is demonstrated in Figs. 4.12 and 4.13. In fact, the evolving "cuspiness" of clusters in the X-ray, reported by several previous studies, can be understood in the context of a cool core with fixed properties embedded in a self-similarly evolving cluster.

- It may seem counterintuitive that a collapsed cluster profile evolves with redshift because it has already decoupled from the Hubble expansion and formed a gravitational potential that is much deeper in comparison to the pull of the Hubble expansion (see Section 2.3). This puzzle is resolved by realizing that the background mass density is decreasing due to the Hubble expansion so that we need to increase the radius of the averaging spheres with time to retain a mean density $\bar{\rho} = 200\rho_{\rm cr}$ that characterizes the virial radius; this effect is called "pseudo evolution".

- The findings of McDonald et al. (2017) suggest that the two populations of clusters (cool cores and non-cool cores) are distributed roughly equally. The fact that the cool cores do not statistically evolve with redshift may suggest that once a cluster has formed a cool core, it remains in this state and does not transform to a non-cool core system and vice versa, a non-cool core cluster only evolves self-similarly with redshift but does not evolve into cool core systems.

- This statistical inference is strengthened by correlating the observed core entropy values of clusters, $K_{e,0}$, with the AGN cavity energy $E_{\rm cav} = 4PV$ (for AGN bubbles filled with a relativistic fluid or magnetic fields, i.e., for $\gamma = 4/3$, see Fig. 4.14). This

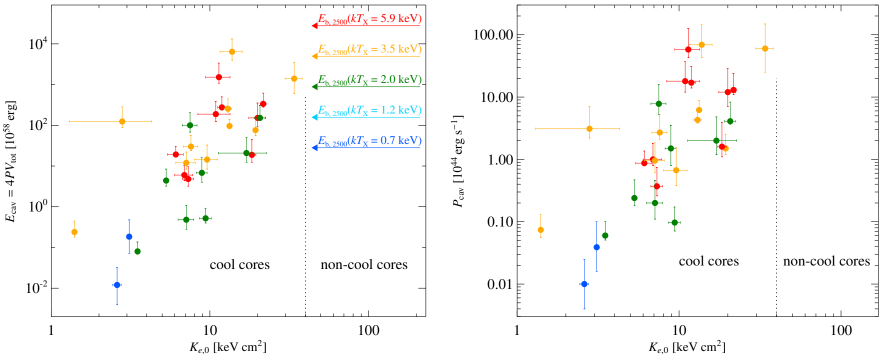


Figure 4.14: Correlations of the observed core entropy values of clusters, $K_{e,0}$, with the AGN cavity energy $E_{\rm cav}$, as inferred from the volume work done by the expanding bubbles (left) and with the cavity power $P_{\rm cav} = E_{\rm cav}/t_{\rm buoyancy}$ (right). Color coding reflects average X-ray temperatures, the lower limit of each color bin is labeled in the upper right of the left panel. Arrows denote the gas binding energy contained within a spherical region of radius $R_{2500} \simeq R_{200}/3$. While some AGN outbursts are energetic enough to unbind all or a substantial fraction of the central gas, the energy is only slowly increased and/or coupled to the surrounding cooling ICM so that its core entropy is only mildly increased as a result of AGN feedback (Pfrommer et al. 2012).)

figure shows that very powerful AGN outbursts with cavity energies up to 10^{62} erg and powers of 10^{46} erg s^{-1} are in some cases energetically capable of unbinding the gas in the entire core regions (within $R_{2500} \simeq R_{200}/3$). However, the mechanical energy of these expanding cavities only heat the cluster core enough to prevent a cooling catastrophe. On the buoyancy timescale, no AGN outburst transforms a cool core to a non-cool core cluster. This is apparent from the low core entropy values (with typically $K_0 \sim 15\,\mathrm{keV\,cm}^2$) of typical cool core clusters.

4.3.3 Intracluster Medium Turbulence

- The X-ray band is uniquely suited to observe and characterize ICM turbulence because of its excellent arcsecond angular resolution and the ability to directly probe the dynamics of the hot thermal plasma. As introduced in Section 3.1.5, turbulence cascades kinetic energy from large to small scales and dissipates it into heat. The source of the ICM turbulence are cluster mergers that result from hierarchical growth as well as AGN feedback in the centers of cool core clusters. Turbulent advection mixes metals and thermal energy and as such this process is critical for evolution of the ICM in galaxy clusters.

- Interestingly, ICM turbulence is stratified which implies that large-scale *radial* gas motions are energetically disfavored because of the stably stratified entropy profile (see Section 3.1.3) so that isotropic three-dimensional on small scales transitions to preferentially non-radial two-dimensional turbulence on scales larger than the pressure scale height (provided turbulence is driven on such large scales). Moreover, ICM turbulence in the centers of cool core clusters is multi-phase because of the short cooling times and feedback heating (see Sections 3.2.1 and 3.2.2) so that the mutual coupling of the hot and cold phase turbulence needs to be understood, i.e., whether the cold filaments are passive tracers of the hot-phase turbulence or whether they actively drive the gas motions in the hot phase.

- Whether a flow is turbulent or not is determined by the Reynolds number,

$$\mathrm{Re} = \frac{Lv}{\nu} = \frac{L}{\lambda_{\mathrm{mfp}}}\frac{v}{v_{\mathrm{th}}}. \tag{4.145}$$

  where L and v are characteristic length and velocity scales of the (macroscopic) system and the kinematic viscosity $\nu \sim \lambda_{\mathrm{mfp}} v_{\mathrm{th}}$ is to order of magnitude the product of the particle mean free path and thermal velocity. Hence, Re is the product of the ratios of macroscopic-to-microscopic length and velocity scales. The transition to turbulence in a pipe flow takes place at a critical Reynolds number around $\mathrm{Re} > \mathrm{Re}_{\mathrm{crit}} \approx 2300$, at which approximate value the laminar-to-turbulence transition is also expected in the ICM.

- The particle mean free path is given by

$$\begin{aligned} \lambda_{\mathrm{mfp}} &= \frac{1}{n\sigma \ln\Lambda} \sim \frac{1}{\pi n \ln\Lambda}\left(\frac{k_{\mathrm{B}}T_{\mathrm{e}}}{Ze^2}\right)^2 && (4.146)\\ &\sim 5\ \mathrm{kpc}\left(\frac{n}{10^{-3}\ \mathrm{cm}^{-3}}\right)^{-1}\left(\frac{k_{\mathrm{B}}T_{\mathrm{e}}}{6\ \mathrm{keV}}\right)^2, && (4.147)\end{aligned}$$

  where we have assumed $Z = 1$, $\ln\Lambda = 39$. Hence, even for trans-sonic flows $v \sim v_{\mathrm{th}}$ during a cluster merger, we obtain for an turbulent energy injection scale of $L \sim 500$ kpc a Reynolds number of $\mathrm{Re} = 100$, which is characteristic for a laminar flow.

- However, the existence of turbulent intracluster magnetic fields complicates this picture. First, it causes incompressible turbulence to become anisotropic towards scales much smaller than the energy injection scale (see Section 3.3.2.6). Second, while the mean free path along the field lines does not change, the charged constituents of the intracluster plasma are confined to orbit around individual field lines with typical proton Larmor radii of

$$r_{\mathrm{L}} = \frac{m_{\mathrm{p}} v_{\perp} c}{ZeB} = 10^5\ \mathrm{km}\left(\frac{v_{\perp}}{10^3\ \mathrm{km\ s}^{-1}}\right)\left(\frac{B}{1\,\mu\mathrm{G}}\right)^{-1} \tag{4.148}$$

Hence, the perpendicular Reynolds number is

$$\mathrm{Re}_{\perp} = \frac{L}{r_{\mathrm{L}}} \frac{v}{v_{\mathrm{th}}} = 10^{14} \frac{v}{v_{\mathrm{th}}}, \tag{4.149}$$

which is very turbulent.

- Isotropic turbulence can either be characterized with second-order velocity statistics (Kolmogorov 1941) or with their Fourier-transformed counterpart, the power spectrum (Oboukhov 1941)

$$E_v(k) = C_{\mathrm{K}} \dot{\epsilon}^{2/3} k^{-5/3}, \tag{4.150}$$

where $E_v(k)$ is the energy spectrum of the three-dimensional velocity field, k is the Fourier wave number, C_{K} is the Kolmogorov constant, and $\dot{\epsilon} = v^3/L = v_l^3/l$ is the energy injection rate into the turbulence which is equal to the energy flow rate at the scale $l = 2\pi/k$. Oboukhov (1949) and Batchelor (1951) showed that turbulent gas pressure fluctuations also obey a scaling law that is a consequence of the turbulent velocity fluctuations:

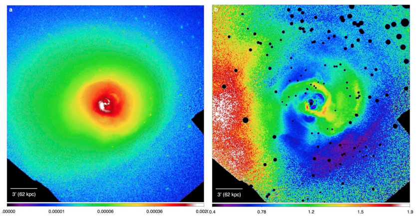
$$E_P(k) = C_P \dot{\epsilon}^{4/3} k^{-7/3}, \tag{4.151}$$

where C_P is the constant of proportionality. Assuming adiabatic pressure fluctuations enables mapping pressure to density fluctuations that are observable through the X-ray surface brightness.

- In the ICM, we cannot directly measure velocity fluctuations and hence, we have to use indirect methods to observationally infer the properties of ICM turbulence. A first method uses pressure or density fluctuations in X-ray maps to infer velocity fluctuations. A theoretical argument, supported by numerical simulations, shows that in relaxed galaxy clusters, where the gas motions are subsonic, the root-mean-square amplitudes of the density and one-component velocity fluctuations are proportional to each other at each scale l within the inertial range,

$$\frac{\delta \rho_k}{\rho_0} \approx \eta_{\mathrm{turb}} \frac{v_k}{c_{\mathrm{s}}}, \tag{4.152}$$

where ρ_0 is the mean gas density, c_{s} the sound speed and $\eta_{\mathrm{turb}} \sim 1$ is the proportionality coefficient set by gravity-wave physics at large, buoyancy-dominated scales. Here we define v_k by $3v_k^2/2 = kE_v(k)$ and $\delta\rho_k/\rho_0$ is defined analogously in terms of the density fluctuation spectrum, but without the factor of 3/2.



(a) Perseus X-ray surface brightness from *Chandra* observations. (b) The same divided by the mean surface-brightness profile, highlighting the relative density fluctuations. Black circles show excised point sources (Zhuravleva et al. 2014).

- Unsharp-masked X-ray images of the Perseus cluster by the *Chandra* telescope show ripple-like structures in the core, reminiscent either of sound waves or stratified turbulence. Subtracting a spherically symmetric beta model of the mean intensity profile from the X-ray surface brightness map yields a map of surface-brightness fluctuations. The power spectra calculated in a set of concentric annuli can be mapped to the power spectrum of density fluctuations and by means of Eq. (4.152) converted to velocity power spectra, which are consistent with the Kolmogorov-Oboukhov prediction of Eq. (4.150) (Zhuravleva et al. 2014). Note that density perturbations could also be caused by contact discontinuities of multiphase gas that is present in the cool core regions, especially in Perseus. Moreover, observations are limited by photon shot noise on small scales which precludes probing scales much smaller than the particle mean free path, also known as the *Kolmogorov scale*, where kinetic energy is dissipated.

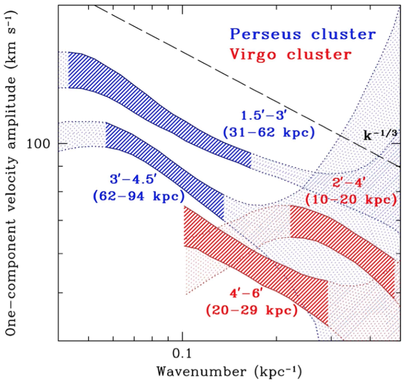
- A second method employs observations by the *Hitomi* X-ray micro-calorimeter with a high spectral resolution of better than several eV. X-ray spectral observations of the bright Perseus cool core region detected Doppler-broadened X-ray lines in a given angular region that corresponds to a physical scale. Separating thermal from turbulent Doppler-line broadening enables to infer a line-of-sight velocity fluctuations on that angular scale of $164 \pm 10\ \mathrm{km\ s^{-1}}$ (Hitomi Collaboration 2016). This implies that the turbulent pressure support in the gas is 4% or less of the thermodynamic pressure, with large-scale shear at most doubling that estimate. Note that current-day X-ray micro-calorimeters have a comparably poor angular resolution, which precludes probing the velocity power spectrum at the Kolmogorov scale.

- A third method has the potential to probe the physics on scales even much smaller than the Kolmogorov scale. The idea is to complement X-ray with spectral data of Hα and molecular filaments, which probe warm ($\sim 10^4$ K) and cold gas (~ 10 K), respectively. These data do not suffer from small X-ray photon statistics on small angular scales and may reveal the plasma physics of thermalization. Spectral shifts of Hα and molecular emission lines enable constructing velocity structure functions (VSFs), which directly probe the character of turbulence in the warm/cold phase across more than two orders of magnitude in scale. The VSF at scale l averages over all possible line-of-sight velocity differences separated by l,

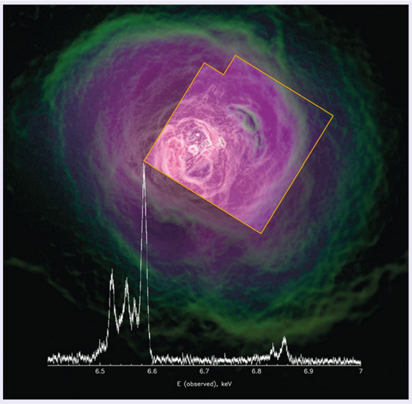
$$\mathrm{VSF}(l) = \langle |\upsilon_z(\boldsymbol{x}+\boldsymbol{l}) - \upsilon_z(\boldsymbol{x})| \rangle = \langle |\delta \upsilon_z| \rangle \propto l^{1/3} \tag{4.153}$$

where in the last step, we adopted the Kolmogorov scaling of velocity fluctuations derived in Eq. (3.87).

- VSF measurements in Perseus and other clusters demonstrate a turbulent velocity field of Hα filaments, which however show steeper slopes $\mathrm{VSF}(l) \propto l^{1/2\ldots 2/3}$ (Li et al. 2020, see Fig. 4.15)



Amplitude of υ_k versus wavenumber k for two different annuli in Perseus (blue) and M87/Virgo (red). The values are obtained from the power spectra of density fluctuations, derived from the X-ray images (Zhuravleva et al. 2014).



Field of view of the *Hitomi* X-ray observations of the Perseus cluster overlaid with the Doppler-broadened emission lines of the X-ray spectrum.

The core of the Perseus galaxy cluster in visible light with Hα filaments (red) that surround the cD galaxy NGC1275, which hosts the supermassive black hole.

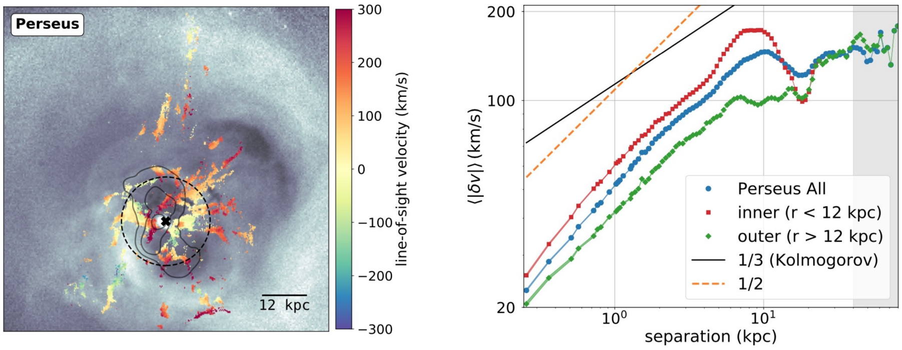


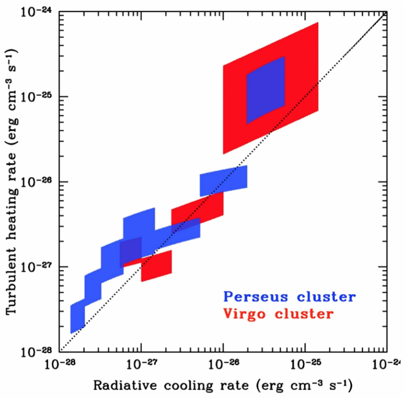
Figure 4.15: Left: velocity maps of the Hα filaments overlaid on the X-ray residual images (gray) in the center of Perseus. The black cross indicates the position of the supermassive black hole, and the black circles denote the separation of the inner and outer regions in our analysis. Black contours show the low-frequency radio synchrotron emission. Right: corresponding VSFs of the filaments (Li et al. 2020). The motion of the filaments is turbulent but it shows a significantly steeper scale dependence than expected from Kolmogorov turbulence. The features in the VSFs correspond to AGN activities.

which are inconsistent with the expectation from Kolmogorov turbulence of Eq. (4.153). The VSFs of the Hα filaments are consistent with those of the molecular gas observed by ALMA. The velocity v of the warm/cold filaments subject to constant gravitational acceleration, g, scales with travel length l as $v \propto (gl)^{1/2}$, which can explain the 1/2 slope of the observed VSFs. Hence, the physical origin of the steeper slope of the cold phase VSF may be (partially) attributed to ballistically moving dense filaments that move under the action of gravity in the central cluster potential. Those precipitating warm and cold dense filaments could be responsible for driving turbulent motions in the hot gaseous phase observable at X-ray energies. This may also explain the radial bias of cold and hot phase velocity distributions.

- The driving scale of the turbulence is consistent with the sizes of X-ray bubbles and (if present) jets so that this is evidence for the picture of black-hole-driven turbulence in the centers of galaxy clusters either directly via stirring of the hot phase or indirectly by pulling up cold gas in the wake of the buoyantly rising bubbles so that it adiabatically cools, precipitates and falls back to the center while stirring the hot gas phase.

- To conclude, turbulent velocity spectra in Perseus inferred by VSFs and *Chandra* surface brightness fluctuation analyses agree with each other. The integrated kinetic energy from *Hitomi* X-ray Doppler line broadening measurements agrees with turbulent amplitudes from VSF analyses. Most importantly, turbulence extends to scales smaller that the Kolmogorov scale, which is providing evidence for collisionless plasma effects such as Braginskii magneto-hydrodynamics (MHD). Open questions remain about the exact relation between the hot and cold phase turbulence and the role of the magnetic field in coupling these two phases. Most importantly, ongoing research will shed light on the question why ICM turbulence is non-Kolmogorov and hopefully unveils the plasma physics beyond the Kolmogorov scale.

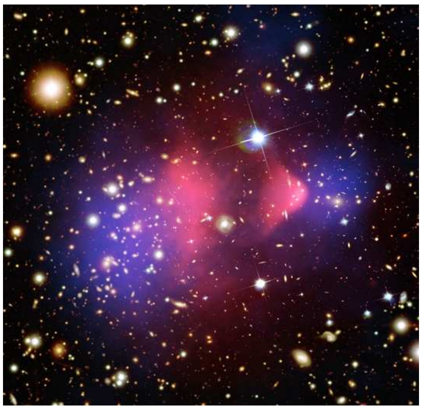
- Finally, turbulent motions in the gas must eventually dissipate into heat. In order to determine whether this heating is sufficient to balance radiative losses and prevent net cooling, one must estimate the turbulent heating rate – and for that, we need to measure the root-mean-square turbulent velocity amplitude v_l as a function of length scale l. The turbulent heating rate in the gas with mass density ρ is $Q_{\rm turb} \sim \rho v_l^3/l$, to within some constant of order unity that depends on the exact properties of the turbulent cascade. The figure on the right shows the comparison of $Q_{\rm turb}$ to the radiative cooling rate, $Q_{\rm cool} = n_{\rm H}^2 \Lambda_0(T)$, where $\Lambda_0(T)$ is the normalized cooling rate (see Eq. 3.185). This shows that $Q_{\rm turb} \sim Q_{\rm cool}$ over nearly three orders of magnitude in Perseus and Virgo. Note that in Virgo and Perseus similar levels of $Q_{\rm turb}$ and $Q_{\rm cool}$ are attained at physically different distances from the cluster centers because of the different density profiles. Note also that such a correlation does not necessarily imply a causal relation: the dependence of $Q_{\rm turb}$ and $Q_{\rm cool}$ on density may (partially) explain this correlation. Another problematic aspect of this solution to the cooling flow problem is that turbulent heating does not provide a thermally stable heating as we have discussed in Section 3.2.6.2.



Turbulent heating versus gas cooling rates in the Perseus and Virgo cores. Each shaded rectangle shows the statistical uncertainties in the heating and cooling rates estimated in a given annulus (top right – the innermost radius; bottom left – the outermost radius).

4.3.4 Merger Shocks and Electron Equilibration

- Combined optical, X-ray and gravitational lensing analyses of the bullet cluster shows a spectacular example of a cluster with the merger axis nearly aligned with the plane of the sky (see figure to the right). As the subcluster ("the bullet") has passed from the left to the right and its dense cool core has driven a curved shock through the main cluster. The bullet can be identified with the cool core region of the merging subcluster while there is a tangential discontinuity separating the bullet from the shocked ICM to the right of it. As we will see in Section 4.6.4, this enables us to probe the collisionless nature of dark matter.



The bullet cluster 1E-0657 (red: X-ray, blue: weak lensing, and galaxies in the optical) showing evidence for collisionless dark matter.

- The ions are dissipatively heated at the shock so that their temperature $T_{\rm i}$ follows the Rankine-Hugoniot jump condition. By contrast, in the classical picture the electrons are only adiabatically compressed at the shock so that we have the following expressions for the change in ion and electron temperatures, respectively:

$$\Delta(k_\mathrm{B}T_\mathrm{i}) \simeq m_\mathrm{i}(\Delta v)^2, \tag{4.154}$$

$$\Delta(k_\mathrm{B}T_\mathrm{e}) \simeq m_\mathrm{e}(\Delta v)^2, \tag{4.155}$$

which is much smaller for the electrons by a factor $m_\mathrm{i}/m_\mathrm{e} = 1836 \times A$, where A is the mass number of the ion. In the classical picture, thermal equilibrium between electrons and ions can only be established via an exchange of energy through Coulomb collisions further downstream the shock. Because we cannot measure T_i in X-rays but only T_e, we need to understand these processes.

- The electron-ion equilibration process can be described by the equation

$$\frac{\partial(T_\mathrm{e} - T_\mathrm{i})}{\partial t} = -\nu_\mathrm{ei}(T_\mathrm{e} - T_\mathrm{i}), \tag{4.156}$$

where the electron-ion equilibration rate via Coulomb collisions (see Section 3.2.5.2) is given by

$$\begin{aligned} \nu_\mathrm{ei} &\approx 4\frac{m_\mathrm{e}}{m_\mathrm{p}}\frac{v_\mathrm{the}}{\lambda_\mathrm{mfp}} & (4.157)\\ &\approx 3.3\times10^{-16}\ \mathrm{s}^{-1}\left(\frac{k_\mathrm{B}T_\mathrm{e}}{10\ \mathrm{keV}}\right)^{-3/2}\left(\frac{n_\mathrm{e}}{10^{-3}\ \mathrm{cm}^{-3}}\right), & (4.158) \end{aligned}$$

the electron thermal speed is here defined as $v_\mathrm{the} = (2k_\mathrm{B}T_\mathrm{e}/m_\mathrm{e})^{1/2}$, λ_mfp is the electron mean free path introduced in Eq. (4.147), and the factor $m_\mathrm{e}/m_\mathrm{p}$ accounts for the energy transfer per collision so that many collisions are needed to reach thermal equilibrium.

- Equation (4.156) is solved by

$$T_\mathrm{e} - T_\mathrm{i} = \mathrm{e}^{-\nu_\mathrm{ei}t}, \tag{4.159}$$

so that we expect the electron temperature to approach that of the ions on time and length scales behind a shock of

$$\begin{aligned} \tau_\mathrm{ei} &= \nu_\mathrm{ei}^{-1} \approx 95\ \mathrm{Myr}\left(\frac{k_\mathrm{B}T_\mathrm{e}}{10\ \mathrm{keV}}\right)^{3/2}\left(\frac{n_\mathrm{e}}{10^{-3}\ \mathrm{cm}^{-3}}\right)^{-1}, & (4.160)\\ L_\mathrm{ei} &= v_\mathrm{post}\tau_\mathrm{ei} \approx 155\ \mathrm{kpc}\left(\frac{v_\mathrm{post}}{1600\ \mathrm{km\ s}^{-1}}\right), & (4.161) \end{aligned}$$

where we adopted characteristic values for the post-shock region of the bullet cluster (and suppressed the dependence on $k_\mathrm{B}T_\mathrm{e}$ and n_e in the last line). This predicts an observable effect for the most extreme mergers.

- However, we are encountering collisionless shocks in galaxy clusters so that we additionally have to consider wave-particle interactions. It turns out, that drifting ions in the shock transition zone drive electromagnetic waves (along the background magnetic field) unstable. These unstable modes can be identified with background ion-cyclotron and resonant Alfvén modes and are expected to grow exponentially fast (Shalaby et al. 2021). As a result, these modes scatter and energize electrons on much faster (plasma) time scales so that the electron temperature is expected to equilibrate with the ions within a factor of two already in the shock transition zone (Shalaby et al. 2022).

4.3.5 Magnetic Draping and Cold Fronts

- Every object that moves super-Alfvénically through a magnetized plasma can very rapidly sweep up a significant magnetic layer which is then "draped" over the object. Initially, the magnetic field grows to the point where it becomes dynamically important so that the magnetic pressure is able to push magnetic field lines over the object. Hence, the magnetic field strength in this draping layer is set by a competition between "plowing up" and slipping around of field lines, and depends primarily on the ram pressure seen by the moving object,

$$P_B = \frac{B^2}{8\pi} = \alpha\rho_0 \upsilon^2, \tag{4.162}$$

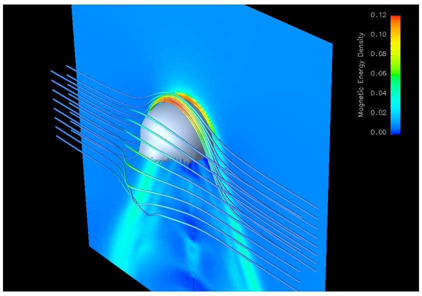
  where P_B is the magnetic pressure in the draping layer, ρ_0 is the mass density of the ambient medium, υ is the velocity of the object, and $\alpha \approx 2$ is a constant of proportionality that is determined from MHD simulations (Dursi & Pfrommer 2008).

- Note that magnetic draping is not similar to ram pressure compression because the density is not increased in the magnetic draping layer as shown by ideal MHD simulations (picture to the right). Clearly, the magnetic flux is still frozen into the plasma but plasma can also move along field lines while field lines get stuck at obstacle.

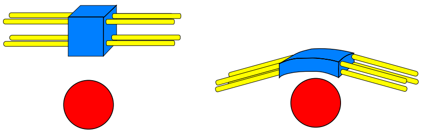
- In order to get a theoretical understanding of the magnetic draping effect, we relate the Alfvénic Mach number $\mathcal{M}_\mathrm{A} = \upsilon/\upsilon_\mathrm{A}$, i.e., the flow speed υ in units of the local Alfvén speed $\upsilon_\mathrm{A} = B_0/\sqrt{4\pi\rho_0}$, to the sonic Mach number $\mathcal{M} = \upsilon/c_\mathrm{s}$, i.e., the flow speed in units of the local sound speed, $c_\mathrm{s} = \sqrt{\gamma P_{\mathrm{th},0}/\rho_0}$, where $\gamma = 5/3$ is the adiabatic index:

$$\mathcal{M}_\mathrm{A}^2 = \frac{\upsilon^2}{\upsilon_\mathrm{A}^2} = \frac{\rho_0\upsilon^2}{2P_{B_0}} = \frac{1}{2}\beta_0\gamma\mathcal{M}^2, \tag{4.163}$$

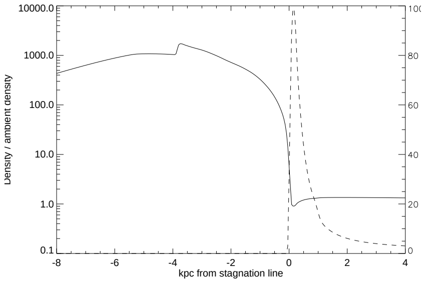
  where $P_{B_0} = B_0^2/(8\pi)$ is the magnetic pressure and $\beta_0 = P_{\mathrm{th},0}/P_{B_0}$ is the plasma beta parameter of the ambient medium (in our case the ICM).



A dense cloud moving upwards through a magnetized plasma sweeps up a strongly magnetized drape. Colors represent the magnetic energy density (Dursi & Pfrommer 2008).



Cartoon showing the distortion of incoming fluid elements and stretching of field lines as a spherical projectile moves upward through the ambient medium (Dursi & Pfrommer 2008).



Comparison of the mass density profile (solid) and magnetic energy density profile (dashed) along the stagnation line in an MHD simulation of magnetic draping. There is no density increase associated with the formation of the magnetic draping layer (Dursi & Pfrommer 2008).

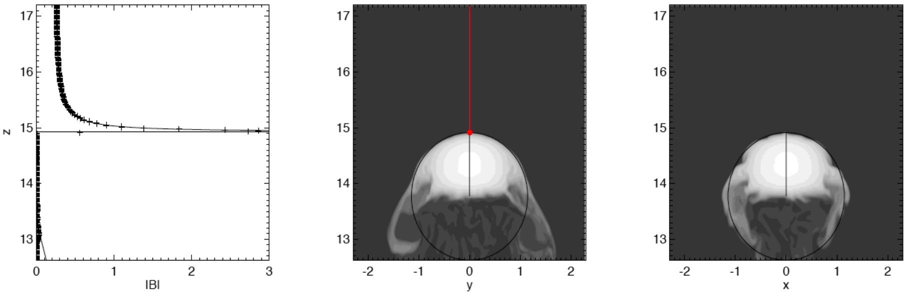


Figure 4.16: Simulations of magnetic draping vs. the analytical solution in the kinematic limit. Left: fitting the peak position and fall-off radius of the analytical prediction of the magnetic field strength B as a function of z coordinate ahead of the draping object. Middle and right: density cut-planes where the circle shows radius and position given by the fit to the magnetic field structure on the left. The stagnation point and the stagnation line is shown in red. There is astonishing agreement of curvature radius at the working surface with potential flow predictions (Dursi & Pfrommer 2008).

- The magnetic field strength in the draping layer obeys the following analytical form in the kinematic regime that neglects magnetic back-reaction (Dursi & Pfrommer 2008, see Fig. 4.16):

  $$B = \frac{B_0}{\sqrt{1-\frac{R^3}{(R+s)^3}}} \approx \sqrt{\frac{R}{3s}}B_0 + O\left(\sqrt{\frac{s}{R}}\right), \tag{4.164}$$

  $$P_B = \frac{B^2}{8\pi} = P_{B_0}\frac{R}{3s} = \alpha\rho_0 v^2, \tag{4.165}$$

  where $s > 0$ is the distance measured from the stagnation point[2] into the upstream of the draping object and R is the object's curvature radius at the stagnation point. Hence, the thickness of the draping layer is given by

  $$l_{\mathrm{drape}} \equiv s = \frac{R\,P_{B_0}}{3\alpha\rho_0 v^2} = \frac{R}{6\alpha\mathcal{M}_{\mathrm{A}}^2} = \frac{R}{3\alpha\beta_0\gamma\mathcal{M}^2} \simeq 100\,\mathrm{pc}, \tag{4.166}$$

  where we adopted typical values for the curvature radius, $R \simeq 30$ kpc, and the plasma beta parameter, $\beta \simeq 50$, and assume a trans-sonic flow, $\mathcal{M}^2 \simeq 1/\gamma$.

- We have seen that in the draping layer, $P_B = \alpha\rho_0 v^2$, is solely given by the ram pressure and *completely* independent of the magnetic field strength in the ICM (provided we have reached a steady state, which takes longer for smaller field strengths). To determine the magnetic energy of the draping layer, we assume sphere with radius R and volume V, and a constant thickness of the drape

---

[2] The stagnation point is a point on the surface of a solid body immersed in a fluid stream which directly faces the stream and at which the stream lines separate.

$l_{\rm drape}$ over the area of the half-sphere, $A = 2\pi R^2$:

$$E_B = \frac{B^2}{8\pi} A\, l_{\rm drape} = \frac{B^2}{8\pi} \frac{A\,R}{6\alpha\,\mathcal{M}_{\rm A}^2} \tag{4.167}$$

$$= \alpha\rho_0 \upsilon^2 \frac{A\,R}{6\alpha} \frac{B_0^2}{4\pi\rho_0\upsilon^2} = \frac{\varepsilon_{B_0} V}{2} = \frac{E_{B_0}}{2}, \tag{4.168}$$

where we have used Eqs. (4.162), (4.163), and (4.166). Hence, we have derived an "Archimedes principle of magnetic draping". While this result derives from the analytic solution of this MHD problem in the kinematic regime, we could have derived it from considering the magnetic flux freezing condition (see Section 3.3.2.5) during the draping procedure in steady state. If Eq. (4.168) were not true, we either would not have reached steady state (for $E_B < E_{B_0}/2$) or we would have encountered a region with an increased magnetic energy density (for $E_B > E_{B_0}/2$) because the field lines that are accumulating in the draping layer are connected to those field lines at a large distance, which subtend the solid angle of the upper half sphere that is being wrapped by the magnetic draping layer.

- This draping effect occurs throughout astrophysics. Solar-wind magnetic field is draped around the magnetopause of Earth and around other moons and planets of the solar system. Magnetic draping at the magnetopause of Earth protects life on Earth from cosmic rays during times of geomagnetic field reversal when our magnetic poles flip sign. Interstellar magnetic fields are also draped around orbiting stars and dense clouds in galaxies. In the context of galaxy clusters, the intracluster magnetic field is draped around orbiting galaxies, dense cores of merging subclusters, and around low-density jets and radio lobes that have been inflated by AGN jets.

- This magnetic draping effect has important consequences for features observed in high-resolution X-ray surface brightness images. There are sharp, abrupt density transitions seen that are balanced by equally strong but oppositely directed temperature jumps so that they represent tangential discontinuities (see Section 3.1.6.3). These abrupt transitions are not expected to be stable against hydrodynamic shear motions at the interface which excite Kelvin-Helmholtz instabilities. These grow in amplitude and scale to eventually mix the hot and cold phases on length scales comparable to the curvature radius of the interface. It turns out that magnetic draping at moving substructures provides enough magnetic tension to suppress interface instabilities for shear velocities $\Delta\upsilon \lesssim \upsilon_{\rm A}$, where $\upsilon_{\rm A}$ is the Alfvén velocity of the draping layer.

- Moreover, sharp discontinuities in density and temperature should be smoothed out by particle diffusion and thermal conduction on the comparable short ion and electron transport times. Again, magnetic draping comes to rescue and suppresses these transport effects because in the weakly collisional ICM, electrons and ions are confined to move on individual magnetic field lines. However, those magnetic fields are precisely draped at the interface of both phases such that the separate the hot/dilute from the cold/dense phase. Hence, electrons and ions are prevented to move across the discontinuities and as such, magnetic draping maintains the sharpness of these interfaces on much longer time scales than otherwise expected. In fact, very high-resolution X-ray images start to resolve the widths of cold fronts and hence have the potential to probe anisotropic transport effects (e.g., heat conduction, viscosity, cosmic ray diffusion).

- As spiral galaxies orbit in galaxy clusters they sweep up magnetic field lines that assemble in form of magnetic draping sheaths. This drape is then lit up with cosmic ray electrons from the galaxies' stars, generating coherent polarized radio synchrotron emission at the galaxies' leading edges, which is observed in the Virgo cluster. This immediately presents a technique for probing local orientations and characteristic length scales of cluster magnetic fields (Pfrommer & Dursi 2010).

- An off-axis minor merger gravitationally attracts the dense cool core of the parent cluster and displaces it from its equilibrium position. Angular momentum transfer from the merging subcluster to the displaced cool core causes it to slosh in the gravitational potential of the parent cluster which is analogous to swirling wine in a glass. This sloshing effect causes the generation of spiral density patterns that are called sloshing cold fronts. Strong shear motions between the interface of the sloshing cool core and the ambient, lower-density external ICM causes small-scale Kelvin-Helmholtz instabilities, which drive a small-scale magnetic dynamo that amplifies magnetic field in the interface until it saturates when its energy balances the kinetic shear energy. This strongly sheared magnetic field should then lie along the spiral pattern and prevent any large-scale Kelvin-Helmholtz instabilities so that these spiral structures are maintained for long (several Gyr) time scales and as such, explain the ubiquity of the spiral patterns frequently found in high-resolution X-ray observations of galaxy clusters.

4.4 Sunyaev-Zel'dovich (SZ) Effect

Here, we present the underlying physics of the Sunyaev-Zel'dovich (SZ) effect. First, we introduce the various contributions of the thermal, kinematic, and relativistic SZ effects. Then we explain how the SZ effect can be used as a cosmological probe by means of cluster number counts, scaling relations and the SZ power spectrum. Finally we show how the SZ effect elucidates cluster astrophysics by providing a new window to AGN bubbles and cluster shocks.

4.4.1 Thermal and Kinematic SZ Effect

- ***The cosmic microwave background (CMB)*** has an almost perfect Planckian spectral distribution that emits an intensity as a function of photon energy $\hbar\omega$,

$$I(x) = i_0 i(x) = i_0 \frac{x^3}{\mathrm{e}^x - 1}, \text{ where} \tag{4.169}$$

$$x = \frac{\hbar\omega}{k_\mathrm{B} T_\mathrm{cmb}}, \tag{4.170}$$

$$i_0 = \frac{2(k_\mathrm{B} T_\mathrm{cmb})^3}{(hc)^2} = 22.8 \text{ Jy arcmin}^{-2}. \tag{4.171}$$

  Here, $T_\mathrm{cmb} = 2.725$ K denotes the average CMB temperature at the present epoch and k_B, h, and c denote Boltzmann's constant, Planck's constant, and the speed of light, respectively. The CMB black body spectrum peaks in the microwave range at a frequency of 160.4 GHz. In the following, we adopt the abbreviation $T \equiv T_\mathrm{cmb}$.

- ***Thermal SZ effect.*** As discussed in Section 1.2.4, the thermal SZ effect arises because CMB photons can inverse Compton scatter off of electrons of the hot, dilute intra-cluster plasma. Thus, at the angular position of galaxy clusters, the CMB spectrum is modulated as photons are redistributed from the low-frequency part of the spectrum below a characteristic crossover frequency ν_c to higher frequencies. Hence, ν_c demarks the transition from a net decrease to a net increase of photon number density. For a non-relativistic electron population in a cluster at rest with respect to the CMB rest frame, $\nu_\mathrm{c} \simeq 217$ GHz, while this characteristic frequency shifts towards higher values for more energetic electrons, as we will see below.

- In order to analyse these distortions quantitatively, we need to consider a transport equation governing the effect of Compton scattering on the photon spectrum. We first consider inverse

Compton scattering of non-relativistic electrons in the limit

$$\frac{\hbar\omega}{c} \ll p \ll m_{\mathrm{e}}c, \tag{4.172}$$

where p is the electron momentum. The change in the photon phase space occupation number $n(\omega, t)$ can be derived with the Kompaneets equation, which is a quantum mechanical extension to the Fokker-Planck equation (Peacock 1999):

$$\frac{\partial n}{\partial t} = \frac{\sigma_{\mathrm{T}} n_{\mathrm{e}} \hbar}{m_{\mathrm{e}} c} \frac{1}{\omega^2} \frac{\partial}{\partial \omega}\left[\omega^4 \left(\frac{k_{\mathrm{B}} T_{\mathrm{e}}}{\hbar} \frac{\partial n}{\partial \omega} + n + n^2\right)\right], \tag{4.173}$$

Here, m_{e} is the electron rest mass and σ_{T} denotes the Thompson cross section. T_{e} and n_{e} are the electron temperature and number density, and t is the time variable.

- We can rewrite this equation by introducing the differential Compton-y parameter and the dimensionless photon frequency[3], x_{e}, via

$$\mathrm{d}y = \frac{k_{\mathrm{B}} T_{\mathrm{e}}}{m_{\mathrm{e}} c^2} n_{\mathrm{e}} \sigma_{\mathrm{T}} c \mathrm{d}t \quad \text{and} \quad x_{\mathrm{e}} = \frac{\hbar\omega}{k_{\mathrm{B}} T_{\mathrm{e}}}, \tag{4.174}$$

so that we arrive at the following compact form of Kompaneets equation:

$$\frac{\partial n}{\partial y} = \frac{1}{x_{\mathrm{e}}^2} \frac{\partial}{\partial x_{\mathrm{e}}}\left[x_{\mathrm{e}}^4 \left(\frac{\partial n}{\partial x_{\mathrm{e}}} + n + n^2\right)\right]. \tag{4.175}$$

Equation (4.175) describes changes to the photon phase space occupation number due to small energy transfers in the inverse Compton scattering process, which is known as the thermal SZ effect (Sunyaev & Zel'dovich 1972). In this limit, the scattering process is a combination of diffusion in momentum space (the first term) and advection, which accounts for adiabatic changes (the second term).

- The meaning of the third, non-linear term in the photon phase space occupation number in Eq. (4.175) becomes clear by considering the equilibrium solution, which is obtained by setting the terms in parenthesis on the right-hand side of Eq. (4.175) equal to zero:

$$\frac{\partial n}{\partial x_{\mathrm{e}}} + n + n^2 = 0. \tag{4.176}$$

This is realized in Nature provided electrons and photons are in thermal equilibrium. Clearly, in this case, we obtain the Bose-Einstein spectrum

$$n(x_{\mathrm{e}}) = \frac{1}{\mathrm{e}^{x_{\mathrm{e}}+\mu_{\mathrm{c}}} - 1}, \tag{4.177}$$

---

[3] Note that this quantity x_{e} differs from the photon energy measured in units of the CMB thermal energy, x defined in Eq. (4.170) by the adopted temperatures.

which can be readily verified. This demonstrates that the non-linear term is responsible for the formation of an Bose-Einstein condensate of a highly degenerated boson gas. The case of a vanishing chemical potential, $\mu_c = 0$, corresponds to a Planckian distribution function.

- In the limit of small x_e, which is appropriate for inverse Compton scattering of CMB photons with hot, non-relativistic ICM electrons, $\partial n/\partial x_e \gg n, n^2$, and the Kompaneets equation yields a linear change in the phase space occupation number,

$$\frac{\partial n}{\partial y} = \frac{1}{x_e^2}\frac{\partial}{\partial x_e}\left(x_e^4\frac{\partial n}{\partial x_e}\right) \tag{4.178}$$

$$= \frac{1}{x^2}\frac{\partial}{\partial x}\left(x^4\frac{\partial n}{\partial x}\right) = 4x\frac{\partial n}{\partial x} + x^2\frac{\partial^2 n}{\partial x^2}. \tag{4.179}$$

- We now insert the Planckian distribution for the CMB photons,

$$n(x) = \frac{1}{e^x - 1} \tag{4.180}$$

into the the linearized Kompaneets equation (4.179) and derive

$$\frac{\partial n}{\partial x} = -n^2 e^x \quad \text{and} \quad \frac{\partial^2 n}{\partial x^2} = -ne^x\left(n + 2\frac{\partial n}{\partial x}\right). \tag{4.181}$$

Substituting these expressions into Eq. (4.179) and rearranging terms yields

$$\frac{\partial n}{\partial y} = \frac{xe^x}{(e^x - 1)^2}\left(x\frac{e^x + 1}{e^x - 1} - 4\right). \tag{4.182}$$

This equation describes a change of the phase-space density (or the occupation number) of CMB photons with the Compton-y parameter as they propagate through a plasma.

- In order to obtain the change of photon intensity with the Compton-y parameter, we need to multiply the Kompaneets equation (4.182) by a factor $i_0 x^3$, as defined in Eq. (4.169). After integrating over y, we find the change in intensity as a result of the thermal SZ effect across the sky spanned by the vector $\boldsymbol{\theta} = (\theta, \phi)$:

$$\Delta I_{\text{tSZ}}(x, \boldsymbol{\theta}) = i_0 y(\boldsymbol{\theta}) g(x), \quad \text{where} \tag{4.183}$$

$$g(x) = \frac{x^4 e^x}{(e^x - 1)^2}\left(x\frac{e^x + 1}{e^x - 1} - 4\right), \quad \text{and} \tag{4.184}$$

$$y(\boldsymbol{\theta}) = \frac{\sigma_T}{m_e c^2}\int n_e(\boldsymbol{r}) k_B T_e(\boldsymbol{r}) c\,dt \tag{4.185}$$

denotes the amplitude of the thermal SZ effect, which is known as the thermal Comptonization parameter y that is defined as the line-of-sight integration of the temperature weighted thermal electron density from the observer to the last scattering surface of the CMB at redshift $z_{\max} = 1090$.

- At small frequencies in the Rayleigh–Jeans tail of the CMB spectrum, we obtain $i(x) \to x^2$ for $x \ll 1$. The thermal SZ spectrum assumes the following simple form for $x \ll 1$ (or $\nu \ll 60$ GHz), which follows from Taylor expanding the thermal spectral distortion in Eq. (4.184):
$$g(x) \to -2x^2. \tag{4.186}$$

- In order to compute the change in thermodynamic temperature as a result of the thermal SZ effect, we need to apply the chain rule to the Kompaneets equation (4.182),
$$\frac{\partial n}{\partial T}\frac{\partial T}{\partial y} = \frac{x\mathrm{e}^x}{(\mathrm{e}^x-1)^2}\left(x\,\frac{\mathrm{e}^x+1}{\mathrm{e}^x-1} - 4\right), \tag{4.187}$$
$$\frac{\partial n}{\partial T} = \frac{\partial n}{\partial x}\frac{\partial x}{\partial T} = \frac{\mathrm{e}^x}{(\mathrm{e}^x-1)^2}\,\frac{x}{T}. \tag{4.188}$$
Combining these results with Eq. (4.183) enables us to determine the relative change $\Delta T/T$ in thermodynamic CMB temperature at position $\boldsymbol{\theta}$ due to the thermal SZ effect,
$$\frac{\Delta T_{\mathrm{tSZ}}}{T}(\boldsymbol{\theta}) = y(\boldsymbol{\theta})\left(x\,\frac{\mathrm{e}^x+1}{\mathrm{e}^x-1} - 4\right) \equiv y(\boldsymbol{\theta})f(x). \tag{4.189}$$

- ***Kinematic SZ effect.*** There is an additional spectral distortion of the CMB spectrum due to the Doppler effect of the bulk motion of baryonic matter streams inside a cluster or of the motion of the cluster as a whole relative to the CMB rest frame. If the component of the cluster's peculiar velocity is projected along the line-of-sight, then the Doppler effect leads to a change in thermodynamic temperature referred to as the kinematic SZ effect,
$$\frac{\Delta T_{\mathrm{kSZ}}}{T}(\boldsymbol{\theta}) = -w(\boldsymbol{\theta}), \tag{4.190}$$
$$w(\boldsymbol{\theta}) \equiv \sigma_{\mathrm{T}} \int \mathrm{d}l\, n_{\mathrm{e}}(\boldsymbol{r})\,\frac{v_r}{c}. \tag{4.191}$$
The amplitude of the kinematic SZ effect is given by the kinematic Comptonization parameter w that is equal to the dimensionless streaming velocity, v_r/c, times the optical depth of free electrons along the line of sight. We have $v_r < 0$ if the gas is approaching the observer, which results in a temperature increase.

- The spectral distortion of the kinematic SZ effect can be obtained by multiplying the change in thermodynamic temperature, Eq. (4.190), by $\partial n/\partial T$ (Eq. 4.188) and by a factor $i_0 x^3 T$ to get
$$\Delta I_{\mathrm{kSZ}}(x,\boldsymbol{\theta}) = -i_0 w(\boldsymbol{\theta}) h(x), \quad \text{where} \tag{4.192}$$
$$h(x) = \frac{x^4\, e^x}{(e^x-1)^2}. \tag{4.193}$$

- ***Comparing the thermal to the kinematic SZ effect.*** The amplitude of the thermal SZ effect, y, is typically more than one order of magnitude larger than the amplitude of the kinematic SZ effect w, which can be seen with an order of magnitude analysis of Eqs. (4.185) and (4.191):

$$y = \sigma_\mathrm{T} \int_0^L n_\mathrm{e} \frac{k_\mathrm{B} T_\mathrm{e}}{m_\mathrm{e} c^2} \mathrm{d}l \sim \sigma_\mathrm{T} \bar{n}_\mathrm{e} L \frac{k_\mathrm{B} T_\mathrm{e}}{m_\mathrm{e} c^2} \sim \tau \frac{k_\mathrm{B} T_\mathrm{e}}{m_\mathrm{e} c^2}, \tag{4.194}$$

$$w = \sigma_\mathrm{T} \int_0^L n_\mathrm{e} \frac{\upsilon_r}{c} \mathrm{d}l \sim \sigma_\mathrm{T} \bar{n}_\mathrm{e} L \frac{\upsilon_r}{c} \sim \tau \frac{\upsilon_r}{c}. \tag{4.195}$$

  This shows that both amplitudes scale to order of magnitude with the Thompson optical depth τ and a factor that describes the mean transfer of energy during an inverse Compton scattering event. This factor can be estimated by using the expression for the sound speed, $c_\mathrm{s}^2 = \gamma k_\mathrm{B} T / (\mu m_\mathrm{p})$ and adopting the standard deviation of radial cluster velocities relative to the CMB rest frame (Schäfer et al. 2006):

$$\frac{y}{\tau} \sim \frac{k_\mathrm{B} T_\mathrm{e}}{m_\mathrm{e} c^2} \sim \frac{m_\mathrm{p}}{m_\mathrm{e}} \frac{c_\mathrm{s}^2}{c^2} \sim 0.02 \left(\frac{c_\mathrm{s}}{10^3 \,\mathrm{km\,s^{-1}}} \right)^2, \tag{4.196}$$

$$\frac{w}{\tau} \sim \frac{\upsilon_r}{c} \sim 0.001 \left(\frac{\upsilon_r}{300 \,\mathrm{km\,s^{-1}}} \right), \tag{4.197}$$

  implying $y/w \sim 20$.

- Important SZ quantities are the (thermal and kinematic) Comptonizations integrated over the cluster face,

$$Y = \int_\Omega \mathrm{d}\Omega \, y = \frac{1}{D_\mathrm{ang}^2} \int_A \mathrm{d}^2 r \, y = \frac{1}{D_\mathrm{ang}^2} \frac{\sigma_\mathrm{T}}{m_\mathrm{e} c^2} \int_V \mathrm{d}^3 r \, P_\mathrm{e}, \tag{4.198}$$

$$W = \int_\Omega \mathrm{d}\Omega \, w = \frac{1}{D_\mathrm{ang}^2} \int_A \mathrm{d}^2 r \, w = \frac{1}{D_\mathrm{ang}^2} \sigma_\mathrm{T} \int_V \mathrm{d}^3 r \, n_\mathrm{e} \frac{\upsilon_r}{c}, \tag{4.199}$$

  where D_ang denotes the angular diameter distance and $P_\mathrm{e} = n_\mathrm{e} k_\mathrm{B} T_\mathrm{e}$ is the thermal electron pressure. Note that V denotes the conical volume bounded by the edge of the cluster and extends from us to the surface of last scattering at $z_\mathrm{max} = 1090$. Because the thermal pressure drops steeply outside the (non-spherical) cluster accretion shock that is situated between 1–3 R_{200}, people very often assume a spherical cluster volume (that reaches, e.g., out to R_{200}) and perform an integration of the thermal pressure.

- The integrated Comptonization is proportional to the thermal energy content of a cluster, $Y \propto E_\mathrm{th}$, which is related to the energy of the gravitational potential in hydrostatic equilibrium and less vulnerable to observational biases. Hence, X-ray observers construct a similar quantity from X-ray inferred electron number densities and temperatures,

$$Y_\mathrm{X} = \frac{1}{D_\mathrm{ang}^2}\frac{\sigma_\mathrm{T}}{m_\mathrm{e}c^2}\int_A \mathrm{d}^3r\, n_\mathrm{X} k_\mathrm{B} T_\mathrm{X}. \tag{4.200}$$

4.4.2 Relativistic SZ Effect

- Averaging the Compton interaction of a photon and a non-relativistic electron of energy $k_\mathrm{B}T_\mathrm{e}$ over scattering angles leads to a relative energy change of

  $$\frac{\langle \Delta E_\gamma \rangle}{E_\gamma} = \frac{k_\mathrm{B}T_\mathrm{e}}{m_\mathrm{e}c^2} \quad \text{for} \quad p \ll m_\mathrm{e}c. \tag{4.201}$$

  This result changes if we consider relativistic electrons. Consider a Compton collision between a photon of energy E and an electron with a Lorentz factor γ_e in the observer's frame Σ. We perform now a Lorentz transformation into the electron's rest frame Σ'. Energy transforms as the time component of the energy-momentum four-vector, so that the photon energy E' in the frame Σ' before scattering is given by $E' = \gamma_\mathrm{e}E(1-\beta_\mathrm{e}\cos\theta)$, where θ is the angle between the incident electron and the photon direction in the observer's frame Σ and $\beta_\mathrm{e} = \upsilon/c$. Thus, in the electron's rest frame Σ', the photon scatters with an energy $E' \simeq \gamma_\mathrm{e}E$ for all but very small angles. If the photon has negligible energy in Σ', i.e. $E' \ll m_\mathrm{e}c^2$, the interaction can be treated in the *Thompson limit* which is characterized by elastic scattering of the photon: $E'_1 \simeq E'$. After transforming back into the observer's frame, using $E_1 = \gamma_\mathrm{e}E'_1(1+\beta_\mathrm{e}\cos\theta'_1)$, the energy of the scattered photon in Σ is given by $E_1 \simeq \gamma_\mathrm{e}^2E(1-\beta_\mathrm{e}^2\cos^2\theta)$. After averaging an isotropic distribution of relativistic electrons over angle, the energy of the inverse Compton scattered photon is therefore increased by[4]

  $$\frac{\langle \Delta E_\gamma \rangle}{E_\gamma} = \frac{4}{3}\gamma_\mathrm{e}^2 \quad \text{for} \quad p \gtrsim m_\mathrm{e}c, \tag{4.202}$$

  which can move the microwave CMB photon to X-ray and γ-ray energies for highly relativistic electrons, $\gamma_\mathrm{e} \gtrsim 10^3$.

- So far, we have discussed the two limiting cases of Compton interactions with non-relativistic and fully relativistic electrons. However, non-thermal electron populations are typically described by power-law spectra that span from non-relativistic to relativistic momenta, with different particle energies dominating

---

[4] Note that this approximation is only valid if the fraction of energy lost by the electron in a single collision $\gamma_\mathrm{e}^2E/(\gamma_\mathrm{e}m_\mathrm{e}c^2) \ll 1$.

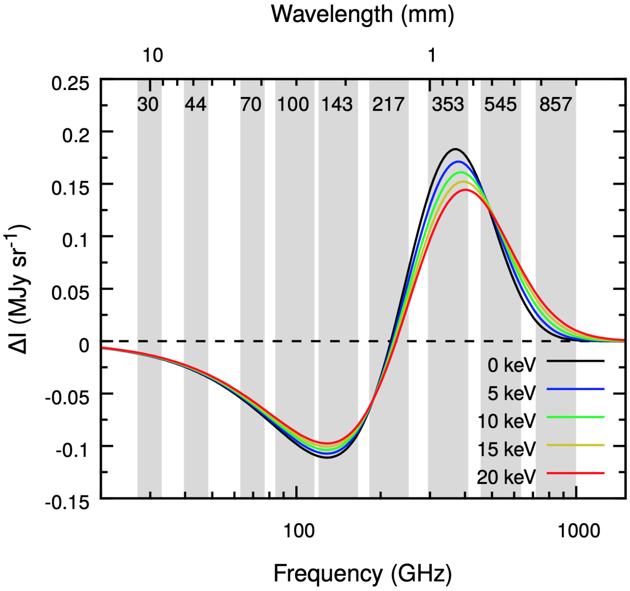


Figure 4.17: Spectrum of the thermal SZ effect with relativistic corrections for a range of electron temperatures at fixed $y = 10^{-4}$. The grey bands indicate the nine *Planck* frequency channels with $\Delta\nu/\nu = 0.2$ for the three low-frequency instruments and $\Delta\nu/\nu = 0.3$ for the six high-frequency channels (Erler 2018).

the non-thermal pressure. If we imagine increasing the characteristic particle energy from the non-relativistic to the relativistic regime, the characteristic crossover frequency ν_c would continuously shift from 217 GHz to higher frequencies. Accordingly, the amplitude of the SZ effect would drop as a result of photon number conservation as photons are redistributed across a larger energy range, see Fig. 4.17.

- These considerations enable us to cast the different inverse Compton processes that modify the CMB spectrum as a result of the SZ effects into the following compact form:

$$\begin{aligned}\delta i(x) &= g(x)\,y\,[1+\delta(x,T_\mathrm{e})] - h(x)\,w \\ &\quad + [j(x) - i(x)]\,\tau_\mathrm{rel}\,, \end{aligned} \tag{4.203}$$

  where the first two terms account for the thermal and kinematic SZ effects, respectively (as discussed in Section 4.4.1) and the third term describes the relativistic SZ effect. For non-relativistic electrons the relativistic correction term to the thermal SZ effect is zero, $\delta(x, T_\mathrm{e}) = 0$, but for hot clusters even the thermal electrons have relativistic corrections, which will modify the thermal SZ effect (see Fig. 4.17).

- The third term in Eq. (4.203) takes account of Compton scattering with relativistic electrons that exhibit an optical depth of

$$\tau_\mathrm{rel} = \sigma_\mathrm{T} \int \mathrm{d}l\, n_\mathrm{e,rel}. \tag{4.204}$$

The flux scattered to other frequencies is $i(x)\tau_{\rm rel}$ while $j(x)\tau_{\rm rel}$ is the flux scattered from other frequencies to $x = h\nu/(k_{\rm B}T)$. It is worth noting, that in the limit of ultra-relativistic electrons and for $x < 10$, one can neglect the flux scattered from other frequencies to x, because $j(x) \ll i(x)$. In the following, we drop this approximation and consider the general case.

- The scattered flux can be expressed in terms of the photon redistribution function for a mono-energetic electron distribution $P(s, p)$, where the frequency of a scattered photon is shifted by a factor s:

  $$j(x) = \int_0^\infty {\rm d}s \int_0^\infty {\rm d}p\, f_{\rm e}(p)\, P(s,p)\, i(x/s). \tag{4.205}$$

  For a given electron spectrum $f_{\rm e}(p)\,{\rm d}p$ with the normalized electron momentum $p = \beta_{\rm e}\gamma_{\rm e}$ and $\int {\rm d}p\, f_{\rm e}(p) = 1$, this redistribution function can be derived following the kinematic considerations of the inverse Compton scattering in the Thomson regime, where $\gamma_{\rm e}\, h\nu \ll m_{\rm e}c^2$. We use the compact formula for the photon redistribution function which was derived by Enßlin & Kaiser (2000):

  $$\begin{aligned} P(s,p) &= -\frac{3|1-s|}{32p^6 s}\left[1 + (10 + 8p^2 + 4p^4)s + s^2\right] \\ &\quad + \frac{3(1+s)}{8p^5}\left\{\frac{3 + 3p^2 + p^4}{\sqrt{1+p^2}}\right. \\ &\quad \left. - \frac{3 + 2p^2}{2p}\left[2\,{\rm arcsinh}(p) - |\ln(s)|\right]\right\}. \end{aligned} \tag{4.206}$$

  The allowed range of frequency shifts is restricted to

  $$|\ln(s)| \le 2\,{\rm arcsinh}(p)\,, \tag{4.207}$$

  and thus $P(s, p) = 0$ for $|\ln(s)| > 2\,{\rm arcsinh}(p)$.

- The spectral distortions owing to the relativistic SZ effect can be rewritten to include a relativistic Comptonization parameter $\tilde{y}$,

  $$\delta i_{\rm rel}(x) = [j(x) - i(x)]\tau_{\rm rel} = \tilde{g}(x)\tilde{y}, \tag{4.208}$$

  where

  $$\begin{aligned} \tilde{y} &= \frac{\sigma_{\rm T}}{m_{\rm e}c^2}\int {\rm d}l\, n_{\rm e}\, k\tilde{T}_{\rm e}\,, && (4.209) \\ k\tilde{T}_{\rm e} &= \frac{P_{\rm e}}{n_{\rm e}}\,, && (4.210) \\ P_{\rm e} &= \frac{m_{\rm e}c^2}{3}\int_0^\infty {\rm d}p\, f_{\rm e}(p)\beta_{\rm e}\, p, && (4.211) \\ \tilde{g}(x) &= [j(x) - i(x)]\tilde{\beta}(k\tilde{T}_{\rm e})\,, && (4.212) \\ \tilde{\beta}(k\tilde{T}_{\rm e}) &= \frac{m_{\rm e}c^2}{\langle k\tilde{T}_{\rm e}\rangle} = \frac{m_{\rm e}c^2 \int {\rm d}l\, n_{\rm e}}{\int {\rm d}l\, n_{\rm e} k\tilde{T}_{\rm e}}\,, && (4.213) \end{aligned}$$

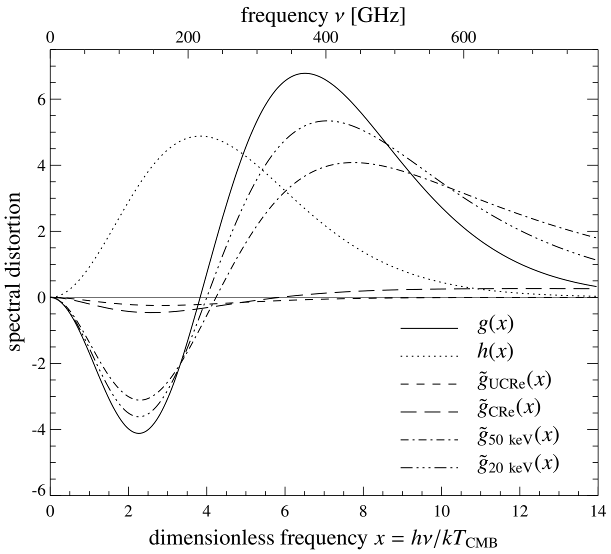


Figure 4.18: Spectral distortions due to the thermal SZ effect, $g(x)$, kinematic SZ effect, $h(x)$, and relativistic SZ effect, $\tilde{g}(x)$. The relativistic SZ effect is different for different relativistic populations: ultra-relativistic cosmic ray electrons (UCRe), power-law cosmic ray electrons (CRe, extending from trans- to fully relativistic energies), and trans-relativistic thermal electrons with temperatures of $kT_{\mathrm{e}} = 20$ keV and 50 keV, respectively (Pfrommer et al. 2005). This shows that higher energetic electrons are able to scatter CMB photons to larger frequencies which causes larger spectral distortions.

where $\beta_{\mathrm{e}} = \upsilon/c = p/\sqrt{1+p^2}$ is the dimensionless velocity of the (relativistic) electron. Here, we introduced the pseudo-thermal beta-parameter $\tilde{\beta}(k\tilde{T}_{\mathrm{e}})$ and the pseudo-temperature $k\tilde{T}_{\mathrm{e}}$, which are both equal to its thermodynamic analog in the case of a thermal electron distribution.

- The frequency dependence of the various SZ effects is shown in Fig. 4.18. In particular, we compare the spectral distortions due to the thermal SZ effect, $g(x)$, the kinematic SZ effect, $h(x)$, and the relativistic SZ effect, $\tilde{g}(x)$, for various scenarios of relativistic electron populations. As we can clearly see, the more energetic the electrons, the larger is the cross-over frequency ν_{c}. However, the change in thermodynamic temperature (or photon intensity) of a specific SZ effect (thermal, kinetic, relativistic) is the product of spectral distortion (shown in Fig. 4.18) and amplitude of that specific effect, namely y, w, and τ_{rel} in Eq. (4.203).
- The amplitude of the relativistic SZ effect, τ_{rel}, is much smaller than that of the thermal and kinetic SZ effects. To see this, we compare the thermal Compton-y parameter to its relativistic analogue in Eq. (4.209),

$$\tilde{y} = \sigma_{\mathrm{T}} \int_0^L n_{\mathrm{CRe}} \frac{k_{\mathrm{B}} \tilde{T}_{\mathrm{CRe}}}{m_{\mathrm{e}} c^2} \mathrm{d}l \sim \sigma_{\mathrm{T}} \bar{n}_{\mathrm{CRe}} L \frac{k_{\mathrm{B}} \tilde{T}_{\mathrm{CRe}}}{m_{\mathrm{e}} c^2}. \qquad (4.214)$$

In steady state, CR electrons with kinetic energies of 0.1 GeV dominate the energy spectrum (see Fig. 4.33) so that $\tilde{T}_{\mathrm{CRe}}/T_{\mathrm{e}} \sim 0.1\,\mathrm{GeV}/10\,\mathrm{keV} \sim 10^4$. The CR electron pressure in units of the thermal pressure reads

$$\frac{P_{\mathrm{CRe}}}{P_{\mathrm{th}}} = \frac{n_{\mathrm{CRe}} k_{\mathrm{B}} \tilde{T}_{\mathrm{CRe}}}{n_{\mathrm{e}} k_{\mathrm{B}} T_{\mathrm{e}}} = \frac{P_{\mathrm{CRe}}}{P_{\mathrm{CRp}}} \frac{P_{\mathrm{CRp}}}{P_{\mathrm{th}}} = K_{\mathrm{ep}} X_{\mathrm{CR}} \lesssim 10^{-4}, \quad (4.215)$$

where we adopted the Milky Way value $K_{\mathrm{ep}} \sim 10^{-2}$ (which indicates the electron-to-proton acceleration efficiency at shocks) and $X_{\mathrm{CR}} \lesssim 10^{-2}$ (because otherwise, we would have observed clusters at gamma rays as a result of hadronic CR-proton interactions, see Section 4.5.1). Hence, the CR-to-thermal electron number density is $n_{\mathrm{CRe}}/n_{\mathrm{e}} \sim K_{\mathrm{ep}} X_{\mathrm{CR}} T_{\mathrm{e}}/\tilde{T}_{\mathrm{CRe}} \sim 10^{-8}$ and we obtain a ratio of the amplitudes of the kinetic to relativistic SZ effect of

$$\frac{w}{\tau_{\mathrm{rel}}} \sim \frac{n_{\mathrm{e}}}{n_{\mathrm{CRe}}} \frac{\upsilon_r}{c} \sim 10^5, \quad (4.216)$$

so that the amplitudes of thermal and relativistic SZ effects differ by a factor of about $y/\tau_{\mathrm{rel}} \sim 2 \times 10^6$.

4.4.3 Self-similar SZ Scaling Relation

- We review the expectations for $Y_{\mathrm{sph}} = Y D_{\mathrm{ang}}^2$ in the idealized case of a cluster in virial equilibrium to help understand how possible deviations from the self-similar Y_{sph}–M relation and the scatter about it may arise. We start with Eq. (4.198), which has been rewritten as

$$Y_{\mathrm{sph}} = \frac{\sigma_{\mathrm{T}}}{m_{\mathrm{e}} c^2} \int_0^{R_{200}} \mathrm{d}V P_{\mathrm{e}} = \frac{(\gamma - 1)\,\sigma_{\mathrm{T}}}{m_{\mathrm{e}} c^2}\, \tilde{x}_{\mathrm{e}}\, X\, \mu\, E_{\mathrm{gas}}, \quad (4.217)$$

  where $\tilde{x}_{\mathrm{e}} = n_{\mathrm{e}}/n_{\mathrm{H}} = (X + 1)/(2\,X) = 1.158$ is the electron-to-hydrogen number density fraction with a hydrogen mass fraction $X = 0.76$, $\mu = 4/(3X+1+4X\tilde{x}_{\mathrm{e}}) = 0.588$ denotes the mean molecular weight for a fully ionized medium of primordial abundance (see Appendix A.1 for a derivation), $\gamma = 5/3$ is the adiabatic index, and we assume equilibrium between the electron and ion temperatures.

- Next, we define the characteristic temperature of the halo as

$$k_{\mathrm{B}} T_{200} = \frac{G M_{200}\, \mu\, m_{\mathrm{p}}}{3 R_{200}} = \frac{\mu\, m_{\mathrm{p}}}{3} \left[10\, G\, M_{200}\, H_0\, E(z)\right]^{2/3}, \quad (4.218)$$

  where M_{200} and R_{200} is the virial mass and radius of the cluster, $H(z) = H_0 E(z)$ is the Hubble function where H_0 denotes it current day value, G is Newton's constant, and m_{p} is the proton rest mass.

This enables us to write the total thermal energy of the halo with Eq. (4.218) as

$$\begin{aligned} E_{\rm gas} &= \frac{3}{2} N_{\rm gas}\, k_{\rm B} T_{200} = (1 - f_*)\, f_{\rm b}\, f_{\rm c}\, \frac{G M_{200}^2}{2\, R_{200}} \\ &= (1 - f_*)\, f_{\rm b}\, f_{\rm c}\, \frac{G}{2} \left[\frac{800\, \pi\, \rho_{\rm cr}(z)}{3}\right]^{1/3} M_{200}^{5/3}, \end{aligned} \qquad (4.219)$$

where $N_{\rm gas}$ is the number of gas particles, $\rho_{\rm cr}(z)$ is the critical density of the universe, $f_{\rm b} = \Omega_{\rm b}/\Omega_{\rm m}$ is the cosmic baryon fraction, $f_* \lesssim M_*/M_{\rm b}$ is the stellar mass fraction within the halo and $f_{\rm c}$ is the correction factor for the fraction of missing baryons at a given overdensity.

- Substituting Eq. (4.219) into Eq. (4.198), we obtain the integrated Compton-y parameter within R_{200},

$$\begin{aligned} Y_{\rm sph} &= \frac{(\gamma - 1)\sigma_{\rm T}}{m_{\rm e} c^2} \tilde{x}_{\rm e} X \mu (1 - f_*) f_{\rm b} f_{\rm c}\, G \left[\frac{\pi}{3} 100 \rho_{\rm cr}(z)\right]^{1/3} M_{200}^{5/3} \\ &= 97.6\, h_{70}^{-1}\, {\rm kpc}^2 E(z)^{2/3} \left(\frac{M_{200}}{10^{15}\, h_{70}^{-1} {\rm M}_\odot}\right)^{5/3} \frac{\Omega_{\rm b}}{0.043} \frac{0.25}{\Omega_{\rm m}} \end{aligned} \qquad (4.220)$$

  In Eq. (4.220), we set $f_* = 0$, $f_{\rm c} = 0.93$ (which derives from non-radiative cluster simulations at R_{200}) and adopted currently favored cosmological parameters. This simple analytical expression for the $Y_{\rm sph} - M$ scaling relation allows one to explore the assumptions underlying its derivation. In particular, Battaglia et al. (2012a) test the influence of the assumptions of spherical gravitational potential, zero non-thermal pressure support, and constant $f_{\rm b}$ (and for simulation with star formation, constant f_*) at R_{200}, on the self-similarity of the SZ scaling relation.

- Figure 4.19 demonstrates that the simulated $Y_{\rm sph} - M$ relation (that includes AGN feedback) is consistent with current data from X-ray and SZ observations. However, at group scales, the simulations by Battaglia et al. (2012a) slightly overpredict the SZ flux due to the too high gas fractions, $f_{\rm gas} = M_{\rm gas}/M_{\rm tot}$, in the simulations compared to X-ray observations. This may signal the omission of potentially relevant physics in the simulations that governs $f_{\rm gas}$ or underestimate the action of AGN feedback on these mass scales.

- The $Y_{\rm sph}$ reported by SZ surveys for known clusters use an X-ray-derived estimate of the aperture size. This is useful because the cluster radii are typically poorly measured in SZ, and so the X-ray aperture fixes the SZ measurement along the otherwise degenerate aperture flux-aperture radius relation. However, this prior introduces correlations between the X-ray and SZ observations, which makes comparisons between these observations difficult to interpret.

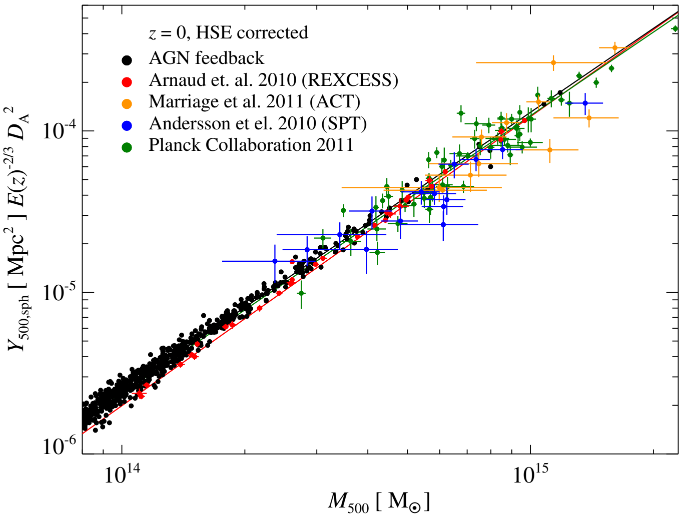


Figure 4.19: The $Y_{\mathrm{sph}} - M$ scaling relation for the AGN feedback simulations of Battaglia et al. (2012a) compared to recent X-ray results from Arnaud et al. (2010) and SZ results from ACT (Marriage et al. 2010), SPT (Andersson et al. 2010), and Planck (Planck Collaboration et al. 2011). Here, a 13% correction to the X-ray inferred hydrostatic cluster mass has been applied.

4.4.4 SZ Power Spectrum

- The angular power spectrum of the thermal SZ effect enables us to probe cosmological parameters associated with the growth of structure. First, we need to define several quantities.

- Fourier decomposition is not defined on the sphere. Instead, one has to project the temperature fluctuations onto another set of basis functions which are orthonormal on the sphere. These are the spherical harmonic functions $Y_\ell^m(\boldsymbol{\theta})$. If $\Delta T(\boldsymbol{\theta})/T$ is the relative temperature fluctuation of the thermal SZ effect at position $\boldsymbol{\theta}$ on the sky, it can be expanded into a series

$$\Theta_{\mathrm{tSZ}}(\boldsymbol{\theta}) \equiv \frac{\Delta T_{\mathrm{tSZ}}}{T}(\boldsymbol{\theta}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_\ell^m(\boldsymbol{\theta}). \qquad (4.221)$$

  with the (generally complex) coefficients $a_{\ell m}$.

- Because of the orthonormality of the spherical harmonics,

$$\int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\theta \sin\theta Y_{\ell_1}^{m_1 *}(\theta,\varphi)\, Y_{\ell_2}^{m_2 *}(\theta,\varphi) = \delta_{\ell_1 \ell_2} \delta_{m_1 m_2}\ , \qquad (4.222)$$

  the expansion coefficients are given by

$$a_{\ell m} = \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \mathrm{d}\theta \sin\theta\, \Theta_{\mathrm{tSZ}}(\theta,\phi) Y_\ell^m(\theta,\phi). \qquad (4.223)$$

- The power spectrum of the temperature fluctuation map $C_{\ell,\mathrm{tSZ}}$ is defined by
$$\left\langle a_{\ell_1 m_1} a^*_{\ell_2 m_2} \right\rangle \equiv \delta_{\ell_1,\ell_2}\delta_{m_1,m_2} C_{\ell,\mathrm{tSZ}}, \tag{4.224}$$
which depends only on the multipole order ℓ because of statistical isotropy. Conventionally, the quantity $\ell(\ell+1)C_\ell$ is shown instead of C_ℓ because it reflects the total power contained in the multipole moment ℓ.

- For sufficiently small angles on the sky (typically $\theta \lesssim 1^\circ$) the curvature of the sky becomes vanishingly small and we can Fourier expand the temperature fluctuations
$$\Theta(\boldsymbol{x}) = \int \frac{\mathrm{d}^2 k}{(2\pi)^2}\,\hat{\Theta}(\boldsymbol{k})\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \quad \hat{\Theta}(\boldsymbol{k}) = \int \mathrm{d}^2 x\,\Theta(\boldsymbol{x})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{4.225}$$
It is convenient to define the power spectrum of temperature fluctuations in the flat-sky approximation,
$$\langle\hat{\Theta}(\boldsymbol{k})\hat{\Theta}^*(\boldsymbol{k}')\rangle \equiv (2\pi)^2 C_{(\ell),\mathrm{tSZ}}\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}'). \tag{4.226}$$
For sufficiently small angles, there is an exact correspondence between the all-sky and flat-sky power spectra (see Appendix C of Hu 2000), so that we have
$$C_{\ell,\mathrm{tSZ}} = C_{(\ell),\mathrm{tSZ}}. \tag{4.227}$$

- Applying the integral representation of the Dirac delta distribution,
$$\int \mathrm{d}^2 x\, e^{\mathrm{i}(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}} = (2\pi)^2\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}'), \tag{4.228}$$
to the definition of the SZ power spectrum in the flat-sky approximation of Eq. (4.226) and substituting the expression for the change of thermodynamic temperature from the thermal SZ effect, Eq. (4.189), we can write down the SZ power spectrum of a single cluster,
$$C_{(\ell),\mathrm{tSZ}} = f(x)^2|\hat{y}_{(\ell)}(M,z)|^2, \tag{4.229}$$
where $f(x)$ is implicitly defined in Eq. (4.189) and $\hat{y}_{(\ell)}(M,z)$ is the cluster form factor, i.e., the two-dimensional Fourier transform of the Compton-y parameter.

- To calculate the SZ power spectrum of the full sky, we need to account for the distribution of clusters in mass and redshift. To this end, we use the halo formalism that accounts for (i) the one-halo contribution of the thermal SZ effect from individual halos, assuming a spatially Poisson distributed population of clusters and (ii) the halo-halo term that accounts for spatial correlation between clusters. Because this second term is subdominant on angular scales of interest here, $\theta < 1.2^\circ$ or multipole moments $\ell > 300$, we neglect this correlation term in the following.

- The one-halo contribution to the thermal SZ angular power spectrum at a multipole moment ℓ is the integral of the squared Fourier transform of the Compton-y parameter over cosmic volume and all halos of mass M that significantly contribute to the SZ power:

$$C_{(\ell),\mathrm{tSZ}} = f(x)^2 \int_0^{z_{\max}} \frac{\mathrm{d}V}{\mathrm{d}z}\mathrm{d}z \int_{M_{\min}}^{M_{\max}} \mathrm{d}M \frac{\mathrm{d}n(M,z)}{\mathrm{d}M} |\hat{y}_{(\ell)}(M,z)|^2, \tag{4.230}$$

where $z_{\max} = 1090$ and $\mathrm{d}n(M,z)/\mathrm{d}M$ is the halo mass function.

- Assuming that the electron pressure profile $P_\mathrm{e}(r)$ is spherically symmetric, we can determine the functional form of $\hat{y}_{(\ell)}(M,z)$ via

$$\hat{y}_{(\ell)} = \frac{1}{D_\mathrm{ang}^2} \int \mathrm{d}^3 r \frac{\sigma_\mathrm{T}}{m_\mathrm{e}c^2} P_\mathrm{e}(r) \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \tag{4.231}$$

$$= \frac{2\pi\sigma_\mathrm{T}}{m_\mathrm{e}c^2} \frac{1}{D_\mathrm{ang}^2} \int_0^\infty \mathrm{d}r r^2 \int_0^\pi \mathrm{d}\theta \, \sin\theta \, P_\mathrm{e}(r) \cos(kr\cos\theta) \tag{4.232}$$

$$= \frac{4\pi\sigma_\mathrm{T}}{m_\mathrm{e}c^2} \frac{1}{D_\mathrm{ang}^2} \int_0^\infty \mathrm{d}r r^2 P_\mathrm{e}(r) \int_0^1 \mathrm{d}\cos\theta \, \cos(kr\cos\theta). \tag{4.233}$$

Note that there should have been an imaginary contribution from the expansion of $\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}$ using Euler's formula, but the integral vanishes identically,

$$\int_0^\pi \mathrm{d}\theta \, \sin\theta \, \mathrm{i} \sin(kr\cos\theta) = \left.\frac{\mathrm{i}\cos(kr\cos\theta)}{kr}\right|_0^\pi = 0. \tag{4.234}$$

- To proceed, we explore the fact that the physical sizes of clusters, R, is small compared to their angular diameter distance, D_ang, so that we can use the small-angle approximation,

$$\frac{2\pi}{\ell} = \vartheta \approx \tan\vartheta = \frac{R}{D_\mathrm{ang}} = \frac{2\pi}{D_\mathrm{ang}k} \tag{4.235}$$

so that we can read off $\ell \approx D_\mathrm{ang}k$ which enables us to rewrite the argument of the cosine in Eq. (4.233),

$$rk = x_r r_\mathrm{s} \frac{\ell}{D_\mathrm{ang}} = x_r \frac{\ell}{\ell_\mathrm{s}}, \tag{4.236}$$

where we introduced a dimensionless radius $x_r = r/r_\mathrm{s}$ where the scale radius is defined via the NFW density profile, $r_\mathrm{s} = r_{200}/c_{200}$, where r_{200} and c_{200} are the virial radius and profile concentration parameter. The scale radius defines the multipole moment in Limber's approximation, $\ell_\mathrm{s} = D_\mathrm{ang}/r_\mathrm{s}$.

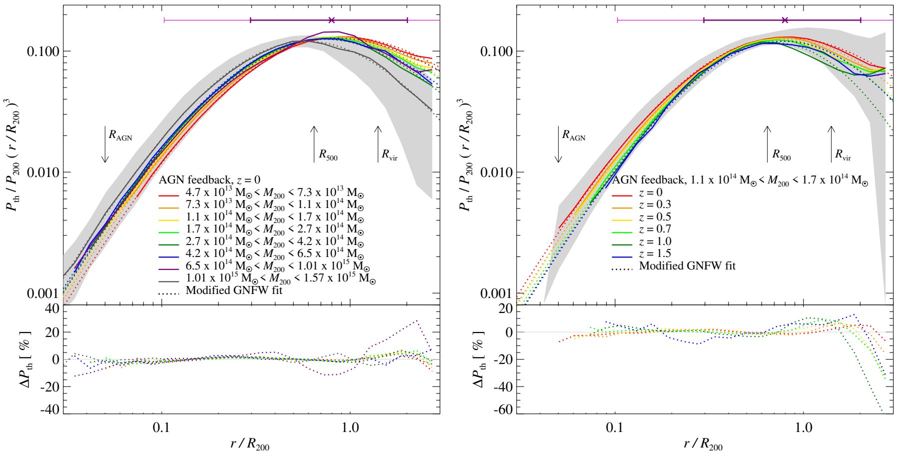


Figure 4.20: The normalized average pressure profiles and *parametrized fits* to these profiles from simulations with AGN feedback scaled by $(r/R_{200})^3$, in mass bins (left panel) and redshift bins (right panel). Here each mass and redshift bin has been independently fit. The grey band shows the standard deviation of the average cluster in the most massive bin (left) and lowest redshift bin (right). In both panels we illustrate the radii that contribute 68% and 95% of the total thermal energy, Y, centered on the median, by horizontal purple and pink error bars. The bottom panels show the percent difference between the fits and the average profiles (Battaglia et al. 2012b).

- Adopting these definitions, we can calculate the integral in $\cos\theta$ via

$$\int_0^1 \mathrm{d}\cos\theta\, \cos\left(\frac{\ell}{\ell_s}\, x_r \cos\theta\right) = \frac{\sin(\ell x_r/\ell_s)}{\ell x_r/\ell_s} \tag{4.237}$$

  and find

$$\hat{y}_{(\ell)} = \frac{4\pi\sigma_T}{m_e c^2} \frac{r_s}{\ell_s^2} \int_0^\infty \mathrm{d}x_r x_r^2 P_e(r) \frac{\sin(\ell x_r/\ell_s)}{\ell x_r/\ell_s}. \tag{4.238}$$

- Defining a virial analogue of the thermal pressure,

$$P_{200} \equiv \frac{GM_{200} 200\, \rho_{cr}(z) f_b}{2R_{200}}, \tag{4.239}$$

  we can fit stacked average thermal pressure profiles $\bar{P}_{th} = \langle P_{th}/P_{200}\rangle$ to a restricted version of the generalized NFW profile,

$$\bar{P}_{fit} = P_0\, (x_r/x_c)^\delta \left[1 + (x_r/x_c)^\alpha\right]^{-\beta},\ x_r \equiv r/R_{200}, \tag{4.240}$$

  where the fit parameters are a core-scale x_c, an amplitude P_0 and a power law index β for the asymptotic fall off of the profile. Because there is substantial degeneracy between fit parameters, fixing $\alpha = 1.0$ and $\delta = -0.3$ provides equally good fits. Figure 4.20 shows that generalized NFW profile fits the average profiles well in the majority of the mass and redshift bins, with deviations within $\sim 5\%$ of the mean. The upturns at large radii are due to contributions from nearby clusters and substructure.

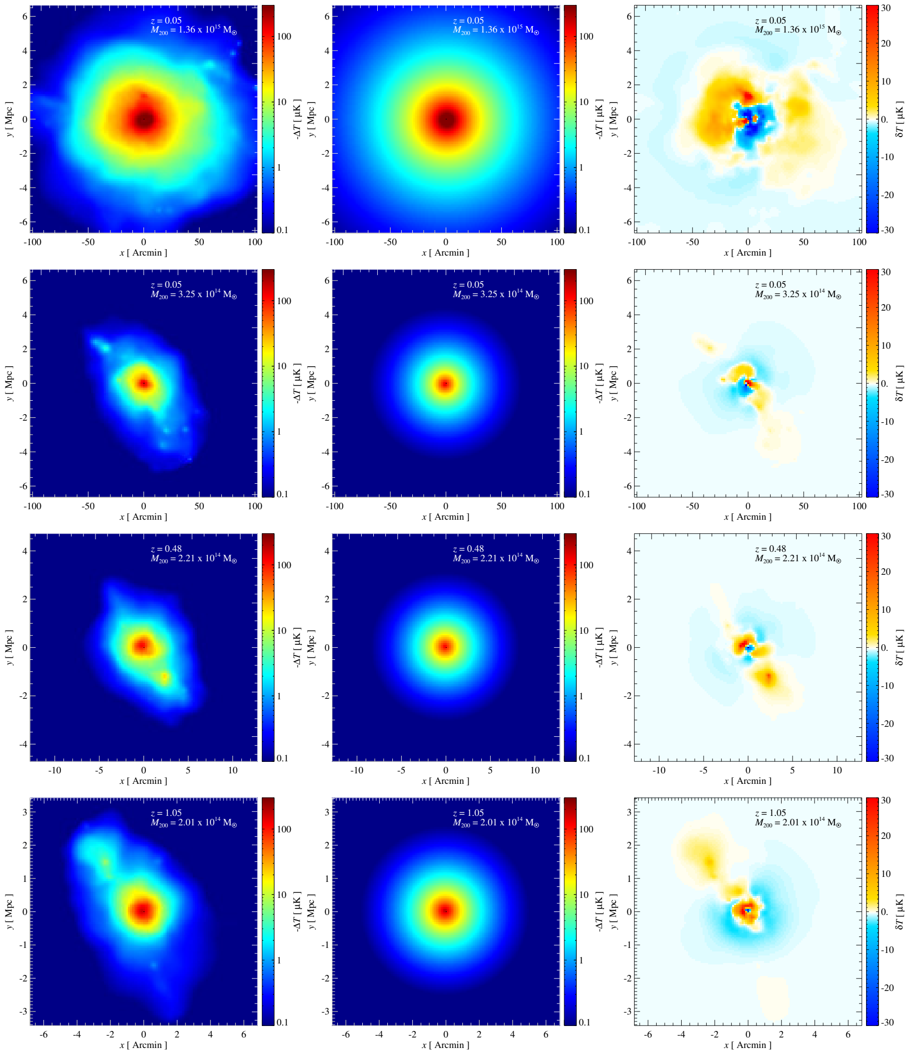


Figure 4.21: A comparison of four projected pressure maps of simulated clusters to the projected pasted-profile maps. From left to right, the panels show the simulated clusters (cut at a spherical radius of $6R_{500}$), the projected pasted profiles from the constrained fit, and the difference map between the two. The maps show the temperature decrement $-\Delta T$ in units of μK, at a frequency of 30 GHz. The difference maps, δT, illustrate the scales and amplitudes of the residuals between the simulated clusters and the projected pasted profiles. Note the color scale is logarithmic for the left two panels (from -0.1 μK to -300 μK), while it is linear for the difference map (from -30μK to 30μK). For all panels the left and top axes are in units of Mpc and the bottom and right axes are in units of arc-minutes (Battaglia et al. 2012b).

- Because of the systematic trend of the pressure profiles with mass and redshift, it is possible to derive a global fit to our pressure profiles as a function of mass and redshift. Treating each parameter as a separable function of mass and redshift gives good results, with the fit parameters constrained to be of the following form: for generic parameter A (representing the fit parameters P_0, β, and x_c), we have

  $$A = A_0 \left(\frac{M_{200}}{10^{14}\,\mathrm{M}_\odot} \right)^{\alpha_m} (1+z)^{\alpha_z} . \qquad (4.241)$$

  This defines a global empirical model for the average electron pressure as a function of cluster radius, redshift, and mass, which is referred to as the constrained pressure profile. Clearly, with fewer degrees of freedom, the constrained fits will naturally not be as accurate as fitting each mass/redshift bin completely independently, but the mean recovered profile is accurate to 10%.

- Figure 4.21 presents projected 30-GHz temperature maps of four sample clusters, their expected maps from the global constrained fit (referred to as "pasted" maps), and the errors in the predicted brightness temperature, which is within ~ 10% of the simulated cluster. Note that this is not a representative sample of clusters and instead shows clusters with different masses across different redshifts. As such, it illustrates the scales of the deviations from the constrained fit, primarily resulting from substructure and miscentering, since the cluster center of mass does not necessarily line up with the peak of the projected pressure. These substructures are significant on scales of tens of arc minutes for nearby massive clusters and scales of arc minutes for higher redshift clusters, corresponding to $\ell \sim 1000 - 10000$.

- Figure 4.22 compares different methods for analytically calculating the thermal SZ power spectrum shows the large influence of low-redshift clusters on the low-ℓ regime of the thermal SZ power spectrum.

- Figure 4.23 shows the contribution of different cluster masses (top) and redshifts (bottom) to the thermal SZ power spectrum. The differences between the simulations and the pasted profile maps result from the absence of substructure and asphericity in the pasted profile maps, which is larger for more massive clusters. The larger differences found between the analytical calculation and the simulations are the result of the mass catalog of the simulations having an excess of high mass clusters and deficit of lower mass cluster compared to the analytic mass function. The agreement between the pasted profile and simulation spectra is excellent below $\ell \sim 5000$ for all redshifts. On smaller scales,

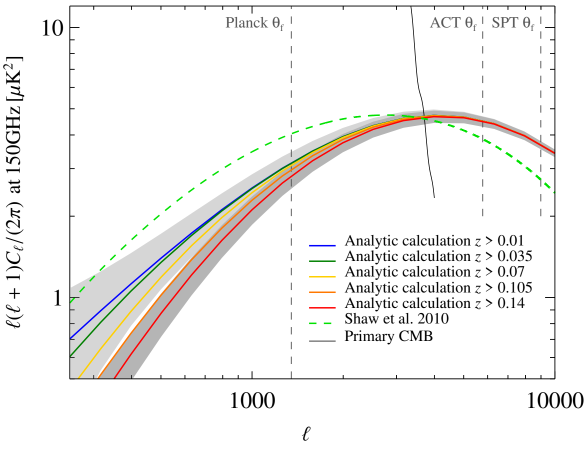


Figure 4.22: The figure shows a comparison between the current analytic calculations for the thermal SZ power spectra (Battaglia et al. 2012b; Shaw et al. 2010) and how the power spectrum changes with the variation of the lower redshift limit of integration. The cosmic variance of 1% of the full-sky power spectrum is illustrated by the grey bands for the highest and the lowest redshift limits of integration. The full-width half-max values for the Planck, ACT and SPT beams as well as the power spectrum of primary CMB fluctuations are also plotted.

cluster substructure contributes similarly across all redshift bins examined.

- Figure 4.24 (left) shows the contribution of cluster masses and redshifts to the thermal SZ power spectrum. At $\ell = 3000$, half the power of thermal SZ power spectrum comes from clusters with $z > 0.6$, and half comes from clusters with $M_{500} < 2 \times 10^{14} M_{\odot}$. The right panel of Fig. 4.24 shows the cumulative contribution of different cluster radii to the thermal SZ power spectrum. On small scales, virtually all of the power at $\ell > 2000$ comes from $r < 2R_{500}$. About 80% of the thermal SZ power is recovered at $\ell = 3000$ when tapering at R_{500}, though the deviations become larger at smaller ℓ. These results emphasize the importance of understanding cluster pressure profiles well past R_{500} in order to do high-precision work with the thermal SZ power spectrum.

4.4.5 SZ Effect of AGN Bubbles and Shocks

- The *Chandra* X-ray Observatory is finding a large number of cavities in the X-ray emitting ICM which often coincide with the lobes of the central radio galaxy. While radio synchrotron emission provides evidence for the existence of cosmic ray electrons and magnetic fields, the detailed composition of the plasma bubble governing its dynamics is still unknown because the cavities are consistent with no X-ray emission. Minimum energy or

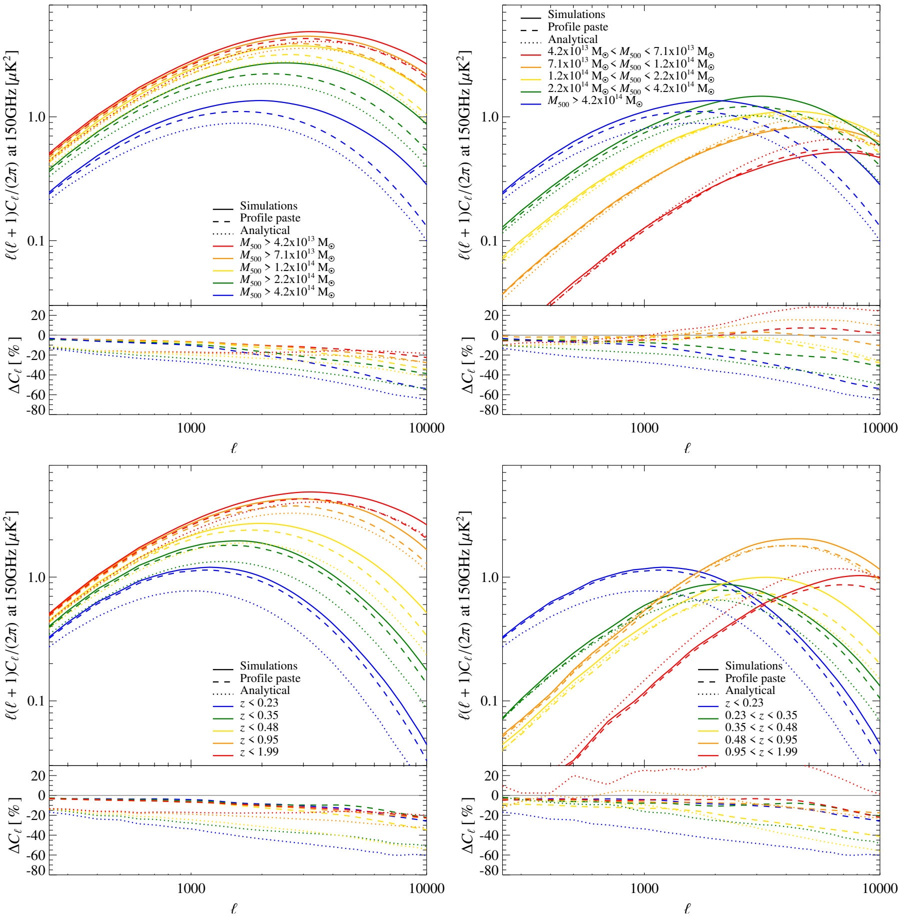


Figure 4.23: The thermal SZ power spectrum sorted into bins of galaxy cluster mass (top) and redshift (bottom). The left panels show the cumulative thermal SZ power spectrum in mass (redshift) bins from the AGN feedback simulations, the pasted profile maps and the analytical calculation. The right panels show the differential thermal SZ power spectrum. The bottom panels show the relative difference, $\Delta C_{(\ell)} = 100\left(C_{(\ell),\,\mathrm{sim}} - C_{(\ell),i}\right)/C_{(\ell),\,\mathrm{sim}}$, where $C_{(\ell),\,\mathrm{sim}}$ is the power spectrum of the simulated maps and $C_{(\ell),i}$ is that of the pasted profile maps and the analytical calculation, respectively (Battaglia et al. 2012b).

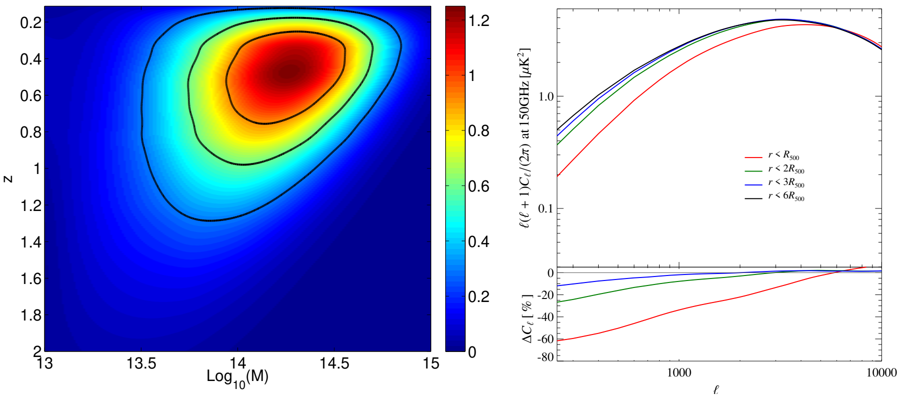


Figure 4.24: The left-hand panel shows the fractional contribution to the analytic SZ power spectrum at $\ell = 3000$ as function of redshift and $\log_{10} M_{500}$. The black contours contain 25, 50, and 75% of the SZ power at $\ell = 3000$. The right-hand panel shows the cumulative contribution of different cluster radii to the thermal SZ power spectrum for the AGN feedback simulations. The thermal pressure distribution has been tapered with an exponential function at varying cluster-centric radii before projection (Battaglia et al. 2012b).

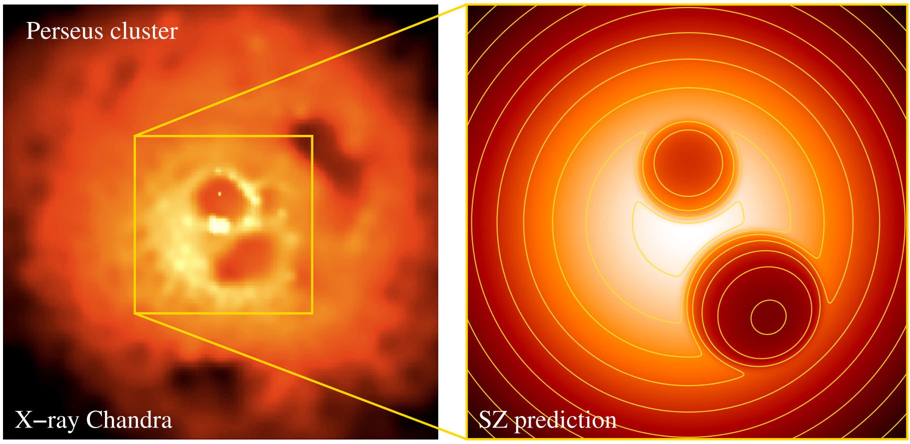


Figure 4.25: Central cool core region of the Perseus galaxy cluster. Left: Chandra X-ray map that shows two pairs of AGN bubbles that have been inflated by AGN jets. Right: synthetic 144-GHz *ALMA* observation of the SZ flux decrement of radio plasma bubbles, which assumes an ultra-relativistic electron population within the bubbles. The image is smoothed to the resolution of the *ALMA* compact core configuration (Pfrommer et al. 2005).

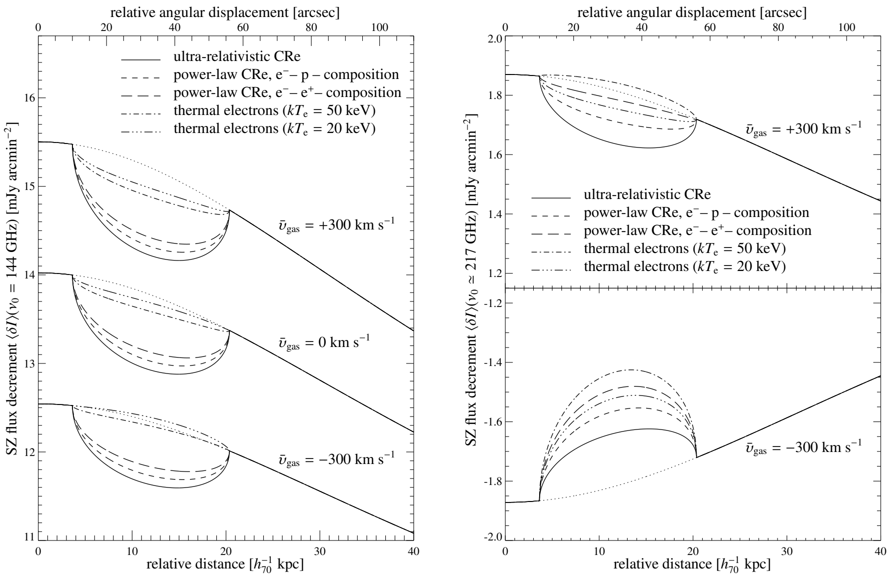


Figure 4.26: Unconvolved SZ flux decrement along an impact parameter through the center of the southern bubble of Perseus (see Fig. 4.25). The *left panel* shows the SZ flux decrement at the *ALMA* frequency band centered on ν_0 = 144 GHz while the *right panel* assumes a central fiducial frequency of $\nu_0 \simeq$ 217 GHz. Compared are five different scenarios of the composition of the plasma bubbles to the undisturbed SZ profile (dotted thin line), respectively. The three (two) set of lines correspond to three (two) differently assumed average bulk velocities along the line-of-sight, $\bar{\upsilon}_{\mathrm{gas}}$, of the thermal gas of Perseus (Pfrommer et al. 2005).

equipartition estimates of the non-thermal pressure in the radio bubbles give values which are typically a factor of ten smaller than the pressures required to inflate and maintain the bubbles as determined from the surrounding X-ray gas. This indicates that the standard minimum energy or equipartition radio arguments are missing the main component of the pressure and energy content of the radio lobes. Possibilities include magnetic fields, cosmic ray proton or cosmic ray electron power-law distributions, or very hot thermal gas. Solving this enigma would yield further insight into physical processes within cool cores, most importantly, how AGN feedback can heat and self-regulate the cooling ICM so that it balances the radiative cooling losses on average. Moreover, this knowledge could provide hints about the composition of relativistic outflows of radio galaxies because plasma bubbles represent the relic fluid of jets.

- Additionally, some of the clusters exhibit cavities in the X-ray emitting intra-cluster medium (ICM) without detectable high frequency radio emission. This category of X-ray cavities is also believed to be filled with radio plasma, but during the buoyant rise of the light radio plasma bubble in the cluster's potential the result-

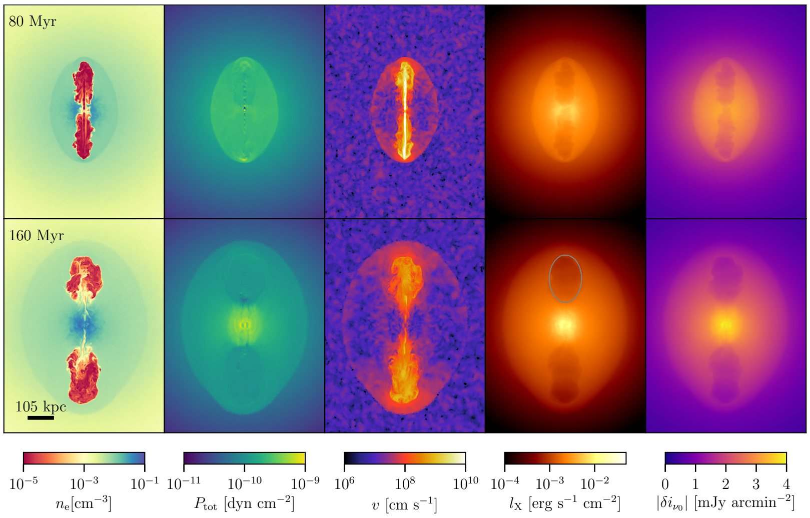


Figure 4.27: Simulation of the most powerful AGN jet outburst in the galaxy cluster MS0735. Shown is the electron density n_e, total pressure P_{tot} (including thermal, cosmic ray and magnetic pressure: $P_{tot} = P_{th} + P_{cr} + P_B$), velocity v, X-ray emissivity integrated along the line-of-sight l_X and the SZ signal $|\delta i_{\nu_o}|$ of the simulation at times 80 Myr and 160 Myr (top and bottom row, respectively). For the SZ effect, we assume a bubble filled with ultra-relativistic electrons and an observing frequency $\nu_0 = 30$ GHz. The jet terminates at 150 Myrs after which buoyantly rising bubbles form that can be observed as cavities in X-ray and SZ images (Ehlert et al. 2019).

ing adiabatic expansion and synchrotron/inverse Compton losses dwindles the observable radio emitting electron population producing a so-called *ghost cavity* or *radio ghost*. Possible entrainment of the ICM into the plasma bubble and subsequent Coulomb heating by cosmic ray electrons generates further uncertainty of the composition of the ghost cavity.

- However, high resolution SZ observations of radio bubbles and radio ghosts in clusters shed light on this puzzle, as the thermal SZ effect directly measures the thermal electron pressure in the gas. Such a measurement directly infers the composition of radio plasma bubbles and radio ghosts while indirectly obtaining indications for a specific underlying jet model (see Fig. 4.25).

- The left-hand panel of Fig. 4.26 shows the profile of the SZ flux decrement through the center of the southern bubble of Perseus at $\nu_0 = 144$ GHz for the different bubble fillings discussed in Fig. 4.18. The depth of the SZ cavity at this frequency is a measure how relativistic the respective electron population is, i.e. a deeper SZ cavity indicates a higher mean momentum of the electron population. A purely leptonic electron-positron filling would have twice the SZ flux in comparison to a electron-proton plasma. This enables us to distinguish a relativistic from a thermal electron population inside the bubble using only a single frequency

SZ observation by either a detection or non-detection of the bubble, respectively.

- The right-hand panel of Fig. 4.26 shows the SZ flux decrement at $\nu_0 = 217$ GHz, which enables to obtain an estimate for the cluster velocity with respect to the CMB rest frame. Thus, the combination of X-ray and SZ observations allows one to circumvent the degeneracy between the effects of the bubble composition and of the bubble extent along the line-of-sight on the SZ measurement.

- The SPT collaboration observed the AGN cavities in the galaxy cluster MS0735, which hosts the most powerful AGN jet outburst ever observed, and found that the cavities have very little SZ-contributing material, if any at all. This suggests a lobe pressure support of diffuse thermal plasma with temperature in excess of several hundreds of keV, or non-thermal relativistic particle populations. Equivalently, this demonstrates that the AGN jet-driven bubbles have a much lower density in comparison to the ambient ICM, with $\rho_{\rm jet}/\rho_{\rm ICM} \gtrsim 10^{-3}$–$10^{-2}$, and likely even smaller, in particular during jet launching. Figure 4.27 shows a simulation of this cluster in which the initial jet inflates lobes which, after termination of the jet, rise buoyantly in the cluster atmosphere. The jet initially drives a shock wave into the ICM. Because the Mach number in the jet direction exceeds that perpendicular to the jet, an ellipsoidal shock emerges that surrounds the lobes and can be clearly seen in the X-ray and SZ maps. The trailing contact discontinuity between jet fluid (red) and shocked ICM (light green-blue) is clearly visible in the electron number density $n_{\rm e}$ maps (left-most panels).

- The SZ effect is able to measure the most powerful merger shocks that have escaped X-ray observations either because they occur in the more dilute cluster outskirts where the X-ray flux has dropped too much to be observable or because a shock has increased the temperature of the ICM to the point that the X-ray instrument response is inefficient to detect the emission. Hence, SZ observations are complementing the much higher resolution X-ray observations of the ICM.

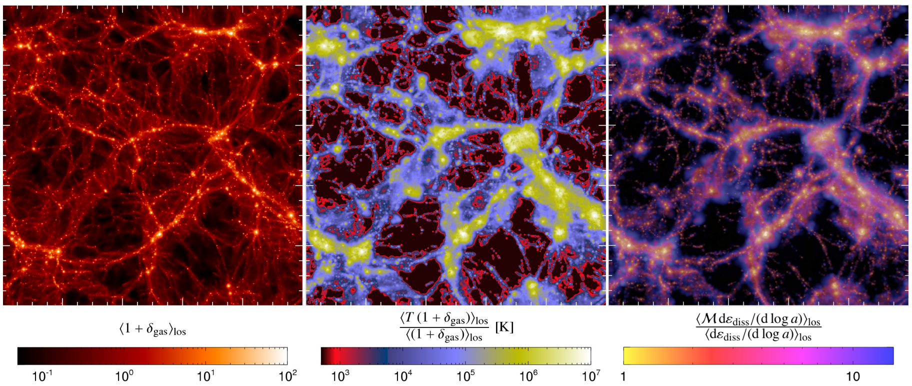


Figure 4.28: The large-scale structure of the universe assembles in form of the "cosmic web". From left to right, we show the projected gas density in a cosmological simulation (with side length $L = 100\,h^{-1}$ Mpc at $z = 0$), the gas temperature of the gravitationally heated intergalactic medium, and structure formation shocks (color coded by Mach number $\mathcal{M}$ and the brightness encoding the dissipated energy, Pfrommer et al. 2006).

4.5 Radio Emission: Shocks and Plasma Physics

4.5.1 Cosmological Shocks

- Cosmological shocks form abundantly during structure formation due to accreting pristine plasma onto cosmic filaments, sheets and halos, as well as due to supersonic flows associated with merging substructures (see Fig. 4.28). Additionally, shocks are driven by supernova explosions or propagating AGN jets in the interstellar and intracluster media. Structure formation shocks propagate through the cosmic tenuous plasma, which is compressed at the shock transition layer while most kinetic energy of the incoming plasma is dissipated into internal energy of the post-shock gas (and the remainder is used to accelerate particles to relativistic energies and to amplify magnetic fields). Because of the large collisional mean free path, the energy transfer proceeds through collective electromagnetic viscosity which is provided by unstable electromagnetic waves driven by drifting charged particles.

- Galaxy cluster mergers are the most energetic events in the Universe (after the Big Bang), which dissipate a gravitational energy of $E_{\mathrm{pot}} \sim GM_{\mathrm{cl}}^2/r_{\mathrm{cl}} \sim 3\times10^{64}$ erg on the free-fall time $t_{\mathrm{ff}} \sim (G\bar{\rho})^{-1/2} \sim 2$ Gyr, where $\bar{\rho} \sim 10^3\rho_{\mathrm{cr}}\Omega_{\mathrm{m}} \sim 3\times10^{-27}$ g cm^{-3} is the mean cluster density (see Section 1.2.6). Thus, despite the enormous energies dissipated in cluster mergers, the associated luminosities are not extraordinary large, amounting to $L_{\mathrm{merger}} \sim E_{\mathrm{pot}}/t_{\mathrm{ff}} \sim 5\times10^{47}$ erg s^{-1}.

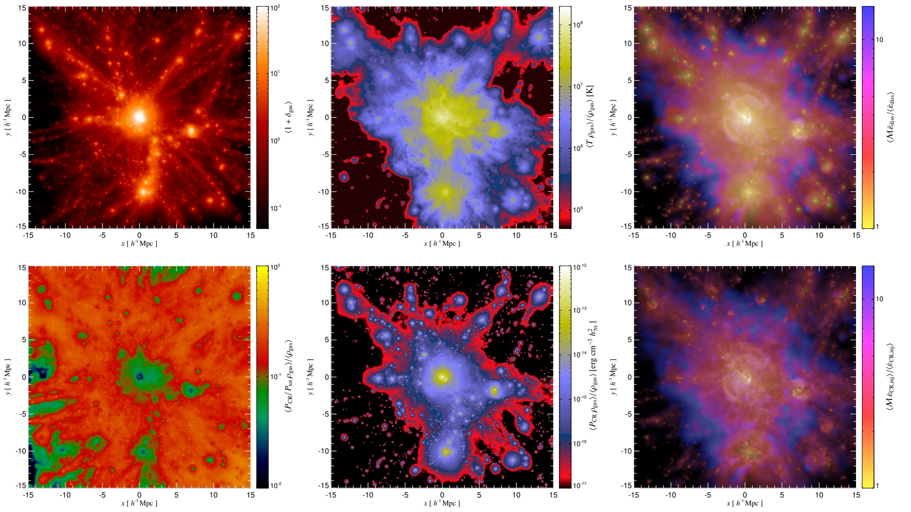


Figure 4.29: Zoom-in simulation of a galaxy cluster of mass $M_{200} = 1.6 \times 10^{15}\,\mathrm{M}_\odot$ in a cosmological setting. The top row shows from left to right the projected gas density, the temperature of the gravitationally heated gas, and structure formation shocks (color coded by Mach number $\mathcal{M}$ weighted by the dissipated energy and the brightness encoding the dissipated energy). The bottom row shows from right to left the shock Mach numbers weighted by the energy of cosmic rays (CRs) accelerated at these shocks, the mass-weighted CR pressure and the CR-to-total gas pressure ratio (Pfrommer et al. 2008).

- Astrophysical collisionless shocks can (i) dissipate kinetic into thermal energy as well as exchange energy between electrons and ions, (ii) accelerate electrons and ions to relativistic energies via diffusive shock acceleration (see Section 3.3.3.2), giving rise a population of cosmic rays (CRs), and (iii) amplify magnetic fields (or generate them from scratch). Efficient particle acceleration at collisionless shocks requires the sonic Mach number to exceed a critical one, $\mathcal{M} \gtrsim 3$ and the upstream magnetic field to point approximately into the direction of the shock normal, which defines a quasi-parallel shock. Before we discuss diffuse radio sources in clusters that these relativistic electrons give rise to, we review the properties of cosmological shocks.

- Figure 4.29 shows a zoom-in simulation of a massive galaxy cluster that has experienced a merger in the recent past as well as two smaller clusters to the lower right. We see that the massive galaxy cluster lies at the node of the cosmic web and contains gas in excess of $k_\mathrm{B}T = 10^8$ K inside the virial radius of $R_{200} = 2.4$ Mpc, which is heated upon passing through the accretion shock. Outside the ICM, we see a supercluster region that is not virialized but was nevertheless heated by structure formation shocks to temperatures of $k_\mathrm{B}T \sim 10^5$–10^7 K, and which forms the warm-hot intergalactic medium that is thought to host most of today's baryons. We identify the strongest shocks (largest Mach numbers) at the borders between cosmic voids and filaments/the supercluster region where the shocks encounter the lowest pre-shock sound speeds, $c_\mathrm{s} \propto (k_\mathrm{B}T)^{1/2}$. However, there is comparatively little energy dissipated in these shocks (that propagate with low velocities sourced by the shallow gravitational potentials in these regions and encounter comparatively dilute structures) as is evident from the expression of the kinetic energy density flux, $\rho v^3/2$. Contrarily, most energy is dissipated within dense regions by weak (small-Mach number) shocks because of the deep gravitational potentials that lead to large velocities and high temperatures of the virialized plasma, which make it difficult to exceed the sound speed by a significant amount.

- Because these weak shocks are very inefficient particle accelerators, we expect strong external shocks to mostly contribute to the generation of CRs. This expectation is confirmed by weighting the shock Mach numbers with the injected CR energy so that only strong shocks (color coded in blue and purple in Fig. 4.29) can be seen in the external regions and in projection onto the cluster. However, while CRs are accelerated at strong cosmological shocks, they are transported into the clusters and form a distribution that resembles that of the gas density. The CR-to-total pressure ratio decreases towards the cluster center to values of order a few per cent because of adiabatic compression. Imagine that CRs are accelerated at strong external shocks with an efficiency of $X_\mathrm{cr,0} = P_\mathrm{cr,0}/P_\mathrm{th,0} = 0.1$. Adiabatic compression of the CR-to-total pressure ratio across a density compression ratio of $r = 10^3$ yields

$$X_\mathrm{cr} = \frac{P_\mathrm{cr}}{P_\mathrm{th}} = \frac{P_\mathrm{cr,0}}{P_\mathrm{th,0}} \frac{r^{\gamma_\mathrm{cr}}}{r^{\gamma_\mathrm{th}}} = X_\mathrm{cr,0} r^{\gamma_\mathrm{cr}-\gamma_\mathrm{th}} = 0.01,$$

  where we adopted the adiabatic indices $\gamma_\mathrm{cr} = 4/3$ and $\gamma_\mathrm{th} = 5/3$. These small values of X_cr are consistent with non-observations of the γ-ray emission from galaxy clusters that would be produced by the hadronic CR proton-proton reaction: $\mathrm{CRp} + \mathrm{p} \rightarrow \pi^0 \rightarrow 2\gamma$ (which also produces secondary electrons, positrons and neutrinos from the decay of charged pions that are generated alongside the hadronic reaction).

- Figure 4.30 shows the cosmological shock statistics, namely the distribution of shock Mach numbers weighted by dissipated energy. The distribution shows an increasing amplitude and shifts to the left with time, i.e., more energy is dissipated at later times and the mean Mach number decreases with time. This simulations includes an instantaneous epoch of reionization at $z = 10$, which increases the mean temperature (and sound speed) of the intergalactic medium to $k_\mathrm{B}T \sim 10^4$ K and thus decreases the max-

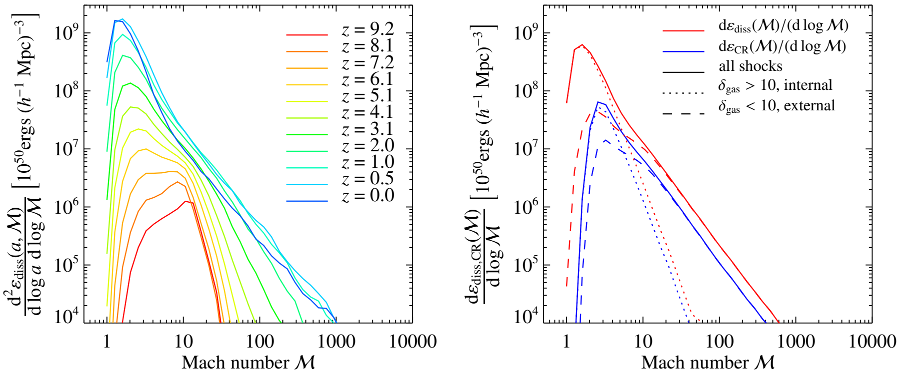


Figure 4.30: The left-hand panel shows the differential Mach number distribution $d^2\varepsilon_{\rm diss}(a,\mathcal{M})/(d\log a\, d\log\mathcal{M})$ of a non-radiative cosmological simulation for different redshifts (color coded). The right-hand panel shows this distribution integrated over the scale factor, $d\varepsilon_{\rm diss}(\mathcal{M})/(d\log\mathcal{M})$ (red) in addition to the Mach number distribution weighted by the injected CR energy rate, $d\varepsilon_{\rm cr}(\mathcal{M})/(d\log\mathcal{M})$ (blue). Internal shocks are shown with dotted lines and external shocks with dashed lines (Pfrommer et al. 2006).

imal Mach number. The effect of a cosmological constant slowing down structure formation is visible at late times in form of a lower amplitude of the high-Mach number tail at $z = 0$. This figure shows that also statistically, more energy is dissipated in weak shocks internal to collapsed structures in comparison to external strong shocks (which are efficient particle accelerators). This explains why the injected CR energy within clusters only makes up a small fraction of the total dissipated energy.

- Figure 4.31 shows the cycling of energy and mass among the different reservoirs in galaxy clusters: the thermal plasma, CRs and stars. Supernova driven shocks within the interstellar medium, and shocks in the ICM driven by AGN jets and resulting from cluster mergers and cosmic accretion of pristine gas dissipate kinetic energy and are able to accelerate CRs to form power-law momentum spectra that extend far into relativistic energies (see Section 3.3.3.2). The various processes shown in red redistribute the energy spatially or among the different reservoirs.

- The various observational windows indicated in yellow in Fig. 4.31 help characterizing these different populations and thus may elucidate the underlying physical processes. While the X-ray emission and the Sunyaev-Zel'dovich effect probe the thermal plasma, optical galaxy spectra directly probe the stellar populations of intra-cluster galaxies and indirectly the cluster's potential through their velocity dispersion. Of particular relevance for deciphering the CR population is the hadronic reaction with gas protons: provided the CR momentum exceeds the threshold

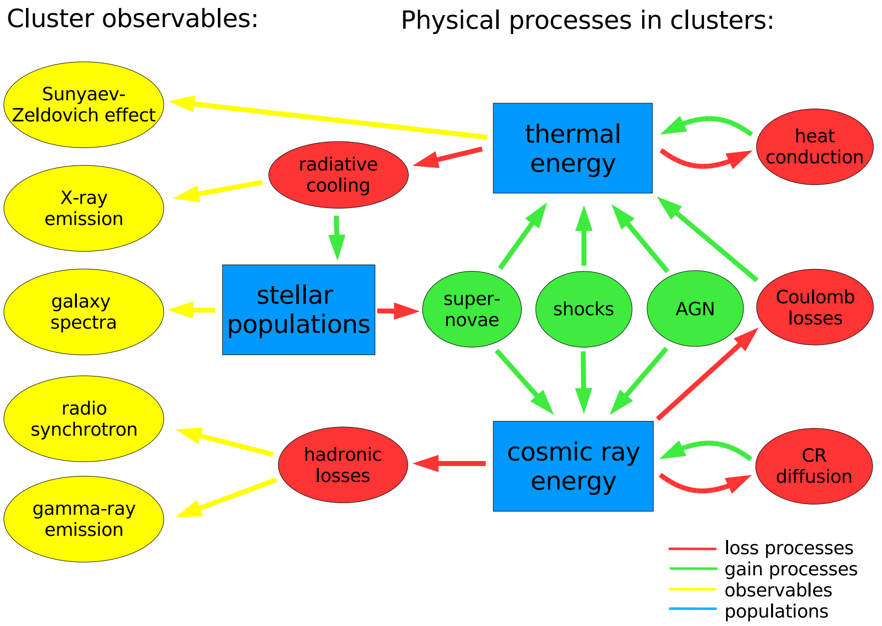
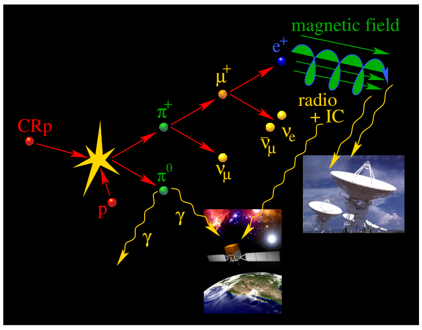


Figure 4.31: Overview over the relevant physical processes in galaxy clusters. The right side shows the interplay of different physical processes highlighting the interplay of the energy reservoirs of the thermal plasma and CR protons (shown in blue) while the left side shows observables (yellow) that inform about the properties of clusters and their dynamical state. Gain processes are denoted in green, while loss or redistribution processes are denoted in red (Pfrommer et al. 2008).

$p \simeq 0.8\,\mathrm{GeV}/c$ of this interaction, they produce pions which decay into secondary electrons, positrons, neutrinos, and γ rays:

$$\begin{aligned}\pi^\pm &\rightarrow \mu^\pm + \nu_\mu/\bar{\nu}_\mu \rightarrow e^\pm + \nu_e/\bar{\nu}_e + \nu_\mu + \bar{\nu}_\mu\\ \pi^0 &\rightarrow 2\gamma\,.\end{aligned}$$

Only CR protons above this kinematic threshold are therefore visible through their decay products via radiative processes, making them directly observationally detectable.



Visualization of the hadronic CR-proton reaction.

- In the high-energy limit, the hadronic reaction has a constant pion multiplicity $\xi(E_\mathrm{p}) \simeq \xi = 2$. This reflects the fact that two leading pion jets are leaving the interaction site in direction of the incident protons diametrically and carrying the high longitudinal momenta owing to Lorentz contraction of the interacting nuclei in the center of mass system and Heisenberg's uncertainty relation. Thus, the energy dependence of the mean pion energy is given by

$$\langle E_\pi \rangle (E_\mathrm{p}) \simeq K_\mathrm{p} \frac{E_\mathrm{p}}{\xi} \simeq \frac{E_\mathrm{p}}{4}, \qquad (4.242)$$

where we adopted the inelasticity $K_\mathrm{p} \approx 0.5$ of the hadronic reaction. Considering the decay kinematics, we can derive the mean

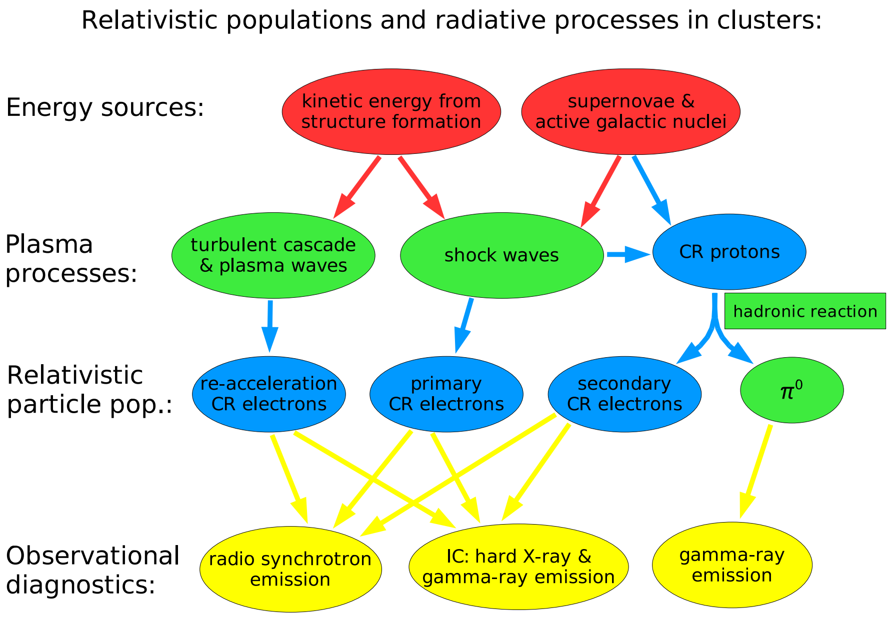


Figure 4.32: Schematic overview over non-thermal radiative processes in galaxy clusters. Various gravitational and non-gravitational energy sources (shown in red) are able to accelerate relativistic particle populations (shown in blue) by means of different plasma processes (shown in green). Non-thermal cluster observables (shown in yellow) are tracers of these CR populations: any CR electron population can emit radio synchrotron radiation as well as inverse Compton emission that extends from the X-ray into the γ-ray regime. In contrast, the characteristic spectral signature accompanying γ-ray emission from hadronic CR interactions is a unique sign of a CR proton population in the intra-cluster plasma (Pfrommer et al. 2008); but this has not yet been observed.

gamma-ray and secondary electron energies, respectively,

$$\langle E_\gamma \rangle = \frac{1}{2}\langle E_{\pi^0} \rangle \simeq \frac{1}{8}\,E_\mathrm{p}, \quad \text{and} \tag{4.243}$$

$$\langle E_{\mathrm{e}^\pm} \rangle = \frac{1}{4}\langle E_{\pi^\pm} \rangle \simeq \frac{1}{16}\,E_\mathrm{p}. \tag{4.244}$$

4.5.2 Radiative Processes and Cooling Times

- Figure 4.32 provides an overview of the various relativistic particle populations and radiative processes in galaxy clusters. Structure formation and AGN jet shocks can directly accelerate *primary CR electrons and protons*. The hadronic reaction of CR protons with the ambient gas protons produces *secondary relativistic electrons and positrons*. These two CR electron populations cool at high energies by synchrotron emission in the ubiquitous intra-cluster magnetic fields and via inverse Compton (IC) interactions with cosmic microwave background (CMB) photons to Lorentz factors $\gamma \sim 100$–300. This makes them invisible for our accessible observational windows in the radio and at γ rays. Continuous in-situ re-acceleration of these 'mildly' relativistic electrons by means of interactions with magneto-hydrodynamic waves resulting from ICM turbulence give rise to a third population of *re-accelerated relativistic electrons*.

- All these three CR electron populations contribute to the observed radio synchrotron emission (which will be detailed below) and should Compton upscatter CMB photons into the X-ray and γ-ray regime. Since the distribution of magnetic field strengths within the ICM is not well known, radio synchrotron emission alone has limited predictive power. Unfortunately, the conceptually simpler IC emission is hard to observe because of the strong radiation background in the soft and hard X-ray regime. So far, clusters have not been observed at γ-ray energies with the exception of the center of the Virgo galaxy cluster. The emission site is spatially coincident with the giant elliptical galaxy M87. Imaging air Cherenkov telescopes detected a time-varying TeV γ-ray flux attributed to the AGN jet while there is a low-flux state reaching from GeV to TeV γ-ray energies, which may or may not signal pion-decay γ-ray emission resulting from hadronic CR-proton interactions. Observational searches for the spatial extent of the TeV γ-ray signal will be critical to prove this exciting possibility, which enables us to estimate the CR pressure contribution to the central region of the Virgo clusters.

- ***Hadronic cooling timescale.*** To quantify these considerations, we turn our attention to the cooling timescales of CR ions/protons. The hadronic cooling timescale above the kinematic threshold for pion production is

  $$t_{\mathrm{pp}} = \frac{1}{0.5\,\sigma_{\mathrm{pp}}\,n_{\mathrm{n}}\,\upsilon_{\mathrm{cr}}} \tag{4.245}$$

  where $\sigma_{\mathrm{pp}} = 32$ mbarn is the inelastic proton cross section with an inelasticity of $\simeq 0.5$, $n_{\mathrm{n}} = \rho/m_{\mathrm{p}}$ is the number density of target nucleons for the hadronic reaction (see Appendix A.1) and $\upsilon_{\mathrm{cr}} \approx c$ is the CR proton velocity.

- ***Coulomb cooling timescale of a CR ion in a thermal plasma.*** When a CR ion encounters the Coulomb field of electrons of the background plasma, it can be deflected. This deflection leads to a momentum transfer from the CR ion to the background electron, causing the ion to decelerate. The calculation is most conveniently performed within the center-of-momentum reference frame, which closely aligns with the rest frame of the CR ion in the fast ion limit. In this particular frame (which we denote by primed quantities), the electron is deflected by an angle θ'_{d} and experiences a perpendicular momentum transfer of $\Delta p'_{\mathrm{e}} = m_{\mathrm{e}} v'_{\mathrm{e}} \theta'_{\mathrm{d}}$ (in the non-relativistic regime). Since the momentum transfer takes place perpendicular to the laboratory-to-CR ion rest frame boost direction, we obtain $\Delta p'_{\mathrm{e}} = \Delta p_{\mathrm{e}}$ in the non-relativistic limit. Therefore, in the laboratory frame, the electron's energy gain is equivalent to the energy loss of the CR ion by the amount

$$\Delta E = \frac{(\Delta p'_{\mathrm{e}})^2}{2m_{\mathrm{e}}} = \frac{m_{\mathrm{e}}}{m_{\mathrm{i}}} \theta'^2_{\mathrm{d}} E, \tag{4.246}$$

where $E = \frac{1}{2} m_{\mathrm{i}} v_{\mathrm{i}}^2$ represents the energy of the CR ion in the laboratory frame.[5] The influence of the ion-ion scattering rate on the energy loss of the incoming CR ion is diminished due to the considerably greater inertia of the background ions compared to that of the background electrons. We define a critical impact parameter b_0 where the electron's kinetic energy equals its mean electrostatic potential energy in the rest frame of the CR ion,

$$b_0 = \frac{2Ze^2}{m_{\mathrm{e}} v_{\mathrm{i}}^2} = \frac{2r_0 c^2}{v_{\mathrm{i}}^2}. \tag{4.247}$$

In this expression, Ze and v_{i} are the charge and velocity of the CR ion, $r_0 = Ze^2/(m_{\mathrm{e}} c^2)$ represents the classical electron radius, m_{e} is the electron mass, and e is the elementary charge. Provided the impact parameter of the interaction is less than b_0, we observe infrequent large-angle scattering events. Since there are many more electrons at distances larger than b_0, small-angle deflections at large impact parameters up to the Debye length, which characterizes the scale at which the charge of a plasma particle is screened, dominate the Coulomb scattering rate by a factor of $2 \ln \Lambda$, where $\ln \Lambda \sim 35$–40 represents the Coulomb logarithm (see Section 3.2.5.2).

The timescale at which the average squared deflection angle $\langle \theta_{\mathrm{d}}^2 \rangle$ in Eq. (4.246) becomes approximately equal to one corresponds to the deflection time, denoted as $t_{\mathrm{d}}^{\mathrm{ei}}$. The Coulomb cooling timescale, $t_{\mathrm{Coul,i}}$, of a CR ion propagating through a plasma is thus given by the particle energy divided by the energy loss rate. In consequence, $t_{\mathrm{Coul,i}}$ is equal to the deflection timescale divided by the mean relative energy transfer to the ambient plasma (see Eq. 4.246):

$$\begin{aligned} t_{\mathrm{Coul,i}} &= \left.\frac{E}{|\dot{E}|}\right|_{\mathrm{Coul,i}} \approx t_{\mathrm{d}}^{\mathrm{ei}} \frac{m_{\mathrm{i}}}{m_{\mathrm{e}}} = \frac{m_{\mathrm{i}}}{m_{\mathrm{e}} n_{\mathrm{e}} v_{\mathrm{i}} \sigma_{\mathrm{ei}}} \\ &= \frac{m_{\mathrm{i}}}{m_{\mathrm{e}} n_{\mathrm{e}} v_{\mathrm{i}} \pi b_0^2 2 \ln \Lambda} = \frac{m_{\mathrm{i}} v_{\mathrm{i}}^3}{8\pi m_{\mathrm{e}} n_{\mathrm{e}} r_0^2 c^4 \ln \Lambda}, \end{aligned} \tag{4.248}$$

---

[5] Please note that the CR ion velocity in the lab frame is equal to the negative velocity of the electron in the CR ion rest frame, denoted as $v_{\mathrm{i}} = -v'_{\mathrm{e}}$.

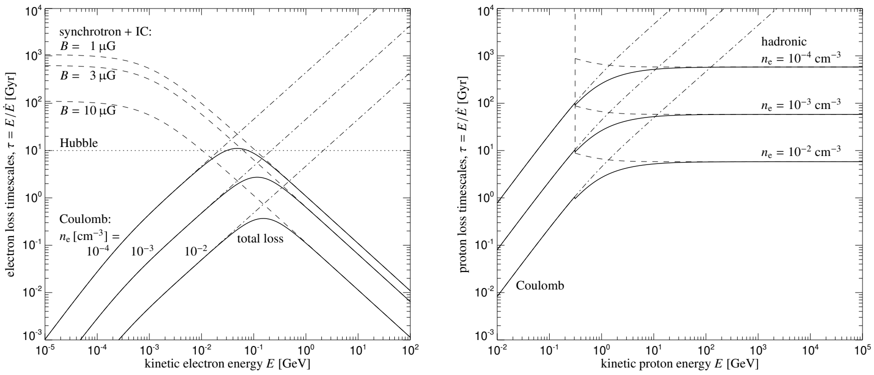


Figure 4.33: Cooling timescales of CRs in the ICM as a function of their kinetic energy. Left: CR electron cooling times due to Coulomb and IC/synchrotron interactions for typical densities and magnetic field strength that range from the central to the peripheral regions in galaxy clusters. Right: CR proton cooling times for the same densities due to Coulomb and hadronic interactions (Enßlin et al. 2011).

Here, n_e represents the electron number density, and σ_{ei} is the Coulomb cross-section. In the second step, we used the non-relativistic limit for simplicity. While both the hadronic and Coulomb cooling timescales are inversely proportional to the gas density, the strong velocity dependence of $t_{\mathrm{Coul,i}} \propto \upsilon_i^3$ implies that for proton energies below approximately 1 GeV, Coulomb interactions are more effective than the hadronic reaction in extracting energy from the CR proton (see the right-hand panel of Fig. 4.33).

- The ***Coulomb interactions of CR electrons*** can be derived analogously to the case of CR ions (see Eq. 4.248). However, in this scenario, we are examining the scattering of a CR electron within the Coulomb field generated by an electron from the background plasma. This interaction results in an enhanced energy transfer due to the fact that both the scattering partners have identical masses, increasing the efficiency of this process compared to electron-ion scattering, despite the nearly identical scattering rates. Consequently, we can derive from Eq. (4.248):

$$t_{\mathrm{Coul,e}} = \left.\frac{E_e}{|\dot{E}_e|}\right|_{\mathrm{Coul,e}} \approx \frac{1}{n_e \upsilon_e \sigma_{ee}} = \frac{1}{n_e \upsilon_e \pi b_0^2\, 2\ln\Lambda} = \frac{\upsilon_e^3}{8\pi n_e r_0^2 c^4 \ln\Lambda}. \tag{4.249}$$

Here, σ_{ee} represents the Coulomb cross-section, and $b_0 = 2r_0c^2/\upsilon_e^2$ is the impact parameter at which the electron's kinetic energy equals its electrostatic potential energy. Interestingly, the Coulomb cooling timescale for CR electrons exhibits a steep decrease at low electron energies, similar to CR ions (as shown in the left-hand panel of Fig. 4.33).

- ***Synchrotron and IC interactions and cooling.*** At high energies, synchrotron interactions with the intra-cluster magnetic field and IC interactions with CMB photons dominate the CR electron cooling.[6] In Section 4.4.2, we derived that the photon energy after an IC collision (averaged over an isotropic distribution of photons) is given by

  $$\langle E_1 \rangle = \frac{4}{3}\gamma_e^2 \langle E \rangle,$$

  where E is the photon energy before the collision and γ_e is the Lorentz factor of the CR electron. The IC energy loss rate of a CR electron is equal to the photon energy gain rate and given by

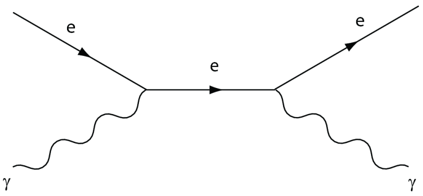
  $$\dot{E}_e = -\sigma_T c n_{ph} \langle E_1 \rangle = -\frac{4}{3}\sigma_T c \varepsilon_{ph} \gamma_e^2 = -\frac{\sigma_T c}{6\pi} B_{ph}^2 \gamma_e^2, \qquad (4.250)$$

  where $\varepsilon_{ph} = \langle E \rangle n_{ph} = B_{ph}^2/(8\pi)$ is the photon energy density and $\sigma_T = 2\pi r_0^2$ is the Thomson cross section.

- Synchrotron and IC interactions resemble each other and are thus described by the same Feynman diagram of the scattering process: while IC emission evokes real photons, synchrotron emission borrows a virtual photon from the magnetic field. In the high-energy regime ($\gamma_e \gtrsim 300$), the energy loss rate of a relativistic electron of energy $E_e = \gamma_e m_e c^2$ due to synchrotron and IC interactions is given by

  $$-\dot{E}_e = \frac{\sigma_T c}{6\pi} \left(B_{cmb}^2 + B^2\right) \gamma_e^2, \qquad (4.251)$$

  where B is the magnetic field strength and $B_{cmb} \simeq 3.2\,(1+z)^2 \mu G$ is the equivalent field of the CMB energy density at redshift z. The first term of Eq. (4.251), $\dot{E}_e \propto B_{cmb}^2$, describes energy loss due to IC scattering off of CMB photons, while the second term $\propto B^2$ describes energy loss due to synchrotron emission. Note that the structural similarity of the two terms is an immediate consequence of the identical scattering process.



Feynman diagram of the electron-photon scattering that simultaneously describes synchrotron and IC interactions.

- The cooling time $t_{cool} = E_e/\dot{E}_e$ of a relativistic electron due to synchrotron and IC interactions is given by

  $$t_{cool} = \frac{E_e}{|\dot{E}_e|} = \frac{6\pi m_e c}{\sigma_T \left(B_{cmb}^2 + B^2\right) \gamma_e} \approx 200\,\mathrm{Myr}, \qquad (4.252)$$

  for $B = 1\,\mu G$ and $\gamma_e = 10^4$ and using Eq. (4.251). A CR electron population that was injected at one epoch and cools for a time $t =$

---

[6] Note that within and close to galaxies, the energy density of starlight photons (in the infrared-to-ultraviolet regime) exceeds that of the CMB. In this case, the CMB energy density would have to be replaced by the total energy density of photons.

t_{cool} shows an exponentially suppressed electron spectrum above the energy/Lorentz factor that corresponds to t_{cool}. In practice, CR electrons are generated over a finite time interval so that the cooled spectrum probes a range of cooling times and associated spectral energy cutoffs. Hence, if we observed such a CR electron population at time $t \gtrsim t_{\mathrm{cool,\,init}}$ (the cooling time of the initially injected CR electron population), we would expect to observe a considerably steepened power-law spectrum in comparison to the injected spectrum.

- The synchrotron frequency in the monochromatic approximation is given by

$$\nu_{\mathrm{syn}} = \frac{3eB}{2\pi\, m_{\mathrm{e}}c}\gamma_{\mathrm{e}}^2 \simeq 1\ \mathrm{GHz}\,\frac{B}{\mu\mathrm{G}}\left(\frac{\gamma_{\mathrm{e}}}{10^4}\right)^2. \qquad (4.253)$$

Combining Eqs. (4.252) and (4.253) by eliminating the Lorentz factor γ_{e} yields the cooling time of electrons that emit at frequency ν_{syn},

$$t_{\mathrm{cool}} = \frac{\sqrt{54\pi m_{\mathrm{e}}c\, eB\nu_{\mathrm{syn}}^{-1}}}{\sigma_{\mathrm{T}}\,(B_{\mathrm{cmb}}^2 + B^2)} \lesssim 190\left(\frac{\nu_{\mathrm{syn}}}{1.4\,\mathrm{GHz}}\right)^{-1/2}\mathrm{Myr}, \qquad (4.254)$$

The cooling time t_{cool} is then bound from above and attains its maximum cooling time at $B = B_{\mathrm{cmb}}/\sqrt{3} \simeq 1.8\,(1+z)^2\mu\mathrm{G}$, independent of the magnetic field.

- Figure 4.33 shows the cooling timescales of CR electrons (left) and CR protons (right) in the ICM as a function of their kinetic energy. It is apparent that CR protons above 10 GeV have live times in the ICM that are at least 60 times longer than CR electrons at any energy. CR electrons can only survive for a Hubble time without re-acceleration within the dilute outskirts of clusters. But in this case, they cool down to Lorentz factors $\gamma_{\mathrm{e}} \sim 100$–300 with kinetic energies $E_{\mathrm{e}} \sim (50$–$150)$ MeV where they cannot be observed on Earth because the plasma cutoff precludes radio waves to propagate through Earth's atmosphere at frequencies below 10 MHz. The GHz-radio emitting electrons have an energy of about 5 GeV in μG fields (Eq. 4.253), and therefore a lifetime of 0.2 Gyr or less. If they are of hadronic origin, their parent CR protons had energies of about 100 GeV (Eq. 4.244), which have considerable longer lifetimes.

4.5.3 Equilibrium Electron Distribution

- We are now discussing the connection between the radio synchrotron spectrum and the radiating CR electron population. In particular, we have to distinguish freshly accelerated electrons and an equilibrium distribution where injection is balanced by radiative losses. In clusters, CR electrons are either directly accelerated at structure formation shocks or injected in hadronic CR proton interactions. This implies a CR electron source function $s_\mathrm{e} = C_\mathrm{inj} E_\mathrm{e}^{-\alpha_\mathrm{inj}}$, with a spectral index $\alpha_\mathrm{inj} \simeq 2$–$2.4$. Note that the test particle limit of diffusive shock acceleration yields a spectral index $\alpha_\mathrm{inj} = 2$ in case of a strong shock (see Eq. 3.323). In steady state, CR electron acceleration balances cooling via synchrotron and IC processes:

$$\frac{\partial}{\partial E_\mathrm{e}} \left[\dot{E}_\mathrm{e}(E_\mathrm{e}) f_\mathrm{e}(E_\mathrm{e}) \right] = s_\mathrm{e}(E_\mathrm{e}), \tag{4.255}$$

where the electron energy loss rate, $\dot{E}_\mathrm{e}$, is given by Eq. (4.251). For $\dot{E}_\mathrm{e}(E_\mathrm{e}) < 0$, this equation is solved by

$$\begin{aligned} f_\mathrm{e}(E_\mathrm{e}) &= \frac{1}{|\dot{E}_\mathrm{e}(E_\mathrm{e})|} \int_{E_\mathrm{e}}^{\infty} \mathrm{d}E'_\mathrm{e} s_\mathrm{e}(E'_\mathrm{e}) \\ &= \frac{C_\mathrm{inj}}{(\alpha_\mathrm{e} - 1)\, |\dot{E}_\mathrm{e}(E_\mathrm{e})|} E_\mathrm{e}^{1-\alpha_\mathrm{inj}} \propto E_\mathrm{e}^{-\alpha_\mathrm{inj}-1} \end{aligned} \tag{4.256}$$

where we assumed synchrotron/IC loss processes in the last step (see Eq. 4.251). Hence, the electron spectral index steepens by unity in steady state, $\alpha_\mathrm{e} = \alpha_\mathrm{inj} + 1$.

- CR electrons with a power-law spectrum, $f_\mathrm{e} = C_\mathrm{e} E_\mathrm{e}^{-\alpha_\mathrm{e}}$ radiate synchrotron emission with a power law in frequency,

$$j_\nu \propto C_\mathrm{e} B^{\alpha_\nu + 1} \nu^{-\alpha_\nu} \tag{4.257}$$

where $\alpha_\nu \equiv \mathrm{d}\log j_\nu / \mathrm{d}\log\nu = (\alpha_\mathrm{e} - 1)/2$. Observationally, the spectral index is determined by comparing radio surface brightness maps at two different frequencies ν_1 and ν_2,

$$\alpha_{\nu_1}^{\nu_2} \equiv \frac{\log(S_{\nu_2}/S_{\nu_1})}{\log(\nu_2/\nu_1)}. \tag{4.258}$$

- Hence, for a steady-state CR electron population that has been accelerated by a strong shock, we expect $\alpha_\mathrm{e} = \alpha_\mathrm{inj} + 1 = 3$ and $\alpha_\nu = (\alpha_\mathrm{e} - 1)/2 = 1$ in the test particle limit. If we resolved the freshly accelerated CR electron population at the shock, we would obtain $\alpha_\nu = (\alpha_\mathrm{inj} - 1)/2 = 0.5$ at the shock in the regime of negligible cooling losses. Observed spectral indices at supernova remnants shocks of $\alpha_\nu \simeq 0.7$ imply $\alpha_\mathrm{inj} \simeq 2.4$, which may require a revision of the theory of diffusive shock acceleration.

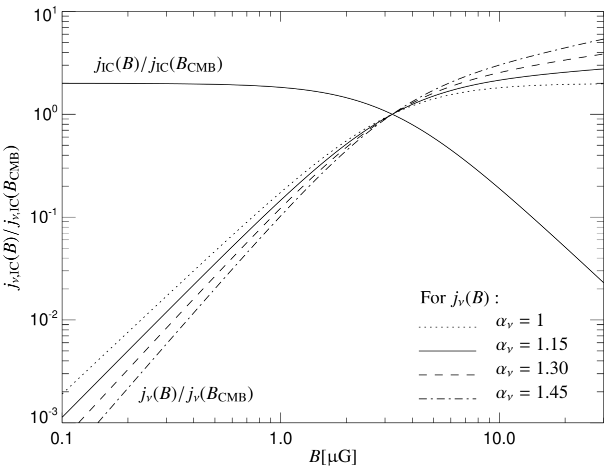


Figure 4.34: The synchrotron and IC emissivity of an equilibrium distribution of CR electrons is shown for various spectral indices α_ν. The normalisation is given by the respective emissivities at the equivalent magnetic field strength of the energy density of the CMB, $B_\mathrm{cmb} = 3.24\,\mu$G. The weak field regime (to the left) is characterised by the dominant IC emission while the strong field regime (to the right) has the synchrotron emission as the dominant electron cooling channel (Pfrommer et al. 2008).

- In steady state, the synchrotron emissivity j_ν and the IC emissivity j_IC are given by

$$j_\nu \propto \frac{C_\mathrm{inj} B^{\alpha_\nu+1}}{B_\mathrm{cmb}^2 + B^2}, \tag{4.259}$$

$$j_\mathrm{IC} \propto \frac{C_\mathrm{inj} B_\mathrm{cmb}}{B_\mathrm{cmb}^2 + B^2}, \tag{4.260}$$

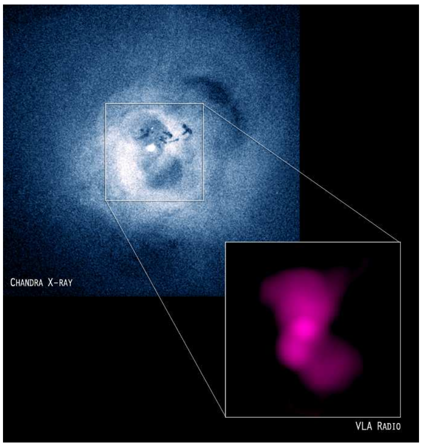
  which is nearly independent of B in the synchrotron cooling regime, $B > B_\mathrm{cmb} \simeq 3.2\,(1+z)^2\,\mu$G, see Fig. 4.34. In the IC cooling regime, $B < B_\mathrm{cmb}$, the IC emission is clearly independent of the magnetic field strength while the synchrotron emission depends on B.

4.5.4 Radio Relics

- Previously, irregularly shaped, diffuse cluster radio sources were called *radio relics*, implicitly assuming that they represent relic bubbles of AGN jets. More sensitive observations with improved angular resolution revealed that radio relics cannot be described by a single population but instead are powered by different physical mechanisms and as such belong to different sub categories. Here, we will provide an taxonomy of diffuse radio relic phenomena in clusters of galaxies and describe the physical origin of each object class (Kempner et al. 2003).

- Radio relics are generally irregularly shaped, have a steep radio spectral index, $\alpha_\nu \sim$ 1–2.5 (adopting the convention $j_\nu \propto \nu^{-\alpha_\nu}$), and are often substantially polarized. While the synchrotron emission is intrinsically polarized, integration of causally unconnected patches along the line of sight depolarizes the synchrotron signal. Observing a highly polarized synchrotron signal implies a narrow emission volume. We distinguish between the following phenomena (and discuss each in detail below):
    - ***Relic AGN bubble.*** This represents an aged radio cocoon with a very steep, cooled spectrum.
    - ***Radio phoenix.*** The passage of a shock across an aged radio cocoon that has already faded out of the radio window revives the CR electron population by adiabatically compressing the bubble magnetic field and electron population so that it becomes observable again, albeit with a low surface brightness.
    - ***Radio gischt or radio shocks.*** This source class shows elongated ($\lesssim$ 2 Mpc sized) radio-emitting structures at the cluster periphery that are in some cases coincident with the weak X-ray shock and are thought to probe freshly injected CR electrons through the process of *diffusive shock acceleration (Fermi I)*.

- ***Relic relativistic plasma in AGN bubbles.*** In our currently favored paradigm, super-massive black holes with $M_\bullet \sim (10^9\text{–}10^{10})\,\mathrm{M}_\odot$ co-evolve with their hosting cD galaxies at the centers of galaxy clusters. The black hole-accretion disk system launches relativistic jets that inflate radio lobes upon interacting with the ambient ICM. The jet/lobes provide $p\mathrm{d}V$ work, push the X-ray emitting gas away, and provide energetic feedback to balance cooling (see Section 3.2.4) as in the AGN in NGC 1275, the cD galaxy in the Perseus cluster (see the X-ray image to the right). The light bubbles are filled with magnetic fields and CR electrons, which emit radio synchrotron radiation, and rise buoyantly on timescales comparable to the sound crossing time $t_\text{sound} \sim 1$ Gyr. This is long in comparison to the synchrotron cooling time of relativistic electrons, see Eq. (4.254). Hence, when relativistic electrons have cooled sufficiently, lobes become invisible in the radio band and form X-ray cavities with so-called "ghost bubbles" as we can observe south and north-west to NGC 1275 (see the composite X-ray radio image of the Perseus cluster).



Chandra X-ray image of the Perseus cluster (blue-white) compared to the VLA radio image (purple) which only shows synchrotron emission from the two central bubbles (NRAO/VLA/G. Taylor).

- ***Radio phoenix.*** Adiabatic expansion of the bubble during its buoyant rise in the cluster potential cause both, the magnetic field strength and the CR electrons to adiabatically cool (in addition to their synchrotron cooling). There is the possibility of

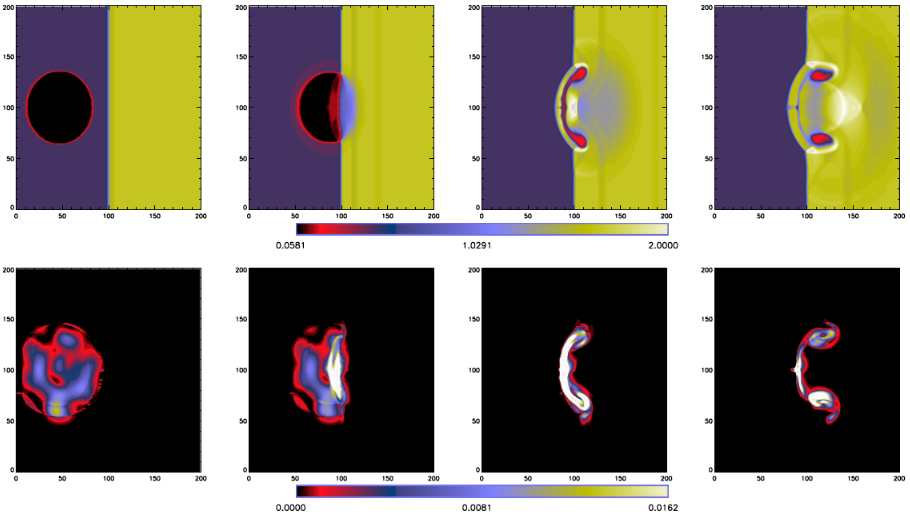
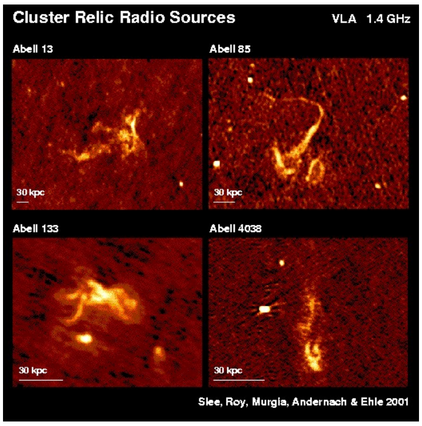


Figure 4.35: Emergence of a radio phoenix by the passage of a shock across an aged radio bubble. Shown is the time evolution (from the initial configuration on the left to the right) of the gas density (top) and magnetic energy density (bottom) in a central slice through the simulation volume (Enßlin & Brüggen 2002). The pre-shock region is on the left-hand side, and the post-shock region on the right-hand side of the stationary shock wave, which is located at the center of the computational box. The shock passage transforms the spherical bubble into a torus, which is best visible in the density plot.

amplifying the magnetic field and energizing the CR electrons. The passage of a shock across an aged radio cocoon transforms the quasi-spherical plasma bubble into a toroidal vortex ring and adiabatically compresses and energizes the aged electron population to emit low surface brightness and steep-spectrum radio emission (Enßlin & Gopal-Krishna 2001, see also radio images to the right). This also adiabatically compresses the magnetic field component perpendicular to the direction of compression so that we expect a turbulent magnetic field to increase according to Eq. (3.299),

$$B \equiv \sqrt{\langle \boldsymbol{B}^2 \rangle} = B_0 \left(\frac{\rho}{\rho_0} \right)^{\alpha_B}, \quad \alpha_B = \frac{2}{3}, \tag{4.261}$$

where B_0 and ρ_0 are the magnetic field strength and mass density just before the onset of adiabatic compression.



VLA radio images of irregularly shaped radio relics that may be representations of the radio phoenix class (Slee et al. 2001).

- The reason for the transformation of a plasma bubble into a toroidal vortex ring can be easiest seen in the rest frame of the shock, where the ram pressure of the pre-shock gas balances the thermal pressure in the post-shock regime. The bubble is filled with hot (relativistic) and more dilute plasma compared to the surrounding ICM. Once the dilute radio plasma of the bubble comes into contact with the shock surface, the ram pressure is reduced at this point of contact due to the smaller density inside the bubble (Enßlin & Brüggen 2002). The shock and the post-shock gas expand into the bubble and propagate with a faster velocity compared to the incident shock in the ICM. Owing to symmetry, the ambient gas penetrates the line through the center of the bubble first and has a smaller velocity for larger impact parameters. This difference in propagation velocities implies a shear flow and eventually causes a vortex flow around the newly formed torus which stabilizes it as it moves now with the post-shock velocity field (see Fig. 4.35).

- Assuming that the bubble was spherical before shock crossing and that the major radius R did not change, we estimate a compression factor[7] through the associated volume change of

$$C = \frac{V_{\mathrm{bubble}}}{V_{\mathrm{torus}}} = \frac{\frac{4}{3}\pi R^3}{2\pi^2 R r_{\mathrm{min}}^2} = \frac{2}{3\pi}\left(\frac{R}{r_{\mathrm{min}}}\right)^2, \qquad (4.262)$$

  where r_{min} is the minor radius of the torus. The radio plasma is adiabatically compressed across the shock passage according to $P_2/P_1 = C^{\gamma_{\mathrm{rel}}}$, where $\gamma_{\mathrm{rel}} = 4/3$ for an ultra-relativistic equation of state. Because of the magnetic field compression, the emerging structure should emit strongly polarized, low surface brightness and steep-spectrum radio emission that ideally shows a toroidal morphology, but which could easily be deformed by turbulent ICM flows to form irregular, filamentary structures (see margin figure on the previous page).

- If the radio bubble is in pressure equilibrium with its surroundings before and after shock crossing, this pressure jump corresponds to the jump at the shock. Applying the standard Rankine Hugoniot jump condition of Eq. (3.119) of an ideal fluid of adiabatic index $\gamma = 5/3$ yields a shock Mach number, and hence to estimate density and temperature jumps across the shock. Application of this method to a vast toroidal radio-emitting structure with diameter 300 kpc at the outskirts of the Perseus cluster enables inferring properties of its accretion shock, including its shock Mach number of $\mathcal{M} \simeq 4.2^{+0.8}_{-1.2}$, density jump $3.4^{+0.2}_{-0.4}$, temperature jump of $6.3^{+2.5}_{-2.7}$, and pressure jump 21.5 ± 10.5 while allowing for uncertainties in the equation of state of the radio plasma and volume of the torus (Pfrommer & Jones 2011).

- ***Radio gischt or radio shocks.*** Cosmic shock waves are ubiquitous and a necessary consequence of cosmic structure formation. The universality of diffusive shock acceleration suggests that merger shocks should amplify magnetic fields and generate

---

[7] Note that we need to distinguish three different compression factors/density contrasts that should not be confused: the compression factor $C = V_{\mathrm{bubble}}/V_{\mathrm{torus}}$ by which the volume of the bubble changes upon shock passage, the compression factor $C_s = \rho_2/\rho_1$ between the post- and pre-shock ICM density at the shock front, and the density contrast $\delta = n_{\mathrm{torus}}/n_{\mathrm{icm}}$ between the radio plasma and the ICM.

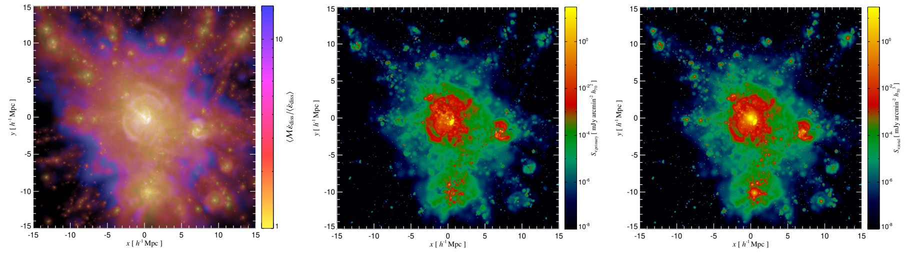


Figure 4.36: Comparison of the Mach number of structure formation shocks (weighted by the dissipated energy, left) to the large-scale "radio web" at 150 MHz of the super-cluster region of a Coma-like cluster (Pfrommer et al. 2008). The middle panel shows the synchrotron emission of *primary CR electrons* that were accelerated directly at structure formation shocks (top left side) as well as additionally the radio emission of *secondary CR electrons* that results from hadronic CR proton interactions with ambient gas protons (right). Note that the smooth centrally peaked "radio halo" emission transitions to the the irregularly shaped primary radio "gischt" emission with a prominent radio relic to the lower right of the cluster (visible as yellow structure in the middle panel).

CR protons and electrons (see Fig. 4.36). The freshly accelerated primary electrons trace merger shocks (Enßlin et al. 1998), which can be observed in form of double radio shocks (see Fig. 4.37). The double radio shocks are strongly polarized with the magnetic field aligned with the shock surface.

- To explore the basics physics of this picture, we assume that the relativistic electrons are accelerated at a merger shock and are advected with the post-shock gas. Assuming that the incoming gas had a pre-shock velocity of $\upsilon_1 = 1200$ km/s in the shock frame, we get a post-shock velocity (Eq. 3.118)

$$\upsilon_2 = \frac{\rho_1}{\rho_2}\,\upsilon_1 = \frac{(\gamma-1)\mathcal{M}_1^2 + 2}{(\gamma+1)\mathcal{M}_1^2}\,\upsilon_1 = 400\left(\frac{\upsilon_1}{1200\,\mathrm{km\ s^{-1}}}\right)\mathrm{km\ s^{-1}} \tag{4.263}$$

for a shock Mach number of $\mathcal{M}_1 = 3$.

- This implies a maximum cooling length $L_{\rm cool,\,max} = \upsilon_2 t_{\rm cool,\,max} = 80$ kpc (Eq. 4.254), which decreases for larger magnetic field strengths to assume a value of

$$L_{\rm cool} = \frac{\upsilon_2\,\sqrt{54\pi m_{\rm e} c\, eB\nu_{\rm syn}^{-1}}}{\sigma_{\rm T}\,(B_{\rm cmb}^2 + B^2)} \approx 30\left(\frac{\nu_{\rm syn}}{1.4\,\mathrm{GHz}}\right)^{-1/2}\mathrm{kpc} \tag{4.264}$$

for 5 μG. Typical radial extends of radio shocks (along the minor axis) are of that size. Hence, one can use the radio shock geometry to estimate magnetic field strengths after carefully accounting for projection effects (van Weeren et al. 2010).

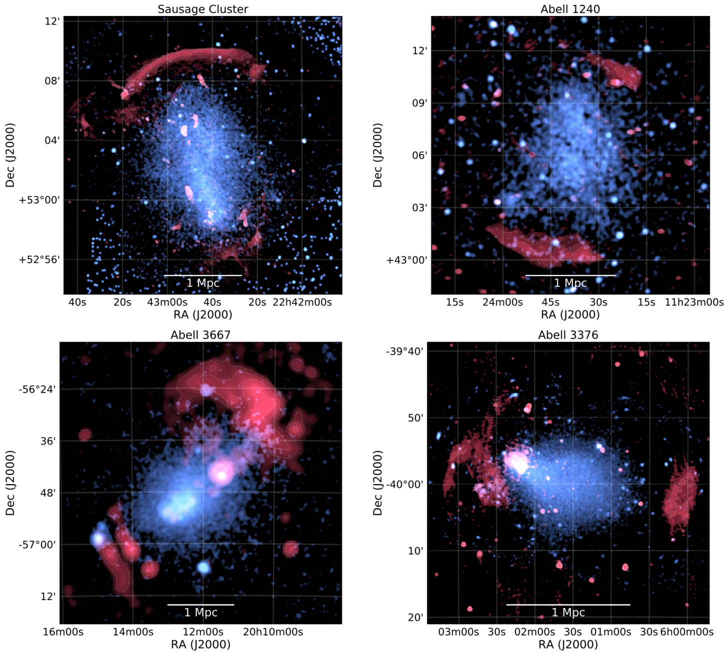


Figure 4.37: Examples of cluster double radio shocks, which trace the merger shock waves in these ongoing cluster collisions. The radio emission is shown in red and the elongated X-ray emission in blue indicates the axis of the cluster merger (van Weeren et al. 2019).

- The radio shock emission at the cluster boundaries is not accompanied by bright thermal X-ray emission and shows astonishing similarities to cosmological cluster simulations (see Fig. 4.38). A spatially resolved spectral analysis of A2256 suggests that the radio shock surface very likely traces the complex shock front, with a broad distribution of Mach numbers propagating through a turbulent and dynamically active ICM (Rajpurohit et al. 2022). The complex filaments are regions where higher Mach numbers dominate the (re-)acceleration of electrons that are responsible for the observed radio emission.

- While the basic picture of radio shocks seems to be in place, there are major open questions remaining, making it an active area of research:

  - What is the seed population of CR electrons that gets accelerated? Direct diffusive shock acceleration from the thermal pool seems to be unlikely because of the small Mach numbers that would required implausible acceleration efficiencies. Instead, the seed electrons could be fossil electrons that were accelerated by strong structure formation shocks in the past during proto-cluster formation and have since cooled and survived at Lorentz factors $\gamma_e \sim 100$–300. Alternatively, the seeds could have been provided by previous AGN outbursts and have survived in giant ghost bubbles.

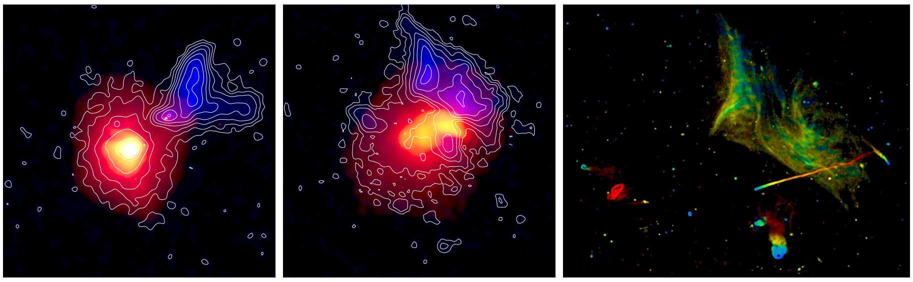
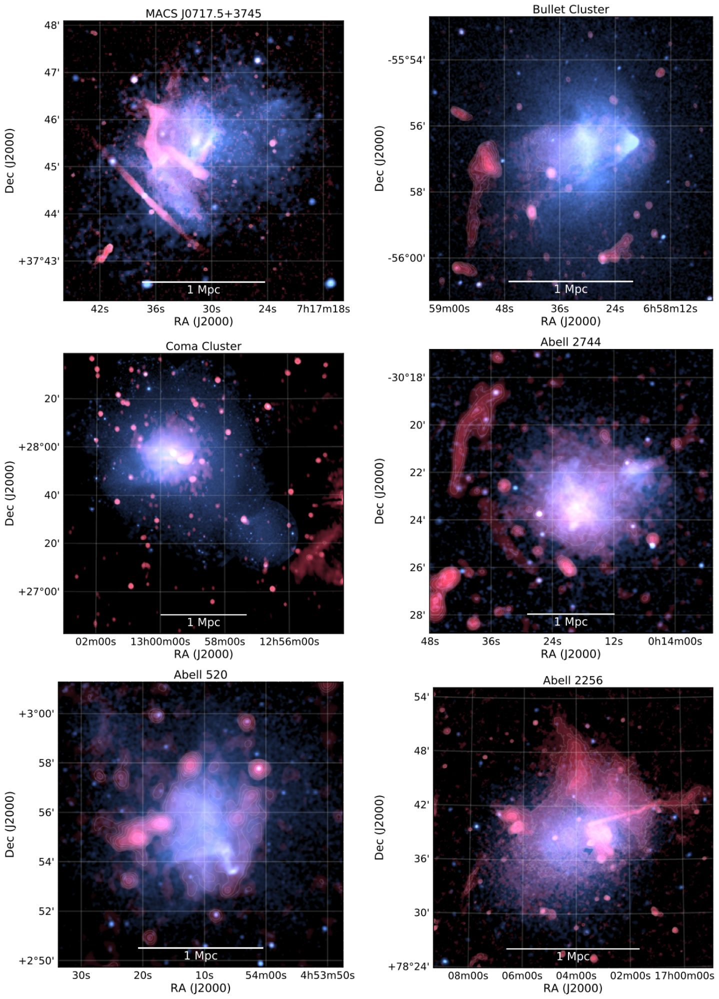


Figure 4.38: Comparison of three-color images of the radio and X-ray emission in a simulation of a merging cluster (left, Pfrommer in prep.) and the radio halo and shock of A2256 (middle, Clarke & Enßlin 2006). The 1.4 GHz synchrotron emission (contours and blue) is contrasted to the X-ray image (red and green). While the radio halo emission resembles the thermal X-ray morphology, the radio shock to the upper right traces the merger shock and only shows weak X-ray emission. Right: higher resolution Jansky VLA image of the A2256 field showing the spectral index α_ν ($S_\nu \propto \nu^{-\alpha_\nu}$) ranging from 2 (red) to 0.4 (blue, Owen et al. 2014). While this reveals the filamentary origin of the radio shock emission, it resolves out the large-scale emission of radio shock and halo.

  - What causes the alignment of the magnetic field with the shock surface as inferred from radio polarization studies? Moreover, quasi-perpendicular shocks are not expected to be efficient electron accelerators provided we can extrapolate the physics of shock acceleration at strong supernova remnant shocks (Winner et al. 2020) to these cluster shocks, which have moderate sonic Mach numbers. More work on the plasma physics of shock acceleration is needed to shed light on this puzzle.
  - Finally, electron cooling length arguments point to strong magnetic fields at the shock position, close to the cluster virial radii (Eq. 4.264). This implies low thermal-to-magnetic pressure ratios $\beta \sim 5$, which is at least a factor of ten lower than the bulk of the ICM. Which physics is responsible for these strong magnetic field strengths at shocks?

4.5.5 Radio Halos

- There is another source class, called *radio halos*, which shows spatially extended regions of diffuse radio emission. The radio halo emission has a regular, cluster-centric morphology and is unpolarized. Because synchrotron emission is intrinsically polarized, this implies the contribution of various magnetic field orientations along the line of sight, and Faraday rotation depolarization. Hence, this suggests that the emission volume of radio halos must have a substantial volume filling fraction. Here,



MACS J0717.5+3745

Bullet Cluster

Coma Cluster

Abell 2744

Abell 520

Abell 2256

Figure 4.39: Examples of radio halos, which are hosted by merging clusters with a disturbed gas morphology. The radio emission is shown in red, the X-ray emission in blue, and the purple color indicates a good spatial correlation of the two emission processes (van Weeren et al. 2019).

we will provide an taxonomy of diffuse radio halo phenomena in clusters of galaxies and describe their physical origins. Similar to radio relics, the spectral index of radio halos is also steep, $\alpha_\nu \sim 1$–1.5. We distinguish between the following radio halo phenomena (and discuss each in detail below):

  - ***Giant radio halos*** occur in merging clusters and extend over the entire cluster, typically from 1–2 Mpc (see Fig. 4.39). The morphology of radio halos and the thermal X-ray emission resemble each other.
  - ***Radio mini halos*** occur in relaxed, cool core clusters (see Fig. 4.40). They are smaller than giant halos and only extend over the cool core region, which is typically 100–500 kpc in size. They are often confined to sloshing cold fronts.

- The radio luminosity of giant halos is strongly correlated with the

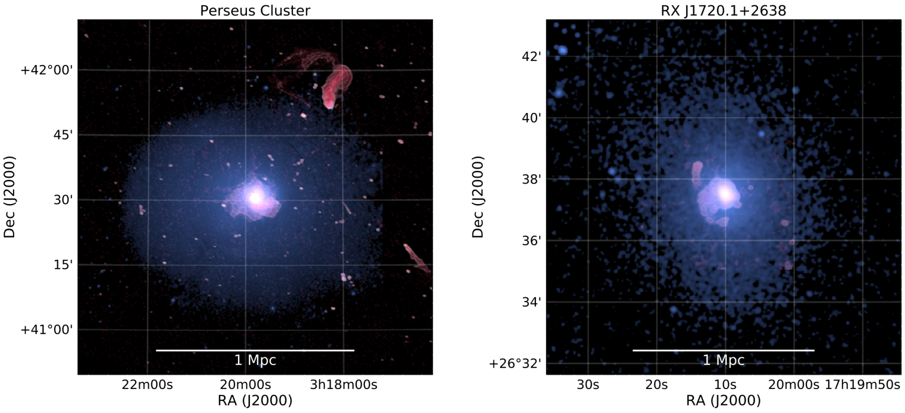


Figure 4.40: Examples of clusters hosting radio mini-halos (the radio emission is shown in red, the X-ray emission in blue, van Weeren et al. 2019). The objects in the outskirts of Perseus are head-tail radio galaxies (see Section 4.5.6).

X-ray luminosity of the cluster (see left-hand panel of Fig. 4.41, Brunetti et al. 2009). While the radio luminosities of mini halos also follow this correlation with X-ray luminosity, which is usually dominated by their cool core region, there seems to be a larger scatter. However, a large fraction of clusters does not exhibit significant radio halo emission as indicated by the upper limits on their radio flux. About half of the radio silent clusters, for which there is *Chandra* data available, show clear evidence for hosting cool cores (with a central entropy $K_0 \lesssim 50\,\mathrm{keV\,cm}^2$) as can be seen in the right-hand panel of Fig. 4.41.

- This finding could either imply that these clusters are in a transitional stage after their giant radio halo was extinguished and the cool core has not yet fully developed so that the associated radio mini halo has not yet been powered on. Because this transitional theory is not supported by X-ray evolution studies of clusters (Section 4.3.2), there is an alternative explanation that proposes two populations of clusters – cool cores and non-cool cores – and the corresponding radio luminosity would sensitively respond to the level of injected turbulence by either the AGN or the cluster merger, respectively. Therefore, massive clusters at the same X-ray luminosity are bimodally distributed with respect to their radio luminosity. Either they exhibit a radio halo or they do not exhibit any detectable diffuse radio halo emission. This indicates the existence of pronounced and rapidly operating switch-on/switch-off mechanisms, which are able to change the radio luminosity by at least a factor of $10-30$ (Brunetti et al. 2009).

- ***Giant radio halos.*** To understand the basic problem of explaining the cluster-wide radio halo emission, we first estimate the synchrotron cooling length of the radiating CR electrons. Assuming that the turbulent-to-thermal pressure ratio is $X_\mathrm{turb} = P_\mathrm{turb}/P_\mathrm{th} \simeq$

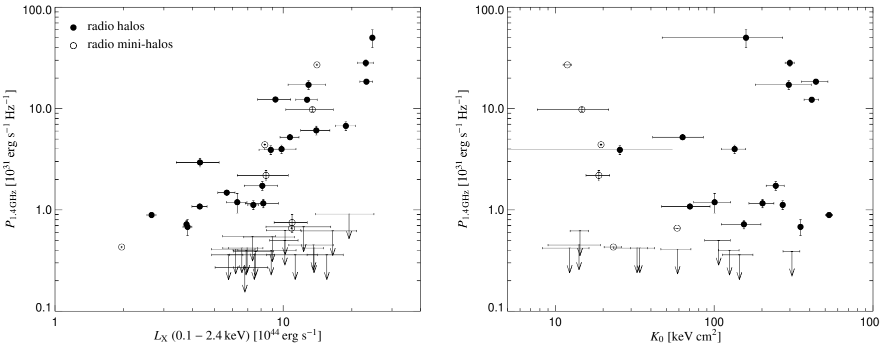


Figure 4.41: Correlation of radio halo luminosities with the properties of clusters hosting radio (mini) halos. Left: Radio halo luminosity vs. X-ray luminosity. Right: Radio halo luminosity vs. the central entropy K_0 (as defined by Eq. 4.143) for the subsample of clusters for which high resolution Chandra data are available (Enßlin et al. 2011).

$(v_{\mathrm{turb}}/c_{\mathrm{s}})^2 = 0.1$, we obtain a typical turbulent ICM velocity of

$$v_{\mathrm{turb}} \simeq \sqrt{X_{\mathrm{turb}}} c_{\mathrm{s}} \simeq 400\,\mathrm{km\ s}^{-1} \left(\frac{X_{\mathrm{turb}}}{0.1}\right)^{0.5} \left(\frac{c_{\mathrm{s}}}{1200\,\mathrm{km\ s}^{-1}}\right). \tag{4.265}$$

The maximum cooling length is $L_{\mathrm{cool,\,max}} = v_{\mathrm{turb}} t_{\mathrm{cool,\,max}} \simeq 80$ kpc at 1.4 GHz (Eq. 4.254). However, the spatial extend of giant radio halos is much larger, $L_{\mathrm{halo}} \sim$ (2–4) Mpc, and the emission is not polarized. This means that the emission is a superposition of causally unconnected patches along the line of sight in combination with Faraday rotation depolarization. Because $L_{\mathrm{halo}} \approx 25 L_{\mathrm{cool,\,max}}$ there must be a volume filling acceleration process of relativistic electrons that effectively counteracts the fast synchrotron electron cooling throughout the cluster volume in merging clusters.

- There have been two general models for radio halos discussed in the literature. (i) In the *hadronic model*, relativistic protons hadronically interact with gas protons and produce secondary electrons/positrons (CRp + p $\rightarrow \pi^{\pm} \rightarrow e^{\pm}$) that power the radio halo. The large CR proton cooling times (Fig. 4.33) imply that CR protons accumulate over the Hubble time in the cluster volume and maintain a volume-filling distribution. (ii) In the *re-acceleration model*, fossil or secondary electrons interact with turbulent magneto-hydrodynamic waves and experience Fermi-II acceleration that energizes them (Section 3.3.3.3) so that they become visible at radio wave lengths.
- All required ingredients of the *hadronic model* are available: shocks to inject CR protons, gas protons as targets, and magnetic fields. The model predicts radio halo luminosities and morphologies as observed without tuning, and returns realistic power-law

spectra. However, the simple version of the hadronic model predicts radio halos in all clusters (which is clearly in contradiction to the population of radio silent clusters, see Fig. 4.41). In principle, accounting for CR streaming and diffusion out of the dense central regions would significantly decrease the radio halo luminosity (Enßlin et al. 2011). However, it is unclear whether CRs can stream fast enough to explain the required fast transition between radio active and silent clusters and the gap in radio luminosity between these two populations. Also, there is a population of radio halos with ultra steep spectra (Brunetti et al. 2008) that cannot be modeled by the hadronic model. Instead, the large CR proton cooling times preclude spectral aging and predict universal radio spectra of the halo population.

- Because of these challenges, we turn our attention to the *re-acceleration model*. Its ingredients are also available: radio galaxies, AGN relics or CR protons to inject CR electrons, cluster mergers to inject ICM turbulence, and plasma waves to re-accelerate mildly relativistic electrons. In this model, the observed convex radio spectra emerge naturally because of the competition of fast synchrotron cooling and inefficient turbulent Fermi-II re-acceleration ($\propto \upsilon_{\rm w}^2/c^2 \sim 10^{-6}$), which cannot any more balance synchrotron cooling at high electron energies (Brunetti & Lazarian 2011). This model naturally predicts a radio halo bimodality as cluster mergers inject turbulence that re-accelerates CR electrons. Turbulence decays away after the merger and the radio halo quickly extinguishes because of fast synchrotron cooling so that radio-silent clusters are simply less turbulent. However, re-accelerating fossil CR electrons from past structure formation shocks with standard assumptions yields too centrally concentrated radio emission. Matching observations requires modifying properties of the CR population (rapid streaming, enhanced CR electron acceleration at shocks) or turbulence (increasing turbulent-to-thermal energy density with radius, Pinzke et al. 2017). Moreover, some clusters such as the bullet cluster show power-law radio spectra which require fine tuning in this model – more work on the plasma physics is needed.

- ***Radio mini halos*** appear to be miniature examples of giant halos. As for the giant halos, the hadronic and the turbulent re-acceleration models are discussed for powering the radio-emitting CR electrons. However, in cool core clusters, there is much less turbulence available. Possible sources include rising AGN bubbles (where the injected turbulence is however confined to the wake) and sloshing cores that can drive a low level of turbulence into the core. In case of the hadronic model, a CR proton-to-thermal pressure ratio of $P_{\rm cr}/P_{\rm th} \simeq 0.02$ is sufficient to explain the observed radio mini halo luminosity, which can easily be delivered by the AGN jet.

4.5.6 Radio Galaxies

- Radio galaxies found in cluster environments often show distorted morphologies due to their interactions with the ICM. We distinguish cluster-central radio galaxies that are responsible for AGN jet feedback (Section 3.2.4) and so-called head-tail galaxies. These galaxies appear with a very luminous head (the AGN in the center of the hosting galaxy) and extended jets that are bent by the ram pressure experienced by the galaxy due to its motion through the ICM (see Fig. 4.42). Typically, the jets are fanning out and expand in the form of lobes in the wake of the galaxy.

- These head-tail galaxy types can be divided into two subclasses: (i) narrow-angle tail jets experiencing high ram pressure and (ii) wide-angle tail jets that experience a weaker ram pressure, which could be caused by low velocities relative to the cluster centre as a result of cluster-cluster mergers. Examples can be found in the outskirts of the Perseus cluster (Fig. 4.40) and in A2256 (Fig. 4.38, where the spectral index gradient along the tail indicates spectral aging, which may partially be counteracted by turbulent re-acceleration.)

- We can derive the jet curvature radius that is caused by the external ram pressure wind due to the motion of the galaxy through the ICM. We denote the mass density, velocity, and radius of the jet by ρ_jet, υ_jet, and r_jet, respectively. The two jets coming out of the active galactic core back to back are assumed to be initially a cylinder of length l_jet. Each jet is then bent over a bending radius r_b by the ram pressure wind of mass density and velocity ρ_ICM and υ. We equate the jet momentum $\rho_\mathrm{jet}\upsilon_\mathrm{jet}\pi r_\mathrm{jet}^2 l_\mathrm{jet}$ with the transverse force due to the ram pressure wind that acts over a jet propagation timescale (along the bended path in steady state), $\rho_\mathrm{ICM}\upsilon^2 2 r_\mathrm{jet} l_\mathrm{jet}\pi r_\mathrm{b}/(2\upsilon_\mathrm{jet})$, to obtain the following equality

  $$\rho_\mathrm{ICM}\upsilon^2\frac{\pi r_\mathrm{b}}{2\upsilon_\mathrm{jet}} = \rho_\mathrm{jet}\upsilon_\mathrm{jet} r_\mathrm{jet}\frac{\pi}{2}. \tag{4.266}$$

  Solving for the ratio of bending-to-jet radius, we obtain

  $$\frac{r_\mathrm{b}}{r_\mathrm{jet}} = \frac{\mathcal{M}_\mathrm{jet}^2}{\mathcal{M}_\mathrm{gal}^2}\frac{\gamma_\mathrm{jet} P_\mathrm{jet}}{\gamma_\mathrm{ICM} P_\mathrm{ICM}}, \tag{4.267}$$

  where we introduced the Mach numbers of the jet and the galaxy, $\mathcal{M}_\mathrm{jet} = \upsilon_\mathrm{jet}/c_\mathrm{jet}$ and $\mathcal{M}_\mathrm{gal} = \upsilon/c_\mathrm{ICM}$, the adiabatic exponents of jet and surrounding ICM, $\gamma_\mathrm{jet} = 4/3$ and $\gamma_\mathrm{ICM} = 5/3$, and

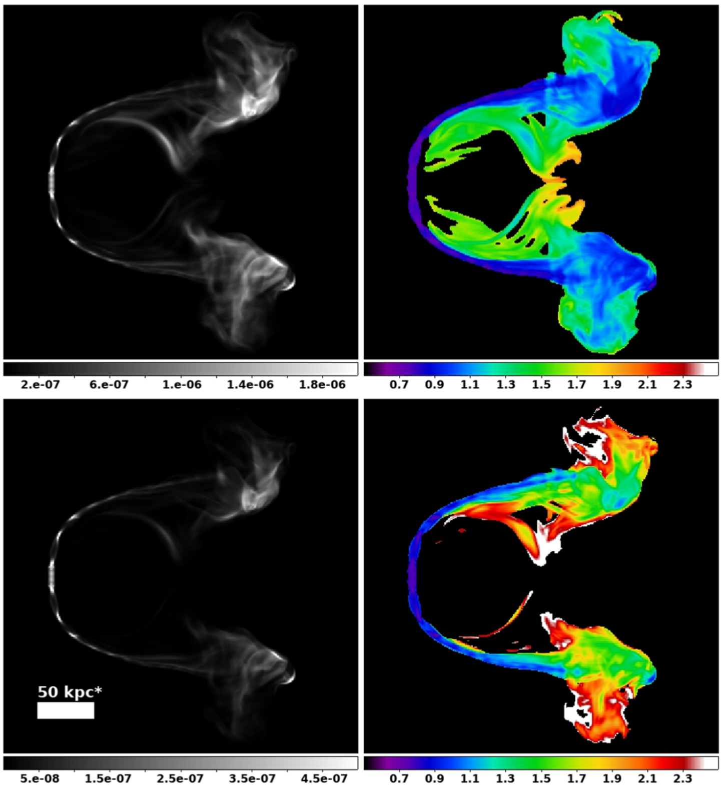


Figure 4.42: Radio images of head-tail simulations (O'Neill et al. 2019) where the jet is launched vertically and the galaxy moves to the left so that the ram pressure wind bends the jets over and turbulence causes them to fan out. Left: total synchrotron intensity at 150 MHz (top) and 950 MHz (bottom). Right: radio spectral index maps, $\alpha_{150\,\mathrm{MHz}}^{325\,\mathrm{MHz}}$ (top) and $\alpha_{950\,\mathrm{MHz}}^{1400\,\mathrm{MHz}}$ (bottom). Multiple jet hot spots in the head on either side along the jet correspond to re-collimation shocks. The bottom images are fainter and show steeper spectra, especially along the tails because they probe higher radio frequencies and as such an older CR electron population.

their pressures, P_{jet} and P_{ICM}. High-resolution radio observations yield constraints on the left-hand side of Eq. (4.267) as well as on P_{jet} while X-ray and Sunyaev-Zel'dovich observations constrain P_{ICM}. Hence, Eq. (4.267) yields a constraint on the ratio $\mathcal{M}_{\mathrm{jet}}/\mathcal{M}_{\mathrm{gal}}$. Provided an independent constraint on $\mathcal{M}_{\mathrm{gal}}$, this enables to solve for $\mathcal{M}_{\mathrm{jet}}$.

- If we take the jet power to be $L_{\mathrm{jet}}/2 \sim \rho_{\mathrm{jet}} v_{\mathrm{jet}}^3 \pi r_{\mathrm{jet}}^2$ (where L_{jet} is the luminosity of the 2 jets and we assume that most of the jet power is kinetic energy), the power can be expressed in terms of the jet pressure and Mach number as

$$\frac{L_{\mathrm{jet}}}{2} \simeq \gamma_{\mathrm{jet}} P_{\mathrm{jet}} \mathcal{M}_{\mathrm{jet}}^2 v_{\mathrm{jet}} \pi r_{\mathrm{jet}}^2. \tag{4.268}$$

Minimum energy density arguments and geometric modeling constraints enable us to constrain the jet luminosity, L_{jet} and jet pressure, $P_{\rm jet}$. Using our inferences on $\mathcal{M}_{\rm jet}$ from Eq. (4.267) hence enables us to solve for the jet velocity, $v_{\rm jet}$.

- Multi-frequency observations of head-tail galaxies yield spectral index maps that help to probe synchrotron cooling and turbulent re-acceleration. Generally, the lengths of a number of observed head tails exceed the synchrotron cooling lengths by a factor of several which empirically points to the requirement of a gentle re-acceleration process (de Gasperin et al. 2017). Moreover, the radio morphology along the tails enables to probe the transition from laminar to turbulent motion and allows for constraints on the effective particle mean free path, which is orders of magnitude below the classical estimate and hence, reinforces the need to consider plasma effects (Müller et al. 2021). Finally, head-tail radio galaxies are local probes of the ICM velocity field and as such are excellent probes of the turbulent cluster weather which enables us to constrain the dynamical state of clusters.

4.6 Cluster Cosmology

4.6.1 Cosmological Parameter Estimates

- Statistical studies of the CMB temperature and polarization fluctuations in combination with other high-precision cosmological probes (such as gravitational lensing or galaxy baryon acoustic oscillation) give rise to the ΛCDM concordance model, a cosmology with cold dark matter and a cosmological constant Λ. This model depends on six very accurately measured parameters: the cold dark matter density parameter $\Omega_c h^2 = 0.120$ (where $h = 0.674$ is the current day Hubble constant in units of 100 km s^{-1}Mpc^{-1}), the baryonic density parameter $\Omega_b h^2 = 0.0224$, the scalar spectral index $n_s = 0.965$ and the primordial amplitude $\ln(10^{10} A_s) = 3.045$ of the power spectrum, the Thompson optical depth to the CMB surface of last scattering $\tau = 0.054$, and the angular scale of the comoving sound horizon at recombination, $\theta_{MC} = 0.0104$, quantifying the distance the photon-baryon perturbations can influence (Planck Collaboration 2018). These results are only weakly dependent on the cosmological model and remain stable, with somewhat increased errors, in many commonly considered extensions.

- While these parameters are derived from physics that probe either the early universe at the time of last scattering at redshift $z = 1090$ or (large) linear scales around half the age of the universe, a complete picture of the universe requires to link this concordance model to the current-day universe in the non-linear regime using cosmological probes such as galaxy clusters. However, the level of precision already reached by the CMB and galaxy baryon acoustic oscillation surveys has substantially "raised the bar" for cluster cosmology.

- Galaxy clusters are particularly sensitive to (i) the amplitude of the (linear) power spectrum on the scale of 8 h^{-1} Mpc because this is the scale that collapses to clusters and (ii) deviations from Gaussian primordial density fluctuations δ (such as a skewness or kurtosis) because matter in the high-density tail of the probability distribution of δ collapses to halos. Because the abundance of clusters is exponentially suppressed, it is exponentially sensitive to changes in the shape of the probability distribution of δ. (iii) Finally, if neutrinos have a non-negligible mass, they account for a small fraction of the mass content attributed to dark matter. Because their masses are a fraction of an eV, they are relativistic when they decouple from the thermal heat bath in the early universe (i.e., they act as hot dark matter) and damp small-scale neutrino density fluctuations: neutrinos cannot be confined into (or kept outside of) regions smaller than the free-streaming length, because their velocity is greater than the escape velocity from gravitational potential wells on those scales. Hence, the power spectrum of such a Λ mixed dark matter model, where mixed refers to the inclusion of some hot dark matter component, does not follow a power law behaviour towards smaller scales but shows an additional suppression by a factor of $\Omega_\nu/\Omega_\mathrm{c}$. Moreover, the growth rate of cold dark matter perturbations is reduced through an absence of gravitational back-reaction effects from free-streaming neutrinos. Hence, non-negligible neutrino masses reduce the abundance of massive clusters for given CMB signal. In particular, assuming a sum of neutrino masses of $\sum m_\nu = 0.3$ eV is equivalent to a $\sim 15\%$ shift in cluster mass.

4.6.2 How Clusters Probe Cosmology

In the following, we will present a number of different methods for estimating cosmological parameters with clusters and the main challenges of the methods, respectively:

- ***Cluster abundance to measure the fluctuation amplitude*** σ_8**.** The larger the amplitude of density fluctuations, the more (filtered) density peaks exceed the critical density contrast for spherical collapse, $\tilde{\delta} > \delta_\mathrm{c}$ (Section 2.4.1), and hence, more halos form per unit volume. However, cluster masses cannot be directly measured but instead must be inferred from cluster observable-mass scaling relations (see Section 3.1.8), which approximately follow the self-similar expectation for X-ray temperature, galaxy velocity dispersion, SZ and X-ray Comptonization (T, σ_gal, Y, and Y_X) but show substantial deviations for the X-ray luminosity, L_X, and optical galaxy number counts, N_gal.

- ***Cluster abundance evolution to measure*** $\Omega_\mathrm{m} = \Omega_\mathrm{c} + \Omega_\mathrm{b}$**.** For ΛCDM models, cluster abundances at a given redshift are given by a combination of the parameters σ_8 and Ω_m. However, cluster evolution sets σ_8 and Ω_m separately so that a measurement of cluster abundances as a function of redshift is able to break the degeneracy and provide individual constraints on σ_8 and Ω_m (see Fig. 4.43). Possible redshift evolutions of the $Y - M$, $L_\mathrm{X} - M$, or $T - M$ relations would bias the inferred cosmological parameters.

- ***Cluster baryon fraction to estimate*** $\Omega_\mathrm{b}/\Omega_\mathrm{m}$**.** Integrating the gas number density and adding the stellar content of cluster galaxies yields the baryon mass inside a cluster. This is compared to the total cluster mass that is either obtained from integrating the hydrostatic mass profile or from cluster scaling relations. Dividing both measures yields the baryon fraction $M_\mathrm{b}/M_\mathrm{m} = \Omega_\mathrm{b}/\Omega_\mathrm{m}$

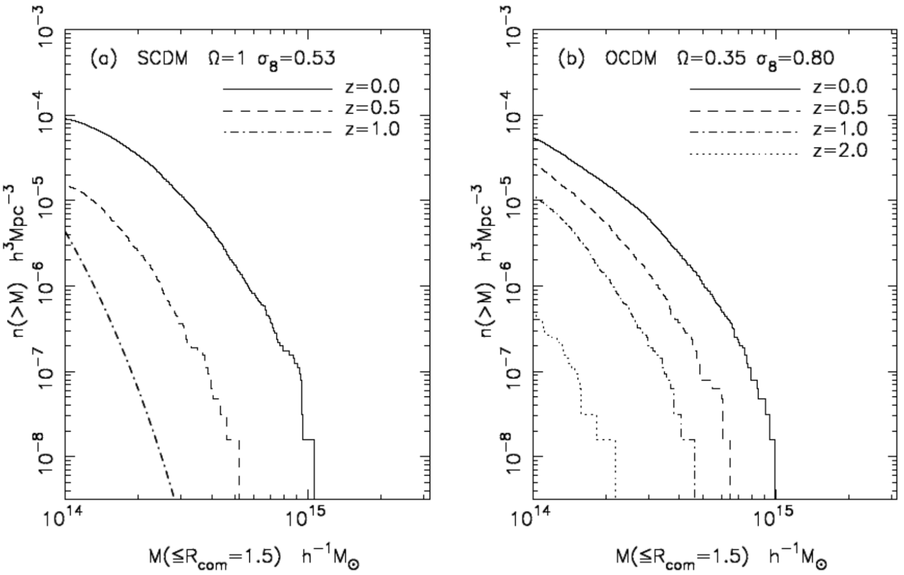


Figure 4.43: Shown is how the evolution of the number density of rich clusters of galaxies breaks the degeneracy between the mass density ratio of the universe, Ω_m, and the normalization of the power spectrum, σ_8, that follows from the observed present-day abundance of rich clusters. The evolution of high-mass (Coma-like) clusters is strong in $\Omega_\mathrm{m} = 1$, low-σ_8 models, where the number density of clusters decreases by a factor of $\sim 10^3$ from $z = 0$ to $z \simeq 0.5$; the same clusters show only mild evolution in low-Ω_m, high-σ_8 models, where the decrease is a factor of ~ 10 (Bahcall et al. 1997).

inside the cluster's virial radius. However, gas clumping biases the X-ray inferred density estimates and the gas at R_{200} can only be observed in very deep X-ray observations so that typically the gas properties at the virial radius need to be extrapolated. A combined X-ray–SZ analysis partially circumvents these problems, but also needs to assume spherical symmetry. This assumption is progressively getting worse towards the virial radius, where the kinetic pressure support increases, the dark matter distribution is becoming more tri-axial and the density- and pressure distribution becomes very clumpy, questioning the accuracy of the estimates.

- ***Cluster distribution to estimate power spectrum and baryon acoustic oscillations in the large-scale structure.*** The spatial distribution of clusters is a highly biased tracer of the total mass distribution and as such excellently suited to probe the power spectrum and/or the spatial distribution of mass. This can be done by (X-ray or SZ) clusters in all-sky surveys or via brightest cluster galaxies (BCGs). The main challenges of this method is the sparse sampling of clusters because of their rarity, which adds a substantial Poisson scatter. Moreover, AGN feedback modifies the power spectrum at surprisingly large scales so that accurate measurements of cosmological parameters are degenerate with uncertainties of baryonic physics.

- ***Cluster thermal SZ power spectrum to measure fluctuation amplitude σ_8.*** Because of the exponentially suppressed abundance of clusters, the SZ power spectrum is a very sensitive probe of σ_8, where the SZ power spectrum amplitude scales as σ_8^7. The challenge of this method consists in controlling the uncertainty of baryonic physics, in particular AGN feedback that redistributes gas towards the outer cluster regions and thus modifies the thermal pressure profile, which adds power on smaller angular scales in comparison to models with less AGN feedback (see Section 4.4.4).

- ***Cluster core structure as a test of the nature of dark matter.*** The dark matter in the cuspy central region is characterized by a cold kinematic phase space distribution. If dark matter has self-interactions, these cause dark matter particles to be heated and their velocity dispersion to be increased, which transforms the steep inner cusp to a flatter central slope. Hence, measurements of the inner dark matter density profile may help elucidating the nature of dark matter. This can be probed via multiple galaxy images as a result of strong gravitational lensing or by kinematic modeling the stars in the central dominant (cD) galaxy through their velocity distribution. Of concern for this method is how the cD galaxy assembly influences the dark matter profile: while strong radiative cooling of the central gas and abundant star formation would cause a strong radial inflow which would steepen the dark matter profile via adiabatic contraction, impulsive outflows of a large amount of gas as a result of strong feedback could non-adiabatically heat the dark matter and cause a cored distribution.

4.6.3 Cluster Abundances

We have seen that in order to obtain accurate cosmological parameters, “precision” cluster abundances are absolutely necessary. In the following we scrutinize this requirement for the methods using cluster abundances.

- Cluster abundances as a function of mass can only be calculated imprecisely from theory and measured precisely from (sufficiently large) simulations. The reason for this is hierarchical structure formation implying a recent formation epoch of clusters so that many of the more massive systems have not had time to virialize and are still in the process of forming. As a result, in the cluster outskirts (which substantially contribute to cluster masses) there is a substantial kinetic pressure support that increases with radius and cluster are becoming more tri-axial as they are connected to individual filaments of the cosmic web.

- However, cluster masses in simulations depend on the mass definition because clusters do not have clear boundaries. In the literature, there are different definitions: (i) the *friends-of-friends mass* links mass elements to a cluster if their mutual distance is closer than a critical linking length. This can yield irregularly shaped clusters that are bounded by an isodensity surface. (ii) The *spherical overdensity mass* algorithm calculates the mean density contained in a sphere. Starting with a small radius that corresponds to a large overdensity, the radius is consecutively increased until the mean density falls below a desired overdensity threshold. Popular choices for the virial mass are M_{200} or M_{500} (corresponding to mean densities that are 200 or 500 times the critical density at the redshift in question) or $M_{200\mathrm{m}}$ (mean densities that are 200 times the cosmic mean density). While the spherical overdensity mass critically depends on the choice of the cluster center (typically the most-bound mass element or highest density point), friends-of-friends masses can produce peanut-shaped clusters in case of a merger so that spherical profiles a larger radii may be ill-defined as their concentric shells contain regions outside the cluster boundaries.

- Simulated cluster abundances depend on baryonic physics. While the baryonic mass content only amounts to $\sim 15\%$, powerful AGN feedback can move a substantial fraction of the gas beyond the (somewhat artificially) chosen virial radius. Thus pulls dark matter gravitationally to larger radii (or at least slow down its assembly) so that masses are typically biased low by $\sim 5\%$.

- Cluster abundances are measured for samples selected observationally by richness/optical flux, X-ray flux, SZ flux, lensing signal. Because of various observational biases, the resulting cluster samples are very different in the redshift and mass distribution. At a given mass, an X-ray cluster sample is biased towards lower redshift because $L_\mathrm{X} \propto D_\mathrm{lum}^{-2}$ (D_lum is the luminosity distance) and is more complete in cool core clusters because of the denser cores that boost the X-ray luminosity in comparison to non-cool core clusters. SZ clusters show a more homogeneous selection in redshift (the limiting mass is only very weakly dependent on redshift) which yields almost volume-limited samples. However, they also contain many more merging, non-virialized clusters because merger shocks boost the thermal pressure and hence the SZ signal. Finally, optically selected samples show a larger Poisson scatter at small (group-scale) objects because of sparse sampling of group/cluster masses with only a few galaxies. Hence, these observational biases need to be accounted for when comparing the resulting samples or when using cluster mass calibrations determined from one method and applying it to a cluster sample inferred from a different method.

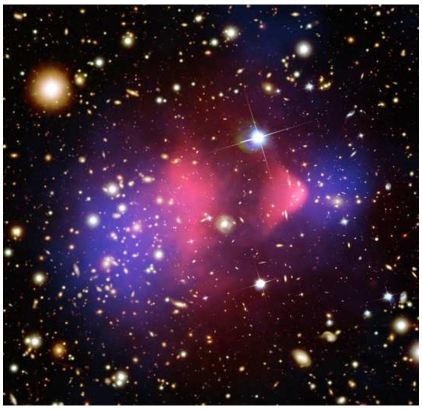
- Abundances are measured as a function of observationally inferred mass, from $N_{\rm gal}$, $\sigma_{\rm gal}$, $L_{\rm X}$, $T_{\rm X}$, $Y_{\rm X}$, Y, and gravitational lensing mass, $M_{\rm lens}$. However, the resulting cluster masses depend sensitively on the mass–observable scaling relations both through sample selection and through mass estimation. A prominent example for such a potential bias is the cluster count as a function of SZ flux Y (or X-ray proxy $Y_{\rm X}$) and z, which implied a lower σ_8 than Planck infers from primary CMB fluctuations (Planck Collaboration 2015). This discrepancy depends critically on the M–Y or M–$Y_{\rm X}$ calibration, which differ for the different sample. Alternatively, the discrepancy could signal the need for an extension of the minimal six-parameter cosmological concordance model and may be caused by neutrino masses (see Section 4.6.1).

- Critically, the normalizations, slopes, (correlated) scatter of the scaling relation, and the (possibly non self-similar) redshift evolution of these parameters is subject to *Malmquist bias*. This bias is referred to as a selection bias or data censoring and affects the results in a brightness-limited survey, where clusters below a certain apparent (X-ray or optical) brightness cannot be detected. The brightness decreases with luminosity distance squared until it falls below the observational threshold. Objects which are more luminous, or intrinsically brighter, can be observed at a greater distance, creating a false trend of increasing intrinsic brightness, and other related quantities, with distance.

4.6.4 Clusters Probe the Nature of Dark Matter

- Because the two parameters concentration c_{200} and halo mass M_{200} of a dark matter density profile are correlated via the c_{200}–M_{200} relation, this is effectively a one-parameter fit (for the mean cluster mass). The largest galaxies populate the largest dark matter halos, the second-largest galaxies reside in the second largest halos, and so on for the next largest galaxies and halos. This so-called *abundance matching* procedure uniquely assigns galaxies to dark matter halos. A more realistic modeling also accounts for a moderate scatter between halo and galaxy size. Hence, an observed distribution of galaxies can be uniquely matched to the distribution of dark matter halos implying a zero-parameter test. Following this procedure, it is shown that the mean density profile of rich clusters on intermediate to large scales has the predicted NFW shape expected for a ΛCDM cosmology (Okabe et al. 2013, see also Fig. 4.44).

- Strong lensing and galaxy kinematics constrain the total mass profile at small cluster-centric radii. Subtracting an estimate of the stellar mass profile constrains the dark matter profile. The central dark matter profile is consistent with the NFW expectation and shows even hints of being steeper on the mass scale of Milky Way galaxies and galaxy groups signaling the effect of adiabatic compression of dark matter during the assembly of the stars. By contrast, dark matter is less concentrated in BCGs of massive galaxy clusters than expected (see Fig. 4.45). If confirmed this could signal velocity dependent dark matter self-interactions that increase the central dark matter velocity dispersion, thus generating a cored density profile.


The bullet cluster 1E-0657 (red: X-ray, blue: weak lensing, and galaxies in the optical) showing evidence for collisionless dark matter.

- While the X-ray emission probes the gas distribution, which is effectively collisional, the individual cluster galaxies are collisionless tracers of the cluster potential. Weak lensing traces the projected total mass distribution that is dominated by cold dark matter. The composite image of the bullet cluster to the right-hand side shows a post-merging system where a small cluster ("the bullet") has passed from the left to the right and caused a disturbance of the main cluster along its path. The location of the weak-lensing mass centroid coincides with the collisionless galaxy distribution while the gas (in form of the bullet) lags behind. This provides evidence that dark matter is a collisionless or weakly interacting component with a long mean-free-path and does not feel the analog of a hydrodynamic pressure force. This observation puts constraints on the cross section of self-interacting dark matter which cannot be too large in order to be consistent with near-collisionless property of the bullet cluster data.

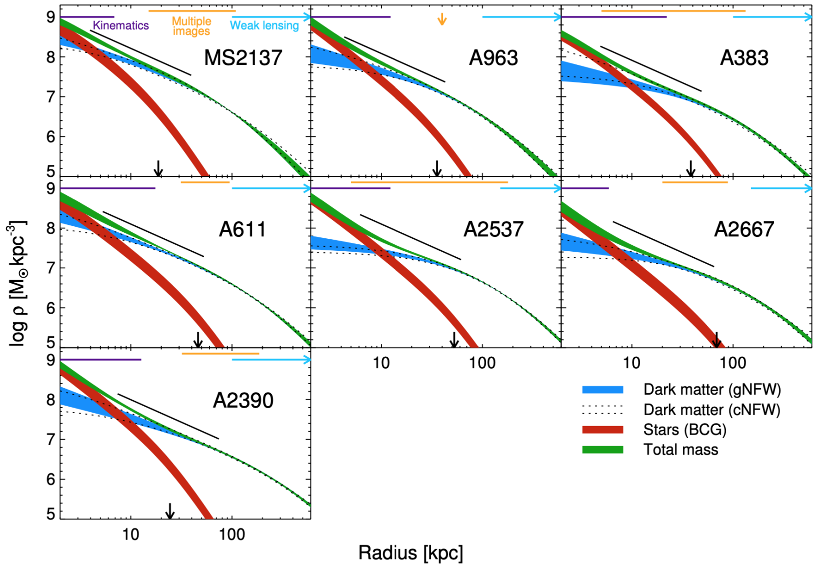


Figure 4.44: Best-fit radial density profiles for the dark matter halo, stars in the BCG, and their total in each cluster. The spatial extent of each data set (galaxy kinematics, strong and weak lensing, respectively) is indicated at the top of each panel. The black line segment has the slope $\rho \propto r^{-1.13}$ that is the average of clusters in the Phoenix dark matter-only simulations. Arrows at the bottom of each panel indicate three-dimensional half-light radius of the BCG (Newman et al. 2013).

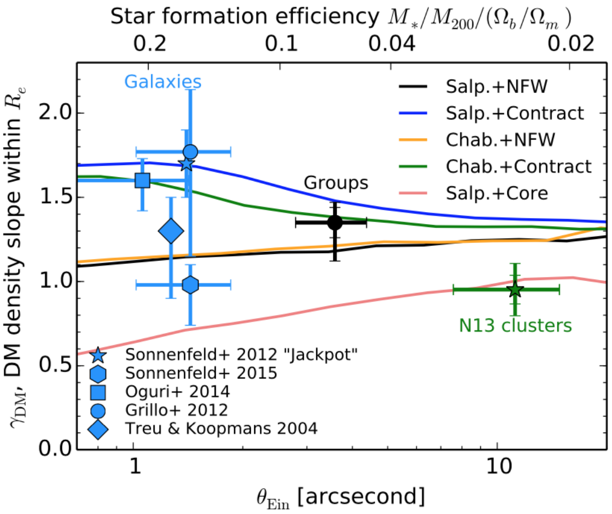


Figure 4.45: Constraints on the average mass-weighted dark matter-density slope within the effective radius, $\gamma_{\rm DM} = \langle d\ln\rho_{\rm DM}/d\ln r\rangle_M$ for galaxy, group, and cluster-scale mass early-type galaxy lenses. Lines assume different initial mass functions (Salpeter, Chabrier) and dark matter density profiles (cored NFW: $\gamma_{\rm DM} = 0$, NFW: $\gamma_{\rm DM} = -1$, adiabatically contracted NFW: $\gamma_{\rm DM} = -1.5$; Newman et al. 2015).

Appendix A

Additional Material

A.1 Equation of State and Mean Molecular Weight

- The ICM is to very good approximation described as an ideal gas. The equation of state for an ideal gas relates gas pressure P, temperature T, and number of particles N within a volume V via

  $$PV = Nk_\mathrm{B}T \quad \text{or} \quad P = nk_\mathrm{B}T, \tag{A.1}$$

  where $n = N/V$ is the number density of the particles.

- The mass density ρ equals the number density n of particles times their mean mass $\bar{m}$,

  $$\rho = \bar{m}n \quad \text{and} \quad \bar{m} = \mu m_\mathrm{H}. \tag{A.2}$$

  Here, we introduced the mean molecular weight μ to relate $\bar{m}$ and the mass of a hydrogen atom m_H. Combining these expressions yields the equation of state in its most common form,

  $$P = \frac{\rho k_\mathrm{B}T}{\mu m_\mathrm{H}}. \tag{A.3}$$

- In general, μ depends on the ionization state and the composition of the medium. To evaluate μ we need to determine the density-averaged particle mass $\bar{m}$ by summing over all particles. Let n_j be number density of atoms of type j, m_j be corresponding mass, and n_e be number density of electrons. Then

  $$\bar{m} = \frac{\sum_j n_j m_j + n_\mathrm{e}m_\mathrm{e}}{\sum_j n_j + n_\mathrm{e}} \approx \frac{\sum_j n_j m_j}{\sum_j n_j + n_\mathrm{e}}, \tag{A.4}$$

  because the electron mass is negligible. The mass of an atom m_j is determined from its mass number A_j (the number of protons plus neutrons), $m_j \approx A_j m_\mathrm{H}$ if we neglect the binding mass. Combining these expressions yields the mean molecular weight

  $$\mu = \frac{\bar{m}}{m_\mathrm{H}} = \frac{\sum_j n_j A_j}{\sum_j n_j + n_\mathrm{e}}. \tag{A.5}$$

- We identify two important limiting cases, neutral gas and fully ionized gas:

  $$\mu = \frac{\sum_j n_j A_j}{\sum_j n_j} \qquad \text{for neutral gas with } n_\mathrm{e} = 0, \tag{A.6}$$

  $$\mu = \frac{\sum_j n_j A_j}{\sum_j n_j(1 + Z_j)} \qquad \text{for fully ionized gas, } n_\mathrm{e} = \sum_j n_j Z_j, \tag{A.7}$$

  where Z_j is the atomic number. In general, n_e is determined by solving Saha's equation.

- To work out the number densities n_j, we need to know the mass fractions of the different elements. We define the mass fraction X_j as the fraction *by mass* of gas that is made up of an element j. For primordial gas, helium has a mass fraction ~ 0.24, meaning 1 g of primordial gas (after completion of Big Bang nucleosynthesis) contains 0.24 g of helium. In the Sun, the helium mass fraction is ~ 0.28. In general, all mass fractions add up to unity,

$$\sum_j X_j = 1. \tag{A.8}$$

We adopt the commonly used conventions:

X_j for hydrogen is written as	X,
X_j for helium is written as	Y,
X_j combined for all remaining elements is	Z (metallicity),

$$X + Y + Z = 1. \tag{A.9}$$

- If we write the number density in terms of mass fractions,

$$n_j = \frac{\rho X_j}{m_{\rm H} A_j} \tag{A.10}$$

we obtain the mean molecular weight for neutral gas,

$$\mu = \frac{\sum_j X_j}{\sum_j \frac{X_j}{A_j}} = \left(\sum_j \frac{X_j}{A_j}\right)^{-1} = \left(X + \frac{Y}{4} + \left\langle\frac{1}{A_j}\right\rangle Z\right)^{-1} \tag{A.11}$$

where $\langle 1/A_j \rangle$ is an average over metals ($\sim 1/15.5$ for solar composition). For *fully ionized gas*, we obtain

$$\mu = \left(\sum_j \frac{X_j}{A_j}(1 + Z_j)\right)^{-1} \approx \left(2X + \frac{3Y}{4} + \frac{Z}{2}\right)^{-1}, \tag{A.12}$$

Here, for each metal we assume $(1 + Z_j)/A_j \approx 1/2$.

- For solar abundance, we have $X = 0.7$, $Y = 0.28$, $Z = 0.02$ so that we get

$$\mu = 1.3 \quad \text{for neutral gas,} \tag{A.13}$$

$$\mu = 0.62 \quad \text{for fully ionized gas.} \tag{A.14}$$

- In the case of *fully ionized primordial gas* consisting mostly of hydrogen and helium with a helium mass fraction $Y = 1 - X \approx 0.24$ we obtain

$$\mu = \left(\sum_{j=1,2} \frac{X_j}{A_j}(1 + Z_j)\right)^{-1} = \frac{4}{8X + 3Y} = \frac{4}{5X + 3} \approx 0.588, \tag{A.15}$$

$$n_{\rm e} = n_{\rm H} + 2n_{\rm He} = \frac{X + 1}{2}\frac{\rho}{m_{\rm H}} \approx 0.88\,\frac{\rho}{m_{\rm H}}. \tag{A.16}$$

- In the case of partially ionized gas, it is conventional to introduce the electron density fraction relative to the hydrogen number density, which can be obtained by means of Saha's equation,

  $$\tilde{x}_\mathrm{e} = \frac{n_\mathrm{e}}{n_\mathrm{H}}, \quad \text{so that} \quad n_\mathrm{e} = \frac{\rho}{m_\mathrm{H}} X \tilde{x}_\mathrm{e}. \tag{A.17}$$

  This enables us to write the mean molecular weight for *partially ionized primordial gas*,

  $$\mu = \left(\sum_{j=1,2} \frac{X_j}{A_j} + X\tilde{x}_\mathrm{e} \right)^{-1} = \left(X + \frac{Y}{4} + X\tilde{x}_\mathrm{e} \right)^{-1} = \frac{4}{3X + 1 + 4X\tilde{x}_\mathrm{e}}. \tag{A.18}$$

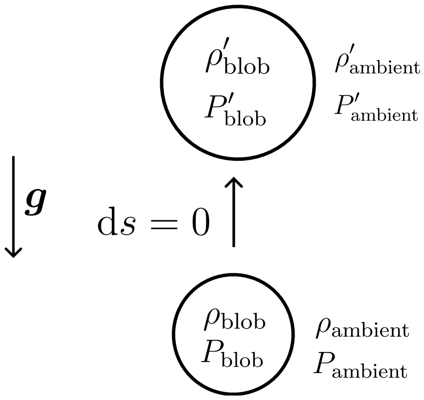
- Finally, in hadronic proton-proton interactions of cosmic rays with atoms of the ambient gas, we need to work out the number density of target nucleons,

  $$n_\mathrm{n} = \sum_j A_j n_j \approx n_\mathrm{H} + 4 n_\mathrm{He} = \frac{\rho}{m_\mathrm{H}} \left(X + 4 \frac{1 - X}{4} \right) = \frac{\rho}{m_\mathrm{H}}, \tag{A.19}$$

  Here, we again assume that we are dealing with a gas that is mainly composed of hydrogen and helium.

A.2 Schwarzschild Criterion for Convective Instability

- Fluid motions driven by thermal gradients (thermal convection) is a common and important phenomenon in clusters. Complementary to our full perturbation analysis of the hydrodynamic equations in Sect. 3.1.3, here we employ considerably simpler thermodynamic arguments to derive the stability conditions for a gravitationally stratified atmosphere, which we encounter in clusters. Unlike the perturbation analysis, the simpler derivation here does not provide insights into the timescale of the buoyant response.



Setup for the Schwarzschild criterion for convective instability.

- We assume a stratified atmosphere in which the gravitational field is pointing downwards and consider an adiabatic ($\mathrm{d}s = 0$) upward displacement of a small volume element (a "blob") so that the pressure and density change according to $P_{\rm blob} \to P'_{\rm blob}$ and $\rho_{\rm blob} \to \rho'_{\rm blob}$. The ambient background pressure and density is in hydrostatic equilibrium, which causes the ambient pressure and density to fall off at larger heights. The change in density of the blob is related to the change of its surroundings by

$$(\mathrm{d}\rho)_{\rm blob} = \left(\frac{\partial\rho}{\partial P}\right)_s \mathrm{d}P \tag{A.20}$$

while the corresponding change in density of the ambient medium is

$$(\mathrm{d}\rho)_{\rm ambient} = \left(\frac{\partial\rho}{\partial P}\right)_s \mathrm{d}P + \left(\frac{\partial\rho}{\partial s}\right)_P \mathrm{d}s, \tag{A.21}$$

where $\mathrm{d}P$ and $\mathrm{d}s$ are the difference of the pressure and the specific entropy of the medium at the new position that arises as a result of mechanical and thermal balance.

- The blob will continue to move under the influence of buoyancy, i.e., it will continue to fall (rise) if its density increase (decrease) exceeds (falls short of) that of the ambient medium

$$\begin{aligned} (\mathrm{d}\rho)_{\rm blob} &> (\mathrm{d}\rho)_{\rm ambient} \quad \text{for instability of a } \textit{downward} \text{ displacement,} \\ (\mathrm{d}\rho)_{\rm blob} &< (\mathrm{d}\rho)_{\rm ambient} \quad \text{for instability of a } \textit{upward} \text{ displacement.} \end{aligned} \tag{A.22}$$

Because we assume small blobs, we can identify the $\mathrm{d}P$ terms in Eqns. (A.20) and (A.21) and we obtain an instability condition for the ambient medium,

$$\begin{aligned} \left(\frac{\partial\rho}{\partial s}\right)_P \mathrm{d}s &< 0 \quad \text{for instability of a } \textit{downward} \text{ displacement,} \\ \left(\frac{\partial\rho}{\partial s}\right)_P \mathrm{d}s &> 0 \quad \text{for instability of a } \textit{upward} \text{ displacement.} \end{aligned} \tag{A.23}$$

- To simplify this, we use Maxwell's relation for thermodynamic potentials, in particular for the specific enthalpy, $h(s, P) = u + P\tilde{V}$ implying $\mathrm{d}h = T\mathrm{d}s + \tilde{V}\mathrm{d}P$, where u is specific internal energy and $\tilde{V} = \rho^{-1}$ is specific volume. These relations state

$$\left(\frac{\partial h}{\partial s}\right)_P = T, \quad \left(\frac{\partial h}{\partial P}\right)_s = \tilde{V} = \rho^{-1}. \tag{A.24}$$

Permutability of the second partial derivatives implies

$$\left(\frac{\partial^2 h}{\partial s \partial P}\right) = \left(\frac{\partial T}{\partial P}\right)_s = \left(\frac{\partial \rho^{-1}}{\partial s}\right)_P = -\frac{1}{\rho^2}\left(\frac{\partial \rho}{\partial s}\right)_P \tag{A.25}$$

- All thermodynamic substances have temperatures that increase upon adiabatic compression, and Eqn. (A.25) implies

$$\left(\frac{\partial T}{\partial P}\right)_s > 0 \quad \Longleftrightarrow \quad \left(\frac{\partial \rho}{\partial s}\right)_P < 0. \tag{A.26}$$

Using this relation, Eqn. (A.23) becomes the Schwarzschild criterion for instability,

$$\begin{aligned} & \mathrm{d}s > 0 && \text{in the direction of gravity,} \\ & \mathrm{d}s < 0 && \text{in the direction opposite of gravity,} \\ \Longrightarrow \quad & \frac{\mathrm{d}s}{\mathrm{d}r} < 0 && \text{for convective instability,} \end{aligned} \tag{A.27}$$

provided the center of gravity is at $r = 0$.

- To remember this result, we associate specific entropy with buoyancy. If the entropy of the cluster atmosphere increases inwards, then more buoyant material underlies less buoyant material and the medium has the tendency to overturn. The overturning mixes entropy – high-entropy (i.e., hotter than average) material rises and low-entropy (colder than average) material sinks in the cluster atmosphere so there is convective transport of heat outward which results in an approximately constant entropy distribution as the final state. If we assemble the intra-cluster gas with pockets of different entropy (which is expected if we have frequent mergers of objects of different mass and characteristic entropy, i.e., of galaxies and groups) then the gas will convectively rearrange itself such that it obeys an increasing entropy profile outwards.

- Note that the derivation assumes no magnetic fields, i.e., that the heat conduction is strictly isotropic and that temperature can be compared along every direction. The presence of magnetic fields modifies this instability criterion and states that the temperature gradient is the source of free energy (rather than the entropy gradient, see Sect. 3.2.6.3).

Acknowledgements

I am indepted to many of my colleagues for inspiring discussions that helped shaping my understanding of cosmological structure formation, galaxy clusters, and the beautiful (astro-)physics that is at work within these objects. In particular, I would like to thank Matthias Bartelmann, Roger Blandford, Torsten A. Enßlin, Tom W. Jones, Rüdiger Pakmor, Ewald Puchwein, Volker Springel, Henk Spruit, G. Mark Voit, David H. Weinberg, and Simon D. M. White (in alphabetic order), from whom I learned much over the last two decades, some of which found its way into these lecture notes. I also thank Larissa Tevlin for preparing some of the electronic versions of the figures.

Potsdam, 1. February 2022, Christoph Pfrommer