Particle acceleration and non–thermal activity during large scale structure formation

Stefano Gabici

Firenze, March 2004

Supervisors: Dr. Pasquale Blasi
Prof. Franco Pacini
Coordinator: Prof. Mario Perinotto
Alla mia Mamma
Contents

Introduction 1

1 Large scale structure formation 5
  1.1 Brief introduction to cosmology and structure formation . . . . . . . 6
  1.1.1 The expanding universe ........................................ 6
  1.1.2 Linear growth of perturbations ............................... 8
  1.1.3 Non–linear evolution of fluctuations: spherical model ....... 11
  1.2 Press–Schechter approach: hierarchical structure formation .... 13
    1.2.1 Comoving number density of dark matter halos .......... 13
    1.2.2 The cloud in cloud problem ................................. 15
    1.2.3 Building the merger tree of a cluster .................... 17
  1.3 Secondary infall onto a gravitationally bound object .......... 20

2 Diffusive shock acceleration: linear theory 23
  2.1 Stochastic acceleration: second order Fermi mechanism ........ 23
  2.2 Acceleration of particles at shock fronts: first order Fermi mechanism 25
    2.2.1 The Bell approach ......................................... 25
    2.2.2 The particle transport equation ........................... 27
    2.2.3 Reacceleration of seed particles .......................... 28

3 Extended Press–Schechter formalism and particle acceleration: a novel approach 29
  3.1 Non–thermal emission from clusters of galaxies: observations and models .................................................. 30
    3.1.1 Cosmic ray confinement in the intracluster medium ...... 32
    3.1.2 Non–thermal emission and cluster mergers ............... 34
  3.2 Merger trees and strength of the shocks in the intracluster medium 35
    3.2.1 Shocks during cluster mergers ............................. 36
    3.2.2 The effects of a local overdensity ......................... 39
    3.2.3 X–Ray observation of shock waves in the intracluster medium 41
  3.3 Particle acceleration at merger shocks and related non–thermal emission 42
    3.3.1 Particle spectra and related non–thermal emission .... 42
4 Gamma ray emission during large scale structure formation 51
4.1 Gamma ray astrophysics: a brief overview ....................... 51
4.2 Gamma rays from large scale structures in the universe: review of past theoretical and observational results ................... 54
4.3 Detectability of clusters with space telescopes .................. 55
4.3.1 Shock acceleration during structure formation ................ 56
4.3.2 Source counts for merging and accreting clusters ............. 59
4.4 The diffuse gamma ray background from large scale structure formation 63
4.5 Comparison with other theoretical results ....................... 65
4.5.1 Semi‐analytical models ........................................ 67
4.5.2 Numerical simulations .......................................... 68
4.6 Clusters of galaxies as TeV sources ............................. 70
4.6.1 Absorption of extragalactic TeV photons by infrared photons . 70
4.6.2 Detectability of clusters with Čerenkov telescopes ............ 71

5 Turbulent Alfvénic reacceleration: protons and electrons 73
5.1 Origin and spectrum of the relativistic particles in the intracluster medium ......................................................... 74
5.1.1 Electrons ......................................................... 75
5.1.2 Protons ......................................................... 76
5.2 Alfvénic reacceleration of relativistic particles .................... 77
5.2.1 From fluid turbulence to Alfvén waves: the Lighthill mechanism 80
5.2.2 Basic equations and time evolution ........................... 81
5.2.3 Turbulent cascade ............................................ 82
5.2.4 Damping processes ........................................... 84
5.3 Quasi stationary solutions ........................................... 87
5.3.1 The spectrum of Alfvén Waves ................................ 87
5.3.2 Electron acceleration ........................................... 89
5.3.3 Proton acceleration ............................................ 90
5.3.4 The Wave–Proton Boiler ....................................... 91
5.4 Non‐thermal emission from galaxy clusters ....................... 93
5.4.1 Cluster mergers and turbulence ................................ 93
5.4.2 Constraining the model parameters ........................... 95
5.4.3 A simplified models for Radio Halos and Hard X‐ray emission 99

6 Future perspectives: non‐linear shock acceleration during structure formation 105
6.1 Non‐linear shock acceleration: a semi‐analytical approach ......... 106
6.2 The problem of multiple solutions .................................. 109
6.3 Comparison with an alternative approach .......................... 111
6.4 Thermal leakage as a recipe for injection: preliminary results ..... 113
6.5 Implications for non‐thermal phenomena during structure formation . 116
Introduction

The universe, when observed on scales much smaller than the Hubble radius, is highly structured. Matter is concentrated in virialized objects, the biggest of them being clusters and groups of galaxies. There is now general agreement on the fact that small structures formed at earlier times, while bigger structures form later as the result of the hierarchical merging of smaller objects. During a merger event, shocks are driven by gravity in the diffuse baryonic component, which is heated up to the observed hot temperature. The presence of this diffuse hot gas in clusters makes them powerful X-ray thermal sources, detectable by space telescopes.

A direct evidence supporting this hierarchical view of structure formation was recently obtained from high resolution X-ray observations of clusters performed by the Chandra and XMM satellites. These observations have shown signatures of ongoing mergers in a number of clusters, such as disturbed morphologies in the brightness distribution and hot shocks in the intracluster medium (Forman et al., 2002).

Besides mergers, the growth of structures proceeds also through the so called secondary infall onto already formed objects (Bertschinger, 1985). In fact, clusters continuously attract unbound matter in their neighbourhood and their mass increases due to accretion of new material. The existence of such non virialized gas outside clusters and groups of galaxies is suggested by cosmological numerical simulations (Cen & Ostriker, 1999). This gas is structured in low density, relatively cold laments whose X-ray emission is expected to be tenuous and hardly detectable by present day instruments. Filaments connect more dense, hot regions, forming the so called cosmic web.

The first evidence for non-thermal emission from large scale structures came from radio observations of the Coma cluster, which is a massive \((M \sim 10^{15} M_\odot)\), X-ray bright \((L_X \sim 10^{45} \text{erg/s})\), nearby \((z \sim 0.0231)\) cluster of galaxies. The first radio map of the Coma cluster region made at 408 MHz revealed the presence of a large size \((\sim 45')\) diffuse source in the middle of the cluster (Large, Mathewson & Haslam, 1959). This source, which is now considered the prototypical example of cluster radio halos, was for the first time classified in this way by Willson (1970). To date, radio halos are observed from several rich clusters, and the detection rate significantly increases with X-ray luminosity, reaching \(\sim 30\%\) if only the brightest clusters, with X-ray luminosity greater than \(10^{45} \text{erg/s}\) are considered (Feretti,
It is now well established that the observed radio emission is truly diffuse and is not associated with single cluster galaxies. Moreover, this emission is believed to be non-thermal synchrotron radiation and this fact constitutes a strong evidence for the presence of relativistic electrons and magnetic field in the intracluster medium. Additional evidence for non-thermal activity in clusters comes from hard X-rays (Fusco-Femiano et al., 2003) and possibly extreme UV/soft-X (Bowyer, 2003) observations, that seem to show an excess with respect to the expected thermal radiation. Hard X-ray observations can be interpreted as inverse Compton scattering off the cosmic microwave background radiation from the same population of electrons responsible for the radio emission (Fusco-Femiano et al., 2003). The origin of the UV/soft-X excess is more controversial and there is still some debate about its thermal or non-thermal origin (Bowyer, 2003).

Even if the existence of relativistic electrons is firmly proved by observations, the origin of these particles and the mechanisms through which they are accelerated are still unknown. Since the majority of the acceleration mechanisms at work in astrophysical sources accelerate protons as well, it is generally assumed that also an hadronic cosmic ray component is present in the intracluster medium. Relativistic protons can interact with thermal protons and produce charged pions which in turn decay, generating relativistic electrons and positrons in the cluster volume. For this reason, non-thermal electrons are generally referred to as primaries if they are directly accelerated through some mechanism, or as secondaries if they are the final products of the decay of charged mesons produced in inelastic collisions between cosmic ray protons and thermal protons which constitute the intracluster medium.

Models explaining radio observations as synchrotron emission from secondary electrons were proposed by Dennison (1980) and extensively studied after the discovery that cosmic ray protons remain diffusively confined within the cluster volume for cosmological times, without losing their energy (Berezinsky, Blasi & Ptuskin, 1997; Völk, Aharonian & Breitschwerdt, 1996). This means that the present day hadronic cosmic ray population in a cluster is the result of the pile up of all the relativistic protons injected in the intracluster medium during the whole cluster lifetime. As a consequence, synchrotron emission from secondary electrons is expected to be amplified. High energy gamma rays can also be produced during inelastic proton–proton scattering, and for this reason, the expected gamma ray luminosity from clusters of galaxies is enhanced as well. However, despite some recent claims, gamma rays from clusters of galaxies have not been detected yet (Reimer et al., 2003).

Harris et al. (1980) proposed that radio halos might be powered by cluster merger events. The connection between cluster mergers and non-thermal activity in the intracluster medium is appealing, especially on the basis of considerations about the energy needed to power the radio halos. In fact, a very small fraction of the huge amount of gravitational energy released during a merger is needed to be converted
Introduction

into relativistic electrons in order to explain observations. Recently, this idea has also received observational support from the fact that clusters hosting a radio halo seem to show the signatures of a recent major merger event (Buote, 2001). Obviously, this correlation is not expected in the context of models based on secondary electrons, in which the level of radio emission depends on the whole history of cosmic ray sources in clusters and not on recent events such as ongoing mergers.

It is well known that collisionless shock waves can accelerate particles up to ultra-relativistic energies via the first order Fermi mechanism (Blandford & Eichler, 1987). For this reason, large scale shocks associated with cluster mergers or secondary infall of matter have been often considered to be the sources of cosmic rays in the intracluster medium. In several earlier works, ad hoc assumptions were made about the values of the shock Mach numbers, in order to reproduce radio observations. As we will demonstrate in the following, these assumptions are incorrect and lead to several predictions which are in strong disagreement with a number of observations. Moreover, models in which particles are accelerated at large scale shocks have problems in reproducing the observational features of radio halos. The details in the radio halo spectra cannot be explained in the context of models involving secondary electrons, while primary electrons cannot explain the $\sim Mpc$ size of halos due to their very short diffusion length. In an alternative and promising approach, electrons are supposed to be continuously reaccelerated by MHD waves injected in the intracluster medium during mergers (Jaffe, 1977; Schlickeiser et al., 1987; Brunetti et al., 2001a). Both the extended size and spectral features of radio halos can be well reproduced by these models.

Recently, Loeb & Waxman (2000) proposed a connection between the observed isotropic extragalactic gamma ray background and the process of large scale structure formation. In particular, they suggested that this background might be due to inverse Compton scattering from relativistic electrons accelerated at merger and accretion shocks in clusters and filaments. More accurate calculations revealed that this mechanism is too inefficient to saturate the whole observed background (Gabici & Blasi, 2003b; Keshet et al., 2003). However, the detection of gamma rays from single forming clusters with next generation space telescopes appears to be feasible (Gabici & Blasi, 2004).

The aim of this work is to study the acceleration of particles during the process of large scale structure formation and the related non-thermal emission from radio waves to gamma rays. The thesis is structured as follows. In chapters 1 and 2 we briefly review the theory of hierarchical structure formation and diffusive shock acceleration respectively. In chapter 3 we present a novel recipe to estimate in a self-consistent way the Mach number of merger related shocks and we use it to investigate the relevance of such shocks in particle acceleration. Merger shocks are found to be weak, and consequently unable to accelerate particles with spectra flat enough to result in appreciable non-thermal radiation.

The presented recipe is generalized in order to include strong accretion shocks in
chapter 4, which is focused on the issue of the detectability of high energy gamma rays from forming structures with next generation GeV and TeV telescopes. Predictions seem to be very optimistic for space borne GeV telescopes such as GLAST and AGILE, while less enthusiastic conclusions can be drawn for ground based Čerenkov telescopes (Gabici & Blasi, 2004). We also give a more realistic estimate of the contribution to the extragalactic diffuse gamma ray background from forming clusters of galaxies, concluding that it can add up to at most 10\% of the observed background (Gabici & Blasi, 2003b), at odds with the above mentioned earlier claims (Loeb & Waxman, 2000).

In chapter 5, we consider the possibility that both radio and hard X-ray radiation from clusters might be produced by electrons reaccelerated by Alfvénic turbulence generated in the cluster volume during major mergers. We consider for the first time the most general situation in which relativistic electrons, thermal protons and relativistic protons exist within the cluster volume and their interactions with waves are treated in a fully time dependent way.

A subjective view of the future perspectives in this field of research is outlined in chapter 6.
Chapter 1

Large scale structure formation

The current, successful standard cosmological model describing our universe is based on the so-called Cosmological Principle. This Principle states that the universe is homogeneous in the large-scale average, that is, every region has the same properties of any other region, and isotropic, that is, the universe looks the same if we turn our telescope towards different directions.

The validity of the isotropy hypothesis can be checked, for example, by means of precision measurements of the angular distributions of the X-ray or microwave background radiations or through deep galaxy counts. These measurements tell us that the universe is very close to isotropy if observed from the Earth. Moreover, if we reasonably think not to be in a very special place in the universe, then homogeneity naturally follows from isotropy.

The homogeneity in the large-scale average can also be constrained by measurements of the large scale fluctuations in the Cosmic Microwave Background radiation (CMB). Large-scale departures from homogeneity, if present, would cause the expansion of the universe across the Hubble length to differ by a fractional amount $\delta$, measured in orthogonal directions. This would produce a quadrupole anisotropy of the same order $\delta$ in the CMB temperature, since the CMB temperature scales with the expansion factor. The CMB is known to be isotropic to better than one part in $10^4$, so that the universe is expected to be very close to homogeneity on scales comparable with the Hubble length, with mass fluctuations of the order $\delta M/M < 10^{-4}$.

These fluctuations become larger and larger if the smoothing radius is progressively reduced and the value $\delta M/M \sim 1$ is reached averaging over a scale which is about one percent of the Hubble length. This is a consequence of the fact that on small scales the universe is structured and the matter is concentrated in stars, galaxies, clusters of galaxies or superclusters.

The observed isotropy of the CMB can also be interpreted in the following, equivalent way: the universe was extremely uniform at recombination, with fluctuations in the energy density and in the gravitational potential of roughly one part in $10^5$. Such small “primordial” fluctuations are generated in the early universe and
1. Large scale structure formation

are believed to be originated by quantum fluctuations during inflation or by topological defects generated during cosmological phase transitions. Once produced, fluctuations evolve mainly due to gravitational instability and grow, leading to the formation of the observed large scale structures.

The matter component in the universe is dominated by non–baryonic particles, called dark matter particles, which have not been directly observed yet. These particles, whose existence can only be deduced by the effects of their own gravity, play an important role during the assembly of structures. In particular, very different scenarios are expected for different assumptions on the temperature of dark matter at the time of its decoupling from the baryonic component.

Dark matter particles are called hot if they are relativistic at decoupling. In this case they can smear out small scale perturbations allowing only large scale fluctuations to evolve. As a consequence big structures form first and subsequently fragment to originate smaller objects. Conversely, in the Cold Dark Matter model, where the dark matter is assumed to be non–relativistic at decoupling, structure formation starts with small scales, and big structures are assembled hierarchically through mergers. These two opposite pictures are often referred to as “top–down” and “bottom–up” scenarios.

In the following we will describe in more detail the process of large scale structure formation in the framework of the Cold Dark Matter model, which successfully reproduces several observational results.

1.1 Brief introduction to cosmology and structure formation

1.1.1 The expanding universe

In General Relativity, an isotropic and homogeneous universe is described by the Robertson–Walker metric:

$$ds^2 = c^2 dt^2 - a^2(t) \left[ \frac{dr^2}{1 - kr^2} + r^2 \left( d\theta^2 + \sin^2 \theta d\phi^2 \right) \right]$$  \hspace{1cm} (1.1)

where $r, \theta$ and $\phi$ are spherical comoving coordinates, $a(t)$ is the cosmic scale factor normalized to his present day value and $k$ represents the spatial curvature, which is 0 for a flat universe and $+(-)1$ for a closed (open) universe.

The substitution of equation (1.1) together with the energy–momentum tensor for a perfect fluid in the Einstein field equations yields:

$$\left( \frac{\dot{a}}{a} \right)^2 + \frac{k c}{a^2} = \frac{8\pi G}{3} \rho + \frac{\lambda c^2}{3}$$  \hspace{1cm} (1.2)

$$\frac{d}{da} \left( ra^3 \right) + 3pa^2 = 0$$  \hspace{1cm} (1.3)
1.1. Brief introduction to cosmology and structure formation

describing respectively the dynamics of the universe and the energy–momentum conservation. Here \( \rho \) represents all the components (except vacuum) that contribute to the energy density of the universe. At the present time, the dominant contribution comes from non–relativistic matter with negligible pressure \( \rho_m \). A radiation component \( \rho_r \), energetically subdominant and presently non interacting with matter, is also present. Due to their different equation of state \( p = p(\rho) \), these components evolve in time in a different way:

\[
\begin{align*}
\rho_m &\propto a(t)^3 \\
\rho_r &\propto a(t)^4
\end{align*}
\]

suggesting that their relative contribution to the total energy of the universe was different in the past, with radiation dominating over matter at earlier epochs. The two contributions are equal at the equipartition redshift that can be estimated to be equal to \( z_{eq} \sim 10^4 \).

Here we are interested in the growth of primordial perturbations and in the subsequent formation of large scale structures in the universe. Since the growth of perturbations inside the horizon is inhibited if radiation dominates the total energy, we will consider in the following only their evolution at redshifts well below the equipartition value. This allows us to neglect the contribution of radiation to \( \rho \).

Equation (1.2) can be rewritten in a more convenient form making use of the quantities:

\[
\begin{align*}
H(t) &= \frac{\dot{a}}{a} \\
\Omega_m &= \frac{\rho}{\rho_{cr}} \\
\Omega_\Lambda &= \frac{\lambda c^2}{3H_0}
\end{align*}
\]

which represent the expansion rate at cosmic time \( t \) and the contribution to the total density from matter and vacuum (or cosmological constant), both evaluated at the present time. These contributions are normalized to the critical density:

\[
\rho_{cr} = \frac{3H_0^2}{8\pi G}
\]

deścribed as the present time density in the Einstein–de Sitter universe \((k = 0, \lambda = 0)\). The Hubble constant \( H_0 \) represents the current value of the expansion rate.

Since strong evidence has been presented in favour of a flat space–time (De Bernardis et al., 2000) we will use in the following \( k = 0 \), so that eq. (1.2) simplifies to:

\[
\frac{H(t)}{H_0} = \left[ \frac{\Omega_m}{a^3} + \Omega_\Lambda \right]^{1/2}
\]
1. Large scale structure formation

As already specified, we are neglecting every contribution from radiation, since we are considering structure formation at redshifts well below the equipartition value. Evaluating equation (1.10) at the present time, we obtain the relation:

$$\Omega_{\text{tot}} = \Omega_m + \Omega_\Lambda = 1 \quad (1.11)$$

The explicit expression for the scale factor can be obtained integrating equation (1.10) for any fixed values of the parameters ($\Omega_m, \Omega_\Lambda$). In the Einstein–de Sitter universe the $\Lambda$–term vanishes and we get the well known result:

$$a(t) = \left( \frac{3H_0 t}{2} \right)^{2/3} \quad (1.12)$$

while in a flat universe with positive cosmological constant the scale factor evolves in time as follows:

$$a(t) = \left( \frac{\Omega_m}{\Omega_\Lambda} \right)^{1/3} \sinh^{2/3} \left( \frac{3H_0 \sqrt{\Omega_\Lambda}}{2} t \right) \quad (1.13)$$

To complete the picture, we need to know the values of the cosmological parameters. The combination of observations of distant type Ia supernovae (Riess et al., 1998; Perlmutter et al., 1999) with CMB data (De Bernardis et al., 2000) has led to the establishment of the so called “concordance model” with:

$$H_0 \sim 70\text{km/s/Mpc} \ , \ \Omega_m \sim 0.3 \ , \ \Omega_\Lambda \sim 0.7 \quad (1.14)$$

These values have been broadly confirmed by the recent WMAP results (Spergel et al., 2003).

1.1.2 Linear growth of perturbations

We study now the evolution of density fluctuations in an expanding universe. To this purpose, it is convenient to write the mass density in the form:

$$\rho(r, t) = \rho_b(t) [1 + \delta(r, t)] \quad (1.15)$$

where $\rho_b(t)$ is the average background density and $\delta(r, t)$ is the density contrast at the comoving position $r$. As already said, we assume that the material pressure is small if compared with the mass density, so that the background density scales as $\rho_b \propto a(t)^{-3}$. Moreover, we start our analysis considering small fluctuations, $\delta << 1$, and neglecting non gravitational forces, such as gas pressure, on the material. Since we are mainly interested in sub–horizon fluctuations and we are considering small perturbations characterized by weak gravitational fields, we can adopt a Newtonian approach. Under these conditions the fluid is described by the continuity and Euler equations that, in comoving coordinates, can be written as follows (Peebles, 1993):

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta) \mathbf{u}] = 0 \quad (1.16)$$

$$\frac{\partial \mathbf{u}}{\partial t} + H(t) \mathbf{u} + \frac{1}{a} (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{a} \nabla \phi \quad (1.17)$$
1.1. Brief introduction to cosmology and structure formation

Here $u$ represents the (small) peculiar velocity with respect to the expanding Hubble flow $H(t)\mathbf{r}a(t)$. The fluid equations can be combined together with the Poisson’s equation for the gravitational potential $\phi$:

$$\nabla^2 \phi = 4\pi G \rho_0 a^2 \delta$$

(1.18)

and linearized dropping all the terms of order $u\delta$ or $u^2$ to obtain:

$$\frac{\partial^2 \delta}{\partial t^2} + 2H(t) \frac{\partial \delta}{\partial t} = 4\pi G \rho_0 \delta$$

(1.19)

This equation can be solved independently for any generic Fourier mode $\delta_k$ of the density contrast:

$$\delta_k(t) = \int d\mathbf{r} \delta(\mathbf{r}, t)e^{-ik\cdot\mathbf{r}}$$

(1.20)

where $k$ is the comoving wavenumber. In general the problem admits two independent solutions, which are decaying and growing with time respectively, with the latter dominating the density evolution. Thus, density fluctuations maintain the same shape in comoving coordinates, and their amplitude increases with time according to a linear growing factor $\delta \propto D(t)$, which is usually set equal to unity at present time.

It is easy to show that in the Einstein–de Sitter model the growing factor is simply proportional to the scale factor:

$$D(t) \propto a(t) \propto t^{2/3}$$

(1.21)

while in a flat universe with non vanishing $\Lambda$, the time evolution is described by (Kitayama, 1997):

$$D(t) \propto \sqrt{1 + \frac{2}{y^3} \int_0^y du \left(\frac{u}{2 + u^3}\right)^{3/2}}$$

(1.22)

$$y = 2^{1/3} \left(\frac{1}{\Omega_m} - 1\right)^{1/3} a(t)$$

(1.23)

The linear growth of fluctuations can be modified at small scales by dissipative processes (free streaming, Silk damping) or by pressure effects (Jeans mass effects), which can affect the density evolution in different ways and with different relative weight depending on the nature of dark matter (Peacock, 1999). All the relevant physics is embedded in the transfer function defined as the ratio of the present day amplitude of the fluctuations to its value at a given time $t$:

$$T_k = \frac{\delta_k(t_0)}{\delta_k(t)D(t)}$$

(1.24)

The behaviour of $T_k$ has been investigated by means of numerical simulations and for a Cold Dark Matter dominated universe the total modification of the fluctuation
1. Large scale structure formation

amplitudes with respect to its primordial shape is described by the following fitting formula (Bardeen et al., 1986):

$$ T_k = \frac{\ln(1 + 2.34q)}{2.34q} \left[ 1 + 3.89q + (16.1q)^2 + (5.46q)^3 + (6.71q)^4 \right]^{-1/4} \tag{1.25} $$

where $q = k/(h\Gamma\text{Mpc})$, $h = H_0/(100\text{km/s/Mpc})$ and $\Gamma$ is the shape parameter (Peacock & Dodds, 1994):

$$ \Gamma = \Omega_m h \left( \frac{T_{CMB,0}}{2.7K} \right)^2 e^{-\Omega_B(1+\frac{h}{m})} \tag{1.26} $$

which takes into account the non vanishing value of the baryon density $\Omega_B$.

If we assume the density fluctuation field to be random and Gaussian, all its properties are described by its power spectrum:

$$ P_k(t) = \langle |\delta_k(t)|^2 \rangle \tag{1.27} $$

whose primordial functional form can be obtained from theory. Conventional inflationary models predict a power law shape for the primordial power spectrum with slope close to unity:

$$ P_k(t_i) \propto k \tag{1.28} $$

This spectrum is known in the literature as the Harrison–Zel’dovich spectrum, and its time evolution is described by equation (1.24).

The only missing ingredient to complete the whole picture is the normalization of the power spectrum. To this purpose it is useful to consider the density field smoothed over a window function $W(r)$, set equal to 1 within a sphere of radius $R$, and 0 elsewhere (spherical top–hat filter). The smoothed field:

$$ \tilde{\delta}(r, t) = \int d\mathbf{r}' \delta(\mathbf{r}') W(|\mathbf{r} - \mathbf{r}'|) \tag{1.29} $$

follows a Gaussian distribution as well, with mass variance:

$$ \sigma^2(M, t) = \frac{1}{(2\pi)^3} \int d\mathbf{k} P(k, t) \hat{W}^2(\mathbf{k}) \tag{1.30} $$

$$ \hat{W}(k) = \frac{3}{(kR)^3} [\sin(kR) - kR \cos(kR)] \tag{1.32} $$

where $M = (4\pi)/3R^3 \rho_b$ is the mass enclosed in the smoothing volume and $\hat{W}(\mathbf{k})$ is the Fourier transform of the smoothing function:

The normalization of the present day power spectrum is conventionally specified by the biasing factor:

$$ b = \sigma_8^{-1} = \sigma^{-1}(R = 8h^{-1}Mpc, t_0) \tag{1.33} $$
1.1. Brief introduction to cosmology and structure formation

whose value $b \sim 1.1$ can be estimated using, for example, CMB data sets (Spergel et al., 2003).

The current mass variance relative to a primordial Harrison–Zel’dovich power spectrum can be well fitted by (Kitayama, 1997):

$$\sigma \propto \left[1 + 2.208m^p - 0.7668m^{2p} + 0.7949m^{3p}\right]^{-\frac{1}{2p}}$$

(1.34)

with $p = 0.0873$ and $m = M(\Gamma h)^3/(\Omega_m h^2)/10^{12}M_\odot$. The above approximation can be used in the range $10^{-7} \leq m \leq 10^5$.

1.1.3 Non–linear evolution of fluctuations: spherical model

In the previous section we have shown how small density perturbations evolve in the linear regime. The density contrast $\delta$ grows in time till its value approaches unity. At this point we can no longer use the linear theory and we have to consider the fully non–linear situation. To this purpose, we can consider an idealized model in which perturbations are supposed to be overdense spherical regions enclosing a mass $M$. The mass distribution inside the sphere is not important. In fact, the spherical symmetry and the absence of shell–crossing ensure that each sphere will evolve as a uniform sphere containing the same mass. The time evolution of the sphere in proper coordinates is described by the Newtonian equation of motion, modified in order to take into account the dynamical effects of the cosmological constant:

$$\frac{d^2r}{dt^2} = H_0^2\Omega_\Lambda r - \frac{GM}{r^2}$$

(1.35)

In an Einstein–de Sitter universe the $\Lambda$–term vanishes and the solution can be written in the parametric form:

$$r = A(1 - \cos \theta)$$

(1.36)

$$t = B(\theta - \sin \theta)$$

(1.37)

with $A^3 = GM^2$. Expanding up to order $\theta^5$ we obtain for small values of $t$:

$$r \sim \frac{A}{2} \left(\frac{6t}{B}\right)^{2/3} \left[1 - \frac{1}{20} \left(\frac{6t}{B}\right)^{2/3}\right]$$

(1.38)

while for the density perturbation we get:

$$\delta(< r, t) \sim \frac{3}{20} \left(\frac{6t}{B}\right)^{2/3} \propto a(t)$$

(1.39)

which is exactly the same result we obtained from linear theory. The further evolution of the sphere proceeds according to equation (1.39) until $\delta$ reaches unity, then
the density contrast grows faster than the linear prediction. The expansion velocity gradually slows down and vanishes for $\theta = \pi$ and $t = \pi B$, when the sphere reaches the turnaround radius. At this point the density contrast is $\delta = 9\pi^2/16 \sim 5.55$, higher than the linear prediction $\delta \sim 1.06$. The sphere starts to contract and, if only gravity is considered, it eventually collapses into a singularity at $\theta = 2\pi$. In a more realistic description, the system undergoes a dissipative process called violent relaxation before reaching the singularity (Lyndel–Bell, 1967). As a consequence, the kinetic energy of the infalling material is converted into random motions and the system virializes. The virialization happens when the gravitational energy is (in modulus) twice the kinetic energy and this allows us to estimate the virial radius as a function of the turnaround radius $r_{ta}$:

$$ r_v = \frac{r_{ta}}{2} \quad (1.40) $$

This expression also remains valid in an approximate way for other cosmological models (Lahav et al., 1991).

The overdensity of a virialized structure can be estimated assuming that the time needed in order to have equilibrium is the collapse time corresponding to $\theta = 2\pi$. Under these assumptions the ratio between the average density inside the virial radius and the background density is:

$$ \Delta_c = \frac{\bar{\rho}_v(t_v)}{\rho_0(t_v)} = 18\pi^2 \sim 178 \quad (1.41) $$

Another useful quantity is the density contrast linearly extrapolated to the collapse time, defined as the density contrast that the sphere would have at $t_v$ according to the linear theory. For an Einstein–de Sitter universe such an extrapolation yields:

$$ \delta_c = \frac{3}{20} (12\pi)^{2/3} \sim 1.69 \quad (1.42) $$

Following the same procedure, it is possible to derive these results for a flat, non-vanishing $\Lambda$ universe. The following expressions for the overdensity and the density contrast extrapolated to the collapse time are taken from the work by Nakamura & Suto (1997):

$$ \Delta_c = \left( \frac{r_{ta}}{r_v} \right)^3 \frac{2w_v}{\chi} \quad (1.43) $$

$$ \delta_c = \frac{3}{5} F \left( \frac{1}{3}, 1, \frac{11}{6}, -w_v \right) \left( \frac{2w_v}{\chi} \right)^{1/3} \left( 1 + \frac{\chi}{2} \right) \quad (1.44) $$

$$ \sim \frac{3(12\pi)^{2/3}}{20} (1 + 0.0123\log_{10} \Omega_v) \quad (1.46) $$
1.2 Press–Schechter approach: hierarchical structure formation

\[ w_v = \frac{1}{\Omega_v} - 1 \]  
\[ \chi = \frac{\Omega_\Lambda H_0^2 r_{\text{iso}}}{GM} \]

where \( F \) is the hypergeometric function of type (2,1) and \( \Omega_v \) is the density parameter evaluated at the virialization redshift \( z_v \):

\[ \Omega_v = \frac{\Omega_m (1 + z_v)^3}{\Omega_m (1 + z_v)^3 + \Omega_\Lambda} \]

1.2 Press–Schechter approach: hierarchical structure formation

1.2.1 Comoving number density of dark matter halos

As seen in the previous section, it is difficult to follow analytically the non–linear evolution of density perturbations except for simple, idealized situations. However, in spite of these difficulties, a number of attempts have been made in order to describe the hierarchical assembly of non–linear structures in the universe in a simple and analytical way. Most notably, starting from the assumption of Gaussianity of the density perturbation field and adopting the critical overdensity for collapse derived in the spherical model presented in §1.1.3, Press & Schechter (1974) proposed a prescription to estimate the mass function of virialized object as a function of cosmic time.

This can be done considering the initial density field smoothed over a spherical region with volume \( V \) containing a mass \( M = V \rho_v \). The condition for an object of mass \( M \) to collapse at a given time \( t \) is that the linearly extrapolated overdensity of the region enclosing that mass exceeds the threshold value \( \delta_c(t) \). Since the smoothed field is assumed to be Gaussian with mass variance \( \sigma^2(M, t) \) given by equation (1.30), the probability that a given point lies in a region with \( \delta > \delta_c(t) \) is:

\[ P(\delta(t) > \delta_c(t)|M) = \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{\delta_c(t)}{\sqrt{2} \sigma(M, t)} \right) \right] \]

Press and Schechter identified this probability with the probability that a given point belongs to a collapsed system of mass greater than \( M \). This can be understood in the following way: suppose to smooth out the density field over a certain region containing a given mass \( M \). Every point characterized by a linear overdensity greater than \( \delta_c \) will have \( \delta = \delta_c \) if some greater smoothing volume is adopted. This bigger volume obviously satisfies the condition for collapse and encloses a mass greater than \( M \). As a consequence, the considered point is counted as part of this more massive object. This is equivalent to say that, at a fixed time, all the existing objects have just reached the critical collapse overdensity \( \delta_c \).
The problem with this approach is that only half of the total mass of the universe is taken into account. Several discussions of this problem can be found in the literature, and a possible way out was proposed, among others, by Peacock & Heavens (1990) and (Bond et al., 1991) which modified the original approach by taking into account in a proper way the underdense regions. However, we will adopt here a more pragmatic approach and simply multiply the probability by a factor of 2, postponing the discussion of the problem to the next section. After this ad hoc assumption, we can finally write the expression for the fraction of the universe condensed into objects having mass greater than $M$:

$$ F(> M, t) = 1 - \text{erf} \left( \frac{\delta_c(t)}{\sqrt{2} \sigma(M, t)} \right) $$

(1.51)

It is interesting to note that we can replace $\delta_c(t)$ and $\sigma(M, t)$ in equation (1.51) with their values extrapolated to an arbitrary reference epoch $t_{ref}$, since they both scale as $D(t_{ref})$. In the following, we use the values extrapolated to the present time, denoted by $\delta_{c0}(t)$ and $\sigma(M)$.

The comoving number density of collapsed dark matter halos as a function of mass and cosmic time can be evaluated differentiating equation (1.51):

$$ \frac{dn}{dM}(M, t) = \left( \frac{\rho_0}{M} \right) \left| \frac{dF(> M, t)}{dM} \right| $$

(1.52)

where $\rho_0$ is the current background density. Written in an explicit form, equation (1.52) becomes (Press & Schechter, 1974):

$$ \frac{dn}{dM}(M, t) = \sqrt{\frac{2}{\pi}} \rho_0 \delta_{c0}(t) \left| \frac{d\sigma(M)}{dM} \right| \exp \left[ -\frac{\delta_{c0}^2(t)}{2\sigma^2(M)} \right] $$

(1.53)

Halo mass functions for different redshifts are plotted in figure (1.1), where a flat cosmology with $\Omega_m = 0.3$ is assumed. The figure well illustrates the hierarchical nature of clustering, showing that small structures form first while bigger structures form at lower redshifts. The position of the exponential cutoff represents the mass of the biggest bound structures at a given time. This mass increases with time and at present it has a value roughly equal to $10^{15} M_\odot$, which is the typical mass of a rich cluster of galaxies. Thus, according with this picture, rich clusters are the biggest virialized structures in the universe and they have been assembled in relatively recent epochs. This assembly proceeds through subsequent mergers of smaller subunits and can be directly observed in the local universe. In fact, both X-ray (Forman et al., 2002) and optical (Solanes, Salvador-Solé & Gonzáles-Casado, 1999) observations of nearby rich clusters show clear signatures of recent merger events in a large fraction of objects, suggesting an high merger rate ($\gtrsim 1$ per Hubble time) for clusters.
1.2. Press–Schechter approach: hierarchical structure formation

Figure 1.1: Comoving number density of dark matter halos according to equation 1.53. Different curves refer to different redshifts, as indicated. We consider here a flat, \( \Lambda \)-dominated universe with \( h = 0.7, \Omega_B = 0.045, b = 1.11 \) and \( T_{\text{CMB}} = 2.726 \).

In the next section, following Lacey & Cole (1993), we will use the Press–Schechter mass function derived above to compute the halo merger rate as a function of halo masses and time. A Monte Carlo method for constructing a simulated merger history of a cluster with a given present mass will be described as well.

1.2.2 The cloud in cloud problem

In the derivation of the halo mass function, Press and Schechter introduced in a somewhat arbitrary way a factor of two in order to account for the entire mass of the universe. As said above, this factor can be justified if underdense regions are properly taken into account (Peacock & Heavens, 1990; Bond et al., 1991). This can be done with the help of figure 1.2, where the density perturbation field extrapolated to the present time is shown, smoothed over spheres of comoving radius \( R \). The perturbation field obviously vanishes if smoothed over a sphere with \( R \to \infty \) and develops fluctuations of increasing amplitude when smoothed over progressively
1. Large scale structure formation

Figure 1.2: Value of the density perturbation field extrapolated to the present time and smoothed on a comoving scale $R$. The field has been sampled at a given point for different values of $R$. The two horizontal dotted lines refer to two different thresholds for collapse at different times $\delta_{c_0}(t_1)$ and $\delta_{c_0}(t_2)$.

smaller regions. Decreasing the smoothing radius $R$ corresponds to add a new shell in $k$-space. Since we are dealing with a Gaussian field, $k$ components are independent from each other, and the trajectory in figure 1.2 is simply a random walk. As a consequence, if $\delta = \delta_{c_0}$ for a given smoothing radius, then the trajectory will evolve above or below the threshold with the same probability. So, it is easy to imagine a trajectory crossing the threshold more than once. If this is the case, according to the standard Press–Schechter prescription, a given point could be considered part of a collapsed system while it should be ascribed to a more massive halo. This is the so called “cloud in cloud problem” which can be easily solved considering only the first upcrossing of the trajectory. The mass of the halo is then equal to the mass enclosed by the largest smoothing sphere for which $\delta = \delta_{c_0}$.

More quantitatively, we can proceed as follows: consider a point which has just reached the threshold $\delta_{c_0}$. Its subsequent behaviour will be completely symmetric, that is: the probability to find the point above or below the threshold for any smaller values of $R$ is the same. Moreover, all the points which are above the threshold correspond to trajectories which crossed the threshold at a higher value of $R$. Due to the symmetry of random walk, all these trajectories will have a symmetric counterpart with respect to $\delta_{c_0}$. So, it is possible to estimate the probability that the threshold has never been crossed simply subtracting from the total distribution all the points which have a symmetric counterpart above the threshold. This probability
1.2. Press–Schechter approach: hierarchical structure formation

is usually called *survival probability* and it is given by:

\[
\frac{dP_s}{d\delta} = \frac{1}{\sqrt{2\pi\sigma}} \left[ exp\left( -\frac{\delta^2}{2\sigma^2} \right) - exp\left( -\frac{(\delta - 2\delta_0)^2}{2\sigma^2} \right) \right] \quad (1.54)
\]

while the probability that the threshold has been crossed at least once is:

\[
1 - P_s = 1 - erf\left( \frac{\delta_0}{\sqrt{2\sigma}} \right) \quad (1.55)
\]

This probability differs by a factor of two from equation (1.50), providing a justification for the *ad hoc* correction factor introduced by Press and Schechter in their seminal work.

1.2.3 Building the merger tree of a cluster

The random–walk model, developed in the previous section to solve the cloud in cloud problem, can be extended to derive the conditional multiplicity function. This function denotes the probability that a point enclosed in a halo with mass \( M_1 \) at a given time \( t_1 \), resided at a previous time \( t_2 \) in a system with mass \( M_2 \). To this purpose, we have to consider once again the smoothed field extrapolated to the present time and two thresholds, \( \delta_0(t_1) \) and \( \delta_0(t_2) \). The problem is how to calculate the conditional probability that the second has not been crossed at \( M_2 \) under the condition that the first has been crossed at \( M_1 \). Once evaluated the conditional probability, it is easy to derive the integral distribution function. The procedure is very similar to the derivation of equations (1.54) and (1.55), the only difference being that in this situation the source of the trajectories is shifted to the point \( (R_1, \delta_0(t_1)) \). This lead to the expression:

\[
F(> M_2|M_1) = 1 - erf\left( \frac{\delta_0(t_1) - \delta_0(t_2)}{\sqrt{2(\sigma^2(M_1) - \sigma^2(M_2))}} \right) \quad (1.56)
\]

which is simply equation (1.55) after the replacements \( \sigma^2 \to \sigma^2(M_1) - \sigma^2(M_2) \) and \( \delta_0 \to \delta_0(t_1) - \delta_0(t_2) \).

The differential conditional probability is then:

\[
P(M_2, t_2|M_1, t_1) = \frac{dF(> M_2|M_1)}{dM_1} \quad (1.57)
\]

and the merger rate at which clusters of mass \( M_1 \) merge at a given time \( t_2 \) is written as a function of the final mass \( M_2 \):

\[
\mathcal{R}(M_1, M_2, t_2) = \left( \frac{dP(M_2|M_1)}{dt_1} \right)_{t_1 \to t_2} \quad (1.58)
\]
where \( P(M_2|M_1) \) is given by the conditional probability definition:

\[
P(M_2|M_1) = \frac{P(M_1|M_2)P(M_2)}{P(M_1)} \quad (1.59)
\]

The full expression for the merger rate at a generic time \( t \) is (Lacey & Cole, 1993):

\[
R(M_1, M_2, t) dM_2 = \sqrt{\frac{2}{\pi}} \left| \frac{d\delta_0(t)}{dt} \right| \left| \frac{1}{\sigma^2(M_2)} \right| \frac{d\sigma(M_2)}{dM_2} \left( 1 - \frac{\sigma^2(M_2)}{\sigma^2(M_1)} \right)^{-3/2} \times
\]

\[
\times \exp \left[ -\frac{\delta_0(t)^2}{2} \left( \frac{1}{\sigma^2(M_2)} - \frac{1}{\sigma^2(M_1)} \right) \right] dM_2
\]

(1.60)

We can use this formalism to simulate one of the possible realizations of a cluster merger tree. It is convenient to determine it moving backward in time, starting with a halo with a given present mass. It would be also possible to proceed in the opposite way, starting with a set of subclusters at an early epoch and evolving them till present time. However, we are interested here in rich clusters, which are rare in the universe and, as a consequence, this latter method results inefficient, since it requires to follow the evolution of several sets of subclusters, the majority of which will never result in a massive objects at present time.

The conditional probability \( P(M_1, t_1|M_2, t_2) \) can be written explicitly as follows (Lacey & Cole, 1993):

\[
P(M_1, t_1|M_2, t_2) = \frac{1}{\sqrt{2\pi}} \frac{M_2}{\sigma^2(M_1) - \sigma^2(M_2)} \delta_0(t_1) \frac{\sigma^2(M_1)}{dM_1} \times
\]

\[
\times \exp \left[ -\frac{(\delta_0(t_1) - \delta_0(t_2))^2}{2(\sigma^2(M_1) - \sigma^2(M_2))} \right] dM_1
\]

(1.61)

This expression can be simplified by replacing the mass and time variables with the functions \( S = \sigma^2(M) \) and \( \omega = \delta_0(t) \). Note that \( S \) and \( \omega \) decrease while \( M \) and \( t \) increase. Let \( K(\Delta S, \Delta \omega) d\Delta S \) be the probability of a change \( \Delta S \) in a temporal step \( \Delta \omega \), then equation (1.61) becomes:

\[
K(\Delta S, \Delta \omega) d\Delta S = \frac{1}{\sqrt{2\pi}} \frac{\Delta \omega}{(\Delta S)^{3/2}} \exp \left[ -\frac{(\Delta \omega)^2}{2(\Delta S)} \right] d\Delta S \quad (1.62)
\]

where \( \Delta S = \sigma^2(M_1) - \sigma^2(M_2) \) and \( \Delta \omega = \delta_0(t_1) - \delta_0(t_2) \). A merger tree can be constructed starting with a cluster with present mass \( M_0 \) and going backward in time according to a positive time step \( \Delta \omega \). The increase in \( S \) can be randomly generated sampling the distribution (1.62). The halo mass at time \( t(\omega + \Delta \omega) < t(\omega) \) is simply given by \( M(S + \Delta S) < M(S) \), and the change in mass can be ascribed to one or more merger events. The time step \( \Delta \omega \) also determines the mass resolution:
1.2. Press–Schechter approach: hierarchical structure formation

Figure 1.3: Merger history of a cluster with a present mass of $10^{15} M_\odot$. The mass (y-axis) suffers major jumps in big merger events. Time is on the x-axis. Figure from Gabici & Blasi (2003).

in fact, a single merger of mass $\Delta M << M$ can be resolved if the condition $(\Delta \omega)^2 \lesssim |d \ln \sigma^2 / d \ln M| (\Delta M / M) S$ is satisfied (Lacey & Cole, 1993).

Salvador–Solé, Solanes & Manrique (1998) introduced a new parameter $\Delta_m = [(M_2 - M_1) / M_1]_{\text{crit}}$ defined as a peculiar value of the captured mass that separates the accretion events from merger events. Events in which a cluster of mass $M_1$ captures a dark matter halo with mass smaller than $\Delta_m M_1$ are considered as continuous mass accretion, while events in which the collected mass is larger than $\Delta_m M_1$ are defined as mergers. The value of $\Delta_m M_1$ is somewhat arbitrary, but its physical meaning can be grasped in terms of modification of the potential well of a cluster, following a merger. A major merger is expected to appreciably change the dark matter distribution in the resulting cluster, while only small perturbations are expected in small mergers, which are then interpreted as events more similar to accretion than to real mergers.

Adopting this effective description of the merger and accretion events, it is easy to use the formalism developed by Lacey & Cole (1993) to construct simulated merger
trees for a cluster with a given mass at the present time. Although useful from a computational point of view, this strategy of establishing a boundary between mergers and accretion events does not correspond to any real physical difference between the two types of events; therefore, in the following we will adopt the name "merger" for both regimes, provided that there is no ambiguity or risk of confusion.

In Figure 1.3 we plotted a possible realization of the merger tree for a cluster with present mass equal to $10^{15} M_\odot$ and $\Delta_{\text{m}} = 0.6$. The history has been followed back in time up to redshift $z = 3$. The big jumps in the cluster mass correspond to merger events, while smaller jumps correspond to what Salvador-Solé, Solanes & Manrique (1998) defined as accretion events. During merger events, the trajectory in the $t - M$ diagram splits into branches of progressively smaller masses and this originates the tree.

### 1.3 Secondary infall onto a gravitationally bound object: the Bertschinger approach

In the previous section we introduced the parameter $\Delta_{\text{m}}$ to distinguish between major merger events and continuous accretion. As already stated, even though this distinction may be useful in several contexts, it still remains somewhat arbitrary for our purposes. Moreover, in the older literature, the concept of continuous accretion onto a collapsed object was discussed in detail (Gunn & Gott, 1972; Bertschinger, 1985), but it had a rather different meaning: once a non-linear object has formed, it will continue to attract matter and consequently its mass will grow by accretion of new material. Since mass falls onto an already formed object, this smooth accretion is often referred in the literature as *secondary infall*.

We summarize here the results obtained by Bertschinger (1985) for an accreting halo in an Einstein–de Sitter universe.

Consider a spherical region with a uniform overdensity $\delta_i < < 1$ at a cosmic time $t_i$. The perturbation will grow in time according with linear theory until the turnaround time, which marks the transition to the non-linear regime. At this time, the matter initially inside $R_i$ ceases expanding altogether and starts to collapse. Shells initially located at radii larger than $R_i$ invert their motion at later times. The evolution of the perturbation after the collapse depends on the collisional or collisionless nature of the matter and on boundary conditions, such as the presence of a black hole in the center.

Since we are interested here in clusters of galaxies, the relevant situation involves an $\Omega < < 1$ collisional gas moving in the potential well generated by the dark matter component, for which $\Omega \sim 1$. After the turnaround time, the system is expected to approach a self-similar form and thus to become independent from initial conditions. The reason for expecting self-similarity is that in an Einstein–de Sitter universe there is no preferred time, since the functional form of the scale factor is a power law in
1.3. Secondary infall onto a gravitationally bound object

time. Moreover, density perturbations grow in time according to the scale factor as well, and they become non–linear when the overdensity is a fixed fraction of the background density (see §1.1.3). So, if there are no characteristic length scales in the primordial fluctuations, it is clear that the infall will approach a self–similar form.

We consider first the evolution of a dark matter fluid element which was located at a radius \( r_i > R_i \) at the initial time. In absence of non gravitational forces, each fluid element is subjected to the Newton’s law:

\[
\frac{d^2 r}{dt^2} = -\frac{GM(< r)}{r^2}
\]

whose solution was already derived in §1.1.3. From equations (1.38) and (1.39) it is possible to obtain the expressions for the initial radius:

\[
r_i = A \left( \frac{6t_i}{B} \right)^{2/3}
\]

and for the initial overdensity:

\[
\delta_i(< r) = 3 \left( \frac{6t_i}{B} \right)^{2/3}
\]

The turnaround radius and time follow from equation (1.36) with \( \theta = \pi \) and can be expressed in terms of the initial radius and time:

\[
r_{ta} = \frac{3}{5} r_i
\]

\[
t_{ta} = 6 \left( \frac{20}{3} \delta_i \right)^{-3/2} t_i
\]

Here we consider spherical and uniform initial overdensities, so that \( \delta_i \propto r_i^{-3} \). Combining this scaling with equations (1.66) we get (Bertschinger, 1985):

\[
r_{ta}(t) = r_{ita} \left( \frac{t}{t_{ita}} \right)^{8/9}
\]

where the subscript \( \text{ita} \) refers to the turnaround of the shell that was at \( R_i \) at the initial time \( t_i \). The same scaling is expected for the virial radius, simply defined as half of the turnaround radius.

After the turnaround, the mass shells start to collapse. During this phase, even a small deviation from spherical symmetry is amplified, shell crossing occurs and the system is expected to virialize through violent relaxation (Lyndel–Bell, 1967). The size of the collapsed halo at a given time is equal to the virial radius \( r_v(t_v) = r_{ta}(t_v)/2 \).
and the density within $r_v$ is a fixed multiple of the background density $\rho \propto t_v^2$, so that the density profile approaches the self-similar form:

$$\rho(r) \propto r^{-9/4}$$  \hspace{1cm} (1.69)

Bertschinger (1985) considered also a baryonic gas component with $\Omega_B << 1$ moving in the potential well generated by the dark matter. He presented his results in terms of the dimensionless quantity $\lambda = r/r_{ta}(t)$. Both dark matter and baryonic components were found to have the same density distribution given by equation (1.69). The main difference is the existence of a spherical shock that forms in the baryonic component due to the collision of the infalling gas with the self generated core. The shock is located at a fixed fraction of the turnaround radius, $\lambda_s = 0.347$, and propagates outwards according to the scaling $r_s \propto t^{8/9}$. The total mass inside a given radius can be expressed as a function of the dimensionless mass $m(\lambda)$, tabulated in (Bertschinger, 1985):

$$M_{tot}(r, t) = \frac{4}{3} \pi \rho_b r_{ta}^3 m(\lambda)$$  \hspace{1cm} (1.70)

At the virial radius we have $\lambda = .5$, which correspond to a value for the dimensionless mass equal to $m(0.5) = 4.06$ (Bertschinger, 1985).

As said in §1.1.3, the region inside the turnaround radius is characterized by an overdensity equal to $9\pi^2/16$. This is also the value of $m(1)$, which represents the dimensionless mass inside the turnaround radius. This allows us to write a more practical expression for the turnaround radius at a given epoch (Gunn & Gott, 1972):

$$r_{ta}(t) = \left(\frac{8GM_{ta}t^2}{\pi^2}\right)^{1/3}$$  \hspace{1cm} (1.71)

where $M_{ta}$ is the mass inside $r_{ta}$.

Since we will be interested in the following in rich clusters of galaxies, we give here some numerical examples specific to this situation. Masses of clusters are generally estimated from the distribution of their X-ray emitting gas if this gas can be considered close to hydrostatic equilibrium (Sarazin, 1988), or through observations of gravitationally lensed background galaxies (Refregier, 2003). These observations tell us that the biggest nearby clusters have masses inside the virial radius of the order of $\geq 10^{15} M_\odot$. The total mass inside the turnaround radius is of the same order of magnitude, being only a factor of $m(1)/m(0.5) = 1.37$ greater. So, using equation (1.71) and assuming an Hubble constant equal to $H_0 = 70 \text{km/s/Mpc}$ we can estimate the turnaround radius to be roughly equal to $\sim 7.6 \text{Mpc}$. The virial radius will be half of this value $\sim 3.8 \text{Mpc}$, while the accretion shock will be located $\lambda_s r_{ta} = 2.6 \text{Mpc}$ away from the center.
Chapter 2

Diffusive shock acceleration: linear theory

It is now commonly believed that particle acceleration can take place at collisionless shock surfaces and this process is often invoked as an explanation for the majority of the non-thermal phenomena observed in astrophysical objects. Since several excellent reviews on this topic can be found in the recent literature (Blandford & Eichler, 1987; Drury, 1983; Gaisser, 1990; Jones, 1994), we give here only a brief overview of the most relevant results which will be useful for the understanding of the following developments of this work. In this chapter, we consider only linear shock acceleration theory, neglecting the backreaction of the accelerated particles onto the shock structure. A more satisfactory approach will be described in chapter 6, where the pressure of the cosmic ray component will be included self-consistently in the calculations.

2.1 Stochastic acceleration: second order Fermi mechanism

Before describing in detail how particles can be accelerated at shocks, it is useful to describe briefly another acceleration mechanism, namely, stochastic or second-order Fermi acceleration. The idea is that particles both gain or lose energy in random elementary interactions, but they gain energy in average. If the average energy gain is proportional to the energy of the particle, $\Delta E = \chi E$, after $n$ interactions the particle energy is:

$$E_n = E_0(1 + \chi)^n$$

(2.1)

where $E_0$ is the initial energy of the particles injected in the accelerator. Equation (2.1) can be rewritten to give us the number of interactions needed to reach a fixed
energy $E$:

$$n = \frac{\ln \left( \frac{E}{E_0} \right)}{\ln(1 + \chi)}$$

(2.2)

We also assume that every particle has a finite probability $P_{\text{esc}}$ to escape from the acceleration region at each interaction, so that the probability of remaining in the accelerator after $n$ interactions is $(1 - P_{\text{esc}})^n$. Thus, the fraction of particles with energy greater than $E$ will be proportional to:

$$N(> E) \propto \sum_{m=n}^{\infty} (1 - P_{\text{esc}})^m = \frac{(1 - P_{\text{esc}})^n}{P_{\text{esc}}}$$

(2.3)

Substitution of equation (2.2) into (2.3) gives the power law spectrum:

$$N(> E) \propto \frac{1}{P_{\text{esc}}} \left( \frac{E}{E_0} \right)^{-\delta}$$

(2.4)

where:

$$\delta = -\frac{\ln(1 - P_{\text{esc}})}{\ln(1 + \chi)}$$

(2.5)

If both $P_{\text{esc}}$ and $\chi$ are much lower than unity, then $\delta \sim P_{\text{esc}}/\chi$. Note that the power law shape of the particle spectrum is a consequence of the fact that we implicitly assumed $P_{\text{esc}}$ to be independent from particle energy.

In a seminal paper, Fermi (1949) proposed a stochastic acceleration mechanism to explain the observed spectrum of galactic cosmic rays. He modelled the turbulent interstellar medium as an ensemble of moving magnetized clouds and considered the interactions between a particle and a moving cloud. These interactions are collisionless, in the sense that particles simply diffuse in the cloud magnetic field without colliding with other particles. In the cloud rest frame no work is done on the particle, which is moving in a static magnetic field, so that the interaction can be simply treated as an elastic scattering.

Consider now a single interaction between a particle with energy $E_1$ and a cloud with Lorentz factor $\gamma$ and velocity $\beta$. For simplicity, the particle is assumed to be already sufficiently relativistic, so that we can assume in the following $E \sim pc$. In the cloud frame the particle has a total energy:

$$E'_1 = \gamma E_1 (1 - \beta \cos \theta_1)$$

(2.6)

where $\theta_1$ is the angle between the cloud and the particle velocities and primes denote quantities measured in the cloud frame. After the interaction the particle maintains the same energy, $E'_2 = E'_1$, and it is scattered in a direction defined by an angle $\theta'_2$. We can now transform this energy back to the lab frame obtaining:

$$E_2 = \gamma E'_2 (1 + \beta \cos \theta'_2)$$

(2.7)
The change in energy after a generic interaction can be written combining equations (2.6) and (2.7):

\[
\frac{\Delta E}{E_1} = \frac{1 - \beta \cos \theta_1 + \beta \cos \theta_2 + \beta^2 \cos \theta_1 \theta_2}{1 - \beta^2} - 1
\]  

(2.8)

In order to calculate the average energy gain per interaction, we have to average this expression over \( \cos \theta_1 \) and \( \cos \theta_2 \). Particles are scattered isotropically in the cloud frame, so that \( \langle \cos \theta_2 \rangle = 0 \). On the other hand, the probability of having a collision between a cloud and a particle is proportional to their relative velocity:

\[
\frac{dn}{d \cos \theta_1} = \frac{1 - \beta \cos \theta_1}{2}
\]  

(2.9)

and the average over this probability distribution is \( \langle \cos \theta_1 \rangle = -\beta/3 \). Thus, the average energy gain per interaction is:

\[
\frac{\Delta E}{E} = \frac{1 + \frac{1}{3} \beta^2}{1 - \beta^2} \approx \frac{4}{3} \beta^2
\]  

(2.10)

which is proportional to the velocity squared, from which the name second order Fermi acceleration. The result derived above holds as long as clouds are not relativistic, namely, \( \beta \ll 1 \). In this situation, particles gain on average a very small amount of energy per interaction, and a long time is needed to accelerate them up to very high energies.

### 2.2 Acceleration of particles at shock fronts: first order Fermi mechanism

The original idea by Fermi can be opportunely modified and applied to the case of a shock wave, resulting in a more efficient acceleration mechanism (Bell, 1978a; Blandford & Ostriker, 1978).

#### 2.2.1 The Bell approach

In order to study the acceleration of particles at collisionless shocks, Bell (1978a) proposed an approach in which the behaviour of each particle is followed. He considered a parallel, plane and non relativistic shock wave and discussed only those particles whose energy is high enough for their Larmor radii to be much greater than the shock thickness, usually assumed to be equal to the Larmor radius of thermal protons. These particles can cross freely from a side to another of the shock surface and their velocity distribution is assumed to be isotropic in both upstream
and downstream regions. Such an isotropization can be physically motivated by the presence of irregularities in the magnetic field structure that can efficiently scatter particles on both sides of the shock.

Consider the fluid flow in the rest frame in which the shock is at rest and let $\beta_1 = u_1/c$ be the supersonic velocity of the fluid in the upstream region and $\beta_2 = u_2/c < \beta_1$ the downstream velocity. Now move to the upstream rest frame. In this frame, the velocity distribution of high energy particles is isotropic and a particle crossing the shock will encounter gas moving towards the shock with a velocity $\beta_r = \beta_1 - \beta_2$. The same situation occurs in the downstream rest frame: the particle velocity distribution is isotropic and a particle crossing the shock front from downstream to upstream will encounter gas having the velocity $\beta_r$. The similarity with the second order Fermi process resides in the fact that each back and forth shock crossing is similar to a particle–cloud interaction. However, this time the distribution of $\cos \theta_1$ and $\cos \theta'_r$ is simply the projection of an isotropic distribution onto the shock plane:

$$\frac{dn}{d \cos \theta} = 2 \cos \theta$$

(2.11)

This distribution has to be averaged over the range $0 \leq \cos \theta \leq 1$ and $-1 \leq \cos \theta \leq 0$ for particles crossing the shock from downstream to upstream and vice versa. It is easy to show that in this case $\langle \cos \theta'_r \rangle = 2/3$ and $\langle \cos \theta_1 \rangle = -2/3$ and equation (2.8) becomes:

$$\frac{\Delta E}{E} = \frac{1 + \frac{2}{3} \beta_r + \frac{2}{3} \beta_r^2}{1 - \beta_r^2} - 1 \sim \frac{4}{3} \beta_r = \frac{4 u_1 - u_2}{3 c}$$

(2.12)

This process is more efficient than the original Fermi mechanism because in this case the average energy gain per interaction is proportional to the relative velocity of the plasma flow. This is a consequence of the fact that during collisions with clouds, particles can both gain and lose energy, while when they cross back and forth a shock they always gain energy. This mechanism is often referred as first order Fermi acceleration, because the average energy increment in one interaction scales linearly with velocity.

The escape probability can be estimated in the following way: the rate at which particles are crossing and recrossing the shock is:

$$\int_0^1 d \cos \theta \int_0^{2\pi} d\phi \frac{\rho_{cr}}{4\pi} \cos \theta = \frac{\rho_{cr}}{4}$$

(2.13)

where $\rho_{cr}$ is the density of high energy particles at the shock. Due to advection, of the $\rho_{cr}/4$ particles that cross the shock from upstream to downstream per unit time, $\rho_{cr} u_2$ escape, while the rest diffuse back upstream. The ratio between these two rates is the escape probability, $P_{esc} = 4u_2/c$, which also determines the slope of the differential energy spectrum of accelerated particles $n(E) \propto E^{-\alpha}$ (see equation 2.5 and following discussion):

$$\alpha = \delta + 1 = \frac{P_{esc}}{\left( \frac{\Delta E}{E} \right)} = \frac{3}{u_2} - 1 + 1 = \frac{r + 2}{r - 1}$$

(2.14)
2.2. Acceleration of particles at shock fronts: first order Fermi mechanism

With the compression factor \( r = \frac{u_1}{u_2} \) related to the shock Mach number through the well known relation (Landau & Lifshitz, 1982):

\[
r = \frac{u_1}{u_2} = \frac{\rho_2}{\rho_1} = \frac{(\gamma + 1)M^2}{(\gamma - 1)M^2 + 2}
\]  

(2.15)

where \( \gamma \) is the adiabatic index, equal to 5/3 for a perfect gas. Equation (2.14) can now be written in a more convenient form:

\[
\alpha = 2\frac{\mathcal{M}^2 + 1}{\mathcal{M}^2 - 1}
\]  

(2.16)

from which it is clear that for strong shocks the slope approaches the asymptotic value \( \alpha = 2 \).

2.2.2 The particle transport equation

The same results obtained above can also be derived adopting a different approach, based on the solution of the transport equation for particles at a shock (Blandford & Ostriker, 1978). For a plane shock wave propagating in the negative \( x \) direction this equation can be written in the following form (Lagage & Cesarsky, 1983):

\[
\frac{\partial f}{\partial t} + \mathbf{u} \cdot \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} \left( D_i \frac{\partial f}{\partial x} \right) - \frac{1}{3} p \frac{\partial f}{\partial p} \Delta u \delta(x) + Q_0 \delta(x)
\]  

(2.17)

where \( p \) is the particle momentum, \( 4\pi \int p^2 f(x, p) dp \) represents the cosmic ray number density, \( i \) is 1 (2) when \( x < 0 \) (\( x > 0 \)), \( D_i \) is the diffusion coefficient and \( Q_0(p) \) represents the injection of particles in the accelerator, assumed to occur only at the shock position. The integration of equation (2.17) around the shock between \( x = 0^- \) and \( x = 0^+ \) yields:

\[
D_1 \left[ \frac{\partial f}{\partial x} \right]_1 = -\frac{1}{3} p \frac{\partial f}{\partial p} \Delta u + Q_0
\]  

(2.18)

where \( f_0(p) = f(0, p) \), and it was made use of the homogeneity condition for the distribution function downstream: \( [D\partial f/\partial x]_2 = 0 \). In order to take into account the boundary condition at upstream infinity, \( [\partial f/\partial x]_{-\infty} = 0 \), it is convenient to integrate equation (2.17) between \( x = -\infty \) and \( x = 0 \) obtaining:

\[
D_1 \left[ \frac{\partial f}{\partial x} \right]_1 = u_1 (f_0 - f_{-\infty})
\]  

(2.19)

The solution at \( x = 0 \) can be easily calculated combining together equations (2.18) and (2.19):

\[
f_0(p) = q p^{-q} \int_0^p dp' p'^{q-1} \left[ \frac{Q_0(p')}{u_1} + f_{-\infty}(p') \right]
\]  

(2.20)
2. Diffusive shock acceleration: linear theory

with:

\[ q = \frac{3u_1}{\Delta u} = \frac{3r}{r - 1} \]  

(2.21)

Consider first the situation in which there are not seed particles at upstream infinity, so that \( f_{-\infty} = 0 \). For a monochromatic injection \( Q_0(p) \propto \delta(p - p_0) \) the particle distribution function is simply a power law in momentum:

\[ p^2 f_0(p) \propto p^{-q} M_{-\infty}^{-3} p^{-2} \]  

(2.22)

Note that this is exactly the same result obtained above adopting the Bell approach.

2.2.3 Reacceleration of seed particles

The situation is different if relativistic particles are already present in the unperturbed upstream medium. These seed particles are reaccelerated by the shock wave (Bell, 1978b; Lagage & Cesarsky, 1983) and their original distribution function \( f_{-\infty}(p) \) changes according to:

\[ f(p) = q p^{-q} \int_0^p dp' p'^{q-1} f_{-\infty}(p') \]  

(2.23)

It is possible to investigate the effects of shock reacceleration by considering particles originally accelerated at a shock front which have subsequently passed through another shock. If the shocks are characterized by two compression factors \( r_1 \) and \( r_2 \), then they can accelerate particle with spectra which are power lows in momentum with slopes \( q_1 \) and \( q_2 \) given by equation (2.21). Since we are interested in reacceleration, we neglect here those particles injected at the second shock. It is easy to show that, if \( q_1 \neq q_2 \), the final spectrum of the particles is:

\[ f(p) \propto \left( \frac{p}{p_0} \right)^{-\alpha_2} \left[ \left( \frac{p}{p_0} \right)^{\alpha_2-\alpha_1} - 1 \right] \]  

(2.24)

where \( p_0 \) is the injection momentum. The distribution function \( f(p) \) vanishes at \( p = p_0 \), while at high energies it has an asymptotic power law shape, with slope equal to the one corresponding to the stronger shock. The total energy in accelerated particles also increases at each reacceleration step. This effect is greater for strong shocks, that can increase the total energy of roughly one order of magnitude (Bell, 1978b), while it is much less important for weak shocks.
Chapter 3

Extended Press–Schechter formalism and particle acceleration: a novel approach

Diffuse synchrotron radio emission is observed from a growing number of rich clusters of galaxies, revealing the presence of ultra–relativistic electrons in the intracluster medium. Understanding the origin of these particles is still an open issue. However, both observations and theoretical considerations seem to suggest a close connection between the non–thermal activity observed from the intracluster medium and major merger events that lead to the formation of rich clusters.

The huge amount of gravitational energy released during mergers is converted into thermal energy through shock waves, which are the main responsible for the heating of the intracluster medium. A small fraction of the particles crossing these shocks may be energized via first order Fermi mechanism up to ultra–relativistic energies and, for this reason, merger shocks are considered amongst the most important sources of cosmic rays in clusters.

In this chapter we present a new, semi–analytical approach to study the acceleration of particles at shocks that form during the formation of large scale structures in the universe. Our method combines the effective description of hierarchical clustering made by Press and Schechter, with the linear theory of diffusive shock acceleration. The link between these two pieces of physics is provided by a recipe that allows us to estimate in a self–consistent way the Mach number of the merger related shocks. Our goal is to evaluate the relevance of merger shocks in the acceleration of the particles responsible for the non–thermal emission observed from several rich clusters of galaxies.

We anticipate here that the most important result is the fact that shocks related to major mergers are found to be only mildly supersonic, having Mach numbers typically equal to $\sim 1.5$. As a consequence, the spectra of the accelerated particles are very steep, and consequently unlikely to result in appreciable non–thermal emis-

sion. Only shocks that form during minor mergers are strong enough to accelerate particles with flat spectra.

3.1 Non–thermal emission from clusters of galaxies: observations and models

Rich clusters of galaxies are strong X–ray sources with luminosities typically in the range $L_X \sim 10^{43} - 10^{45}\,\text{erg/s}$. The X–ray emission is well explained as thermal Bremsstrahlung radiation of the very hot ($T \sim 10^8\,\text{K}$), low–density ($n_e \sim 10^{-3}\,\text{cm}^{-3}$), highly ionized intracluster electron gas.

It is now well established that magnetic field and relativistic particles are present in the intracluster medium. The main evidence for this fact comes from radio observations, that reveal the presence of large–scale diffuse and unpolarized radio halos in $\sim 30\%$ of the X–ray brightest clusters (Feretti, 2003). This emission is not associated with single cluster galaxies and it is interpreted as synchrotron radiation from $\sim \text{GeV}$ electrons gyrating in the intracluster magnetic field. The observed radio spectra are typically steep, with spectral index $\alpha \gtrsim 1$. Moreover, in a few cases, the evidence for a radial spectral steepening has been claimed (Giovannini et al., 1993; Feretti et al., 2004).

The first halo was detected in the Coma cluster (Willson, 1970) and is considered now the prototype of this class of radio sources. When observed at $1.4\,\text{GHz}$, it is $\gtrsim 1.5\,\text{Mpc}$ wide and its total radio luminosity is $L_{\nu} \sim 10^{40}\,\text{erg/s}$ (Deiss et al., 1997). By assuming energy equipartition between magnetic field and relativistic particles, a magnetic field value equal to $0.5\,\mu\text{G}$ can be derived (Giovannini et al., 1993).

Non–thermal emission is also observed from a number of rich clusters of galaxies as an excess over the thermal emission in the hard–X (Fusco–Femiano et al., 2004) and possibly in the soft–X-ray/UV (Bowyer, 2003) bands. For the Coma cluster, the detected fluxes in these bands correspond to source luminosities equal to $\sim 10^{43}\,\text{erg/s}$ (Fusco–Femiano et al., 2004) and $\sim 10^{42}\,\text{erg/s}$ (Bowyer, Berghöfer & Korpela, 1999) respectively. While it is known that radiative processes related to relativistic electrons are responsible for this radiation, the sites and nature of the acceleration of such particles are not known.

Several models have been proposed in order to explain radio observations, based on acceleration of electrons at shock surfacers that naturally form during merger events (Roettiger, Burns & Stone, 1999; Sarazin, 1999; Takizawa & Naito, 2000; Fujita & Sarazin, 2001) or models in which electrons are secondary products of hadronic interactions (Dennison, 1980; Colafrancesco & Blasi, 1998; Blasi & Colafrancesco, 1999; Dolag & Ensslin, 2000) and finally models in which electrons are continuously reenergized by turbulence (Schlickeiser et al., 1987; Brunetti et al., 2001a; Ohno, Takizawa & Shibata, 2002).

Hard X–ray and EUV radiation in excess of the thermal emission may be gen-
3.1. Non–thermal emission from clusters of galaxies: observations and models

erated by inverse Compton scattering of relativistic electrons off the photons of the cosmic microwave background radiation. When applied to the Coma cluster, these models require values of the volume averaged magnetic field of order $\sim 0.1 \mu G$ (Fusco–Femiano et al., 1999), which are much smaller than those measured through Faraday rotation, which are typically of several $\mu G$ (Clarke, Kronberg & Böringer, 2001). This conclusion can be possibly avoided only by constructing models in which a cutoff in the electron spectrum is tuned up in order to reduce the corresponding synchrotron emission. In these cases the magnetic field can be as high as $0.3 - 0.4 \mu G$ (Brunetti et al., 2001a). In the case of a secondary origin for the radiating electrons, the small magnetic fields imply a large cosmic ray content in the intracluster gas. In the case of the Coma cluster, the gamma ray upper limit found by Sreekumar et al. (1996) is exceeded by the gamma ray flux from the decay of neutral pions, as shown by Blasi & Colafrancesco (1999).

The hard X–ray excess might also be the result of Bremsstrahlung emission from a population of thermal electrons whose distribution function is slightly different from a Maxwell–Boltzmann distribution (Ensslin, Lieu & Biermann, 1999; Blasi, 2000; Dogiel, 2000; Sarazin & Kempner, 2000). A tail might in fact be induced in the Maxwell–Boltzmann distribution by the presence of MHD waves that resonate with part of the thermal electrons (Blasi, 2000). This model requires an energy input comparable with the energy budget of a cluster merger, and implies a substantial heating of the intracluster gas (this was shown by Blasi (2000) by solving the full Fokker–Planck equations, including Coulomb scattering). If the process lasts for too long a time (larger than a few hundred million years) the cluster is heated to a temperature well in excess of the observed ones, and the model fails. In this case the arguments presented by Petrosian (2001) apply.

The presence of tails in the Maxwell–Boltzmann electron distribution can be tested through observations of the Sunyaev–Zeldovich (SZ) effect, as proposed by Blasi, Olinto & Stebbins (2000) (see Ensslin & Kaiser (2000) for a general discussion of the SZ effect including nonthermal effects). Clearly, by simply observing radio radiation and hard X–ray radiation from clusters, it is extremely difficult, if not impossible to discriminate among classes of models. The study of the SZ effect allows one to partly break the degeneracy. An even more powerful tool is represented by gamma ray astronomy. Some of the models in the literature predict gamma ray emission to some extent, while others (this is the case of nonthermal tails in the Maxwell–Boltzmann distribution) do not make precise predictions about the gamma ray emission, and in fact do not require it. Clusters of galaxies are among the targets for observations by the GLAST satellite and the new generation ground based Cherenkov arrays. These observations will open a new window onto the non–thermal processes occurring in the intracluster gas, and will allow one to understand the origin of the observed radiation at lower frequency.
3.1.1 Cosmic ray confinement in the intracluster medium

It was first realized by Berezinsky, Blasi & Ptuskin (1997) and Völk, Aharonian & Breitschwerdt (1996) that cosmic rays are diffusively confined in the cluster volume for cosmological times. This finding is of great importance for a better understanding of the non-thermal activity in clusters of galaxies and deserves a quantitative description.

The propagation of particles in the intracluster medium can be described by means of an energy dependent diffusion coefficient \( D(E) \). The minimum possible diffusion coefficient along the field lines can be estimated assuming that particles are scattered by magnetic field irregularities with a mean free path comparable with their Larmor radius \( R_L \). This situation is known as Bohm diffusion, and the related diffusion coefficient is given by:

\[
D_B(E) = \frac{1}{3} R_L(E) c \propto E
\]  

where the particle velocity is set equal to \( c \), since we are interested here in relativistic particles.

More generally, the functional form and normalization of \( D(E) \) can be derived from the magnetic field spatial structure according to:

\[
D(E) = \frac{1}{3} R_L(E) c \frac{B^2}{8\pi} \int_{1/R_L}^{\infty} dk P(k)
\]  

where \( P(k) \) is the spectrum of the magnetic field irregularities. If a Kolmogorov spectrum is assumed:

\[
P(k) = P_0 k^{-5/3}
\]  

the diffusion coefficient has a weaker dependence on energy: \( D(E) \propto E^{1/3} \). The normalization constant \( P_0 \) can be obtained from the condition:

\[
\int_{1/l_{\text{max}}}^{\infty} dk P(k) \sim \frac{B^2}{8\pi}
\]  

where \( l_{\text{max}} \) is the size of the largest eddy in the magnetic field. For typical intracluster medium conditions the diffusion coefficient can be written in the form (Blasi, 2001):

\[
D(E) = \begin{cases} 
3.3 \times 10^{22} \left( \frac{B}{1\mu G} \right)^{-1} E(\text{GeV}) cm^2/s & \text{Bohm} \\
2.3 \times 10^{29} \left( \frac{B}{1\mu G} \right)^{1/3} \left( \frac{l_{\text{max}}}{20 \text{ kpc}} \right)^{2/3} E(\text{GeV})^{1/3} cm^2/s & \text{Kolmogorov}
\end{cases}
\]  

Following Berezinsky, Blasi & Ptuskin (1997), we can use these expressions to estimate the escape time for a relativistic particles with energy \( E \) from a Coma-like cluster with mass \( \sim 10^{15} M_\odot \) and virial radius \( R_{cl} \sim 3 \text{ Mpc} \):

\[
\tau_{\text{esc}} \sim \frac{R_{cl}^2}{6D(E)}
\]
3.1. Non-thermal emission from clusters of galaxies: observations and models

\[
E_{\text{max}}^{\text{conf}} \approx \begin{cases} 
1.4 \times 10^{10} \left( \frac{R_{\text{cl}}}{3 \text{ Mpc}} \right)^2 \left( \frac{B}{1 \mu\text{G}} \right) E(\text{GeV})^{-1} \text{Gyr} & \text{Bohm} \\
2.0 \times 10^3 \left( \frac{R_{\text{cl}}}{3 \text{ Mpc}} \right)^2 \left( \frac{B}{1 \mu\text{G}} \right)^{-1/3} \left( \frac{\text{Bohm}}{20\text{kpc}} \right)^{-2/3} E(\text{GeV})^{-1/3} \text{Gyr} & \text{Kolmogorov}
\end{cases}
\] (3.6)

All the particles having escape times longer than the cluster life time remain trapped within the cluster volume.

The cluster life time can be assumed to be roughly equal to the Hubble time \( t_H \sim 13\text{Gyr} \), so that the maximum particle energy for which the confinement is effective is:

\[
E_{\text{max}}^{\text{conf}} \approx \begin{cases} 
10^9 \left( \frac{R_{\text{cl}}}{3 \text{ Mpc}} \right)^2 \left( \frac{B}{1 \mu\text{G}} \right) \text{GeV} & \text{Bohm} \\
4 \times 10^6 \left( \frac{R_{\text{cl}}}{3 \text{ Mpc}} \right)^6 \left( \frac{B}{1 \mu\text{G}} \right)^{-1} \left( \frac{\text{Bohm}}{20\text{kpc}} \right)^{-2} \text{GeV} & \text{Kolmogorov}
\end{cases}
\] (3.7)

It is clear that the determination of this maximum energy strongly depends on the diffusion coefficient, which is rather uncertain. However, equation (3.7) tells us that the bulk of the cosmic rays remains confined in the intracluster medium for cosmological times, regardless of the assumptions made on the diffusion coefficient.

This argument is of special importance for particles which are not affected by energy losses. For cosmic ray protons propagating in the intracluster medium, the main channel of energy losses is provided by inelastic scattering with thermal protons. The inclusive cross section for the reaction considered is \( \sigma_{pp} = 3.2 \times 10^{-26}\text{cm}^2 \) and the energy loss time for typical intracluster medium conditions is:

\[
\tau_{\text{loss}}^p = \frac{1}{n_{\text{gas}} \sigma_{pp} c} \approx 34 \left( \frac{n_{\text{gas}}}{10^{-3}} \right)^{-1} \text{Gyr}
\] (3.8)

which is longer than the Hubble time, leading to the important conclusion that protons are unaffected by energy losses. It follows that all the relativistic protons injected in the intracluster medium during the whole cluster lifetime remain confined and accumulate there. Thus, the cosmic ray energy density in a cluster is an increasing function of time (Berezinsky, Blasi & Ptuskin , 1997; Völk, Aharonian & Breitschwerdt , 1996)

This fact has important consequences for the expected non-thermal emission from clusters. In fact, due to confinement, the probability of having inelastic proton–proton scattering is enhanced. The process is efficient enough to the continuous production of neutral and charged pions, which in turn decay into gamma rays, electrons and positrons (secondary electrons) through the decay chain:

\[pp \rightarrow \pi^0 + \pi^+ + \pi^- + \text{anything}\]

\[\pi^0 \rightarrow \gamma\gamma\]

\[\pi^\pm \rightarrow \mu^\pm \nu_\mu \quad \mu^\pm \rightarrow e^\pm \nu_\mu \nu_e\]

The gamma ray production rate due to neutral pion decay and the synchrotron emission from secondary electrons are amplified as well.
The situation is very different for relativistic electrons, that lose their energy via inverse Compton scattering on short time scales (Rybiki & Lightman, 1979):

\[ \tau_{\text{loss}}^e = \frac{E}{4 \sigma_T c \left( \frac{E}{m_e c^2} \right)^2 U_{\text{cmb}}} \sim 1E(\text{GeV})^{-1} \text{Gyr} \quad (3.9) \]

where \( U_{\text{cmb}} \) is the energy density in the cosmic microwave background and the other symbols have their usual meanings. If the intracluster magnetic field is greater than \( \sim 3\mu G \), synchrotron losses become dominant and the energy loss time is even shorter. GeV electrons gyrating in a \( \mu G \) magnetic field emit synchrotron radiation in the radio band and, during one energy loss time, they can diffuse over a distance:

\[ l_{\text{diff}}(E) = \sqrt{6D(E)\tau_{\text{loss}}^e} \quad (3.10) \]

which ranges from tens of parsecs to tens of kiloparsecs, depending on the assumptions made on the diffusion coefficient. For these reason primary electrons cannot diffuse away from the sites where they are produced and are unlikely to be responsible of the observed \( \sim \text{Mpc} \) size diffuse radio emission in clusters.

### 3.1.2 Non–thermal emission and cluster mergers

As stressed above, there is at present no compelling evidence in favor of any of the proposed acceleration sites for the non–thermal particles in clusters. Nevertheless, energetic events in the history of a cluster represent good candidates, and cluster mergers, that build up the cluster itself hierarchically, fit the description. In fact, during mergers, clusters collide at velocities \( v \lesssim 2000 \text{km/s} \), releasing a huge amount of gravitational energy roughly equal to:

\[ E_m \sim \frac{GM_1M_2}{d} = 1.4 \times 10^{64} \left( \frac{M}{5 \times 10^{14} M_\odot} \right)^2 \left( \frac{d}{1.5 \text{Mpc}} \right)^{-1} \text{erg} \quad (3.11) \]

where a major merger between two rich clusters with typical mass \( 5 \times 10^{14} M_\odot \) is considered. The duration of the merger event is \( t_m \sim d/v \sim 10^9 \text{yr} \), so that the rate of energy release is \( L_m = E_m/t_m \sim 2 \times 10^{47} \text{erg/s} \). This means that only a very small fraction of the total energy involved in a merger event needs to be converted into relativistic particles in order to explain observations. Moreover, a possible observational evidence for a correlation between major merger and radio halos has recently been found by Buote (2001). In particular, the correlation exists between the radio emission at 1.4GHz and the degree of departure from virialization in the shape of clusters, interpreted as a consequence of a recent or ongoing merger that visibly changed the dark matter distribution in the cluster core.

It seems therefore reasonable to associate the existence of non–thermal particles to some process occurring during these mergers. This argument, first proposed by
3.2. Merger trees and strength of the shocks in the intracluster medium

Harris et al. (1980), is made stronger by the fact that mergers are also thought to be responsible for the heating of the intracluster gas. In fact, shocks that are formed in the baryon components of the merging clusters are able to convert part of the gravitational energy of the system into thermal energy of the gas, as shown by direct observations (Markevitch, Vikhlinin & Forman, 2003). It has been claimed that if these shocks are strong enough, they can efficiently accelerate particles by first order Fermi acceleration (Fujita & Sarazin, 2001; Miniati et al., 2001a,b; Blasi, 2000).

The consequences of these shocks on the non-thermal content of clusters of galaxies may be dramatic, and deserve to be considered in detail. Both electrons and protons (or nuclei) are accelerated at the shock surfaces during mergers, but the dynamics of these two components is extremely different: high energy electrons have a radiative lifetime much shorter than the age of the cluster, so that they rapidly radiate most of their energy away. On the other hand, protons lose only a small fraction of their energy during the lifetime of the cluster, and their diffusion time out of the cluster are even larger, so that they are stored in clusters for cosmological times (Berezinsky, Blasi & Ptuskin, 1997; Völk, Aharonian & Breitschwerdt, 1996). In other words, while for high energy electrons only the recent merger events are important to generate nonthermal radiation that we can observe, in order to determine the proton population of a cluster (that can generate secondary electrons) we need to take into account the all history of the cluster.

3.2 Merger trees and strength of the shocks in the intracluster medium

As seen in chapter 1, the hierarchical growth of clusters can be described semi-analytically in the framework of the so-called extended Press–Schechter formalism (Press & Schechter, 1974; Lacey & Cole, 1993). Merger trees for clusters with a given present mass can be built by means of Monte Carlo simulations. This formalism was extensively used in the past to investigate the intracluster medium heating and other thermal properties of clusters (Cavaliere, Menci & Tozzi, 1999; Randall, Sarazin & Ricker, 2002). We suggest here that this approach can be effectively generalized in order to describe the acceleration of particles at merger shocks and the consequent implications for cluster non-thermal emission. The novelty in our proposal resides in the combination of the Press–Schechter formalism with a recipe which allows us to estimate in a self-consistent manner the Mach number of each merger related shock. This new approach can be used as a powerful tool to investigate the non-thermal aspect of structure formation and to make predictions about the number of objects showing non-thermal activity at a given frequency, from radio waves to gamma rays.

3.2.1 Shocks during cluster mergers

During the merger of two clusters of galaxies, the baryonic component, feeling the gravitational potential created mainly by the dark matter component, is forced to move supersonically and shock waves are generated in the intracluster medium.

In this section we describe in more detail the physical properties of such shocks, with special attention for their Mach numbers and compression factors. To this purpose we use an approach introduced in its original version by Takizawa (1999). We assume to have two clusters, as completely virialized structures, at temperatures $T_1$ and $T_2$, and with masses $M_1$ and $M_2$ (here the masses are the total masses, dominated by the dark matter components). The virial radius of each cluster can be written as follows

$$r_{\text{vir},i} = \left( \frac{3M_i}{4\pi \Delta c_\Omega m \rho_{\text{cr}} (1 + z_{f,i})^2} \right)^{\frac{1}{3}} = \left( \frac{GM_i}{100\Omega_m H_0^2 (1 + z_{f,i})^2} \right)^{\frac{1}{3}},$$

(3.12)

where $i = 1, 2$, $\rho_{\text{cr}} = 1.88 \times 10^{-29} h^2$ g cm$^{-3}$ is the current value of the critical density of the universe, $z_{f,i}$ is the redshift of formation of the $i$–th cluster, $\Delta c \sim 200$ is the density contrast for the formation of the cluster and $\Omega_m$ is the matter density fraction. In the right hand side of the equation we used the fact that $\rho_{\text{cr}} = 3H_0^2/8\pi G$, where $H_0$ is the Hubble constant. The formation redshift $z_f$ is on average a decreasing function of the mass, meaning that smaller clusters are formed at larger redshifts, consistently with the hierarchical scenario of structure formation. There are intrinsic fluctuations in the value of $z_f$ from cluster to cluster at fixed mass, due to the stochastic nature of the merger tree.

The relative velocity of the two merging clusters, $V_r$, can be easily calculated from energy conservation:

$$-\frac{GM_1 M_2}{r_{\text{vir},1} + r_{\text{vir},2}} + \frac{1}{2} M_r V_r^2 = -\frac{GM_1 M_2}{2R_{12}},$$

(3.13)

where $M_r = M_1 M_2 / (M_1 + M_2)$ is the reduced mass and $R_{12}$ is the turnaround radius of the system, where the two subclusters are supposed to have zero relative velocity. In fact the final value of the relative velocity at the merger is quite insensitive to the exact initial condition of the two subclusters. In an Einstein–De Sitter cosmology this spatial scale equals twice the virial radius of the system. Therefore, using eq. (3.12), we get:

$$R_{12} = 2 \left( \frac{M_1 + M_2}{M_1} \right)^{1/3} \frac{1 + z_{f,1}}{1 + z_f} r_{\text{vir},1}.$$

(3.14)

where $z_f$ is the formation redshift of the cluster with mass $M_1 + M_2$. This expression remains valid in approximate way also for other cosmological models (Lahav et al., 1991). The sound speed of the $i$–th cluster is given by

$$c_{s,i}^2 = \frac{\gamma_g (\gamma_g - 1) GM_i}{2r_{\text{vir},i}},$$

(3.15)
3.2. Merger trees and strength of the shocks in the intracluster medium

where we used the virial theorem to relate the gas temperature to the mass and virial radius of the cluster. The adiabatic index of the gas is \( \gamma_g = 5/3 \). The Mach number of each cluster can be written as follows (Gabici & Blasi, 2003):

\[
\mathcal{M}^2 = \frac{4(1 + \eta)}{\gamma(\gamma - 1)} \left[ \frac{1}{1 + \frac{1+z_{f1}}{1+z_{f2}} \eta^{1/3}} - \frac{1}{4\frac{1+z_{f1}}{1+z_{f2}} (1 + \eta)^{1/3}} \right]
\]

\[
\mathcal{M}^2 = \eta^{-2/3} \frac{1+ z_{f1}}{1+ z_{f2}} \mathcal{M}^2_1,
\]

where \( \eta = M_2/M_1 < 1 \). The procedure illustrated above can be applied to a generic couple of merging clusters, and in particular it can be applied to a generic merger event in the history of a cluster with fixed mass at the present time. The merger history (and indeed many realizations of the history) for a cluster can be simulated as discussed in §1.2.3. In particular, for a cluster with mass \( M_0 \) at the present time we simulate 500 realizations of the merger tree and calculate the Mach numbers associated with the merger events. A value \( \Delta_m = 0.05 \) is assumed. Note that this value is much lower than in Fujita & Sarazin (2001). This simply implies that we follow the histories of very small halos of dark matter, rather than the big ones only. The results of our calculations of the Mach numbers are plotted in figure 3.1.

The striking feature of this plot is that for major mergers, involving clusters with comparable masses (\( \eta \sim 1 \)), the Mach numbers of the shocks are of order unity. In other words the shocks are only moderately supersonic. In order to achieve Mach numbers of order 3 – 4 it is needed to consider mergers between clusters with very different masses (\( \eta \sim 0.05 \)), which, in the language of Salvador-Solé, Solanes & Manrique (1998) and Fujita & Sarazin (2001) would not be considered as mergers but rather as continuous accretion. These events are the only ones that produce strong shocks, and this is of crucial importance for the acceleration of suprathermal particles, as discussed below. In figure 3.2 we plotted histograms representing the frequency of shocks with given Mach number. It is easy to see that the greatest part of the merger shocks are weak, with a peak in Mach number at about 1.4. This result is in some contradiction with the Mach number distribution found by Miniati et al. (2000), which is a bimodal distribution with one peak at Mach numbers of \( \sim 1000 \), and one at Mach numbers of \( \sim 5 \). While the former peak cannot be obtained within our approach, because it is caused by shocks in the cold outskirts of clusters, not directly related to merger events, the latter shocks should be the same as those described in our paper. In fact, they are located, according to Miniati et al. (2000), within \( 0.5h^{-1} \) Mpc from the cluster center, namely in the virialized region, so that the arguments presented here do apply. It is worth noticing in passing that the simulations of Miniati et al. (2000) have an intrinsic cutoff at Mach numbers smaller than \( \sqrt{3} \), so that the peak we find at 1.5 is out of the available range. It is also worth noticing that the spatial resolution achieved in these simulations is of \( 0.315h^{-1} \) Mpc, very close to the size of the region (\( 0.5h^{-1} \) Mpc) where the shocks
need to be identified. Moreover, the simulated region is a cubic box of comoving size $50h^{-1}$ (Miniati et al., 2000), which is too small to contain rich clusters, which are very rare objects. In fact, only groups and small clusters with temperatures in the range $0.3 - 3KeV$ are simulated by Miniati et al. (2000). It is worth noticing that groups might have very different merger rates with respect to rich clusters and, for this reason, it is not clear if the properties of the shocks found in mergers between relatively small clusters can be extrapolated to the case of more massive objects.

Recent numerical simulations (Ryu et al., 2003) seem to find more weak shocks than in Miniati et al. (2000), in closer agreement with our findings. The weakest shocks identified by Ryu et al. (2003) have a Mach number $\sim 1.5$, very close to the value corresponding to the peak of the Mach number distribution obtained from our semi-analytical approach. On the other hand, the numerical simulations of Ryu et al. (2003) also find a class of high Mach number shocks that are related to gravitationally unbound structures, not included in our semi-analytical calculations.
3.2. Merger trees and strength of the shocks in the intracluster medium

Figure 3.2: Differential (left panel) and integral (right panel) distribution of the Mach numbers of merger related shocks.

It follows that a more rigorous comparison between the results of simulations and analytical calculations is needed (Gabici & Blasi, in preparation).

The presence of the shocks at very large Mach numbers in the intergalactic medium well outside virialized structures like clusters of galaxies can be expected simply on the basis of the propagation of shocks moving with speed typically $10^8 \text{cm/s}$ in a cold medium, with say $T \sim 10^4 K$. These shocks can accelerate particles efficiently: if these particles are electrons, their radiation is expected to be concentrated around the shocks, and well outside the intracluster medium. If the accelerated particles are protons, they are expected to be advected inside the cluster by the accretion flow, where their inelastic collisions can generate secondary radiation.

3.2.2 The effects of a local overdensity

The simple method introduced in our paper in order to evaluate the Mach number of merger related shocks suffers of the limitation of being applicable only to binary mergers. The cases in which a cluster is merging with another cluster in a deeper gravitational potential well generated by a collection of nearby structures cannot be treated in the context of this method.

In this case, the relative velocity between the two clusters, and also the related shock Mach numbers may be larger (or smaller) than those estimated here. Although a rigorous evaluation of the probability of occurrence of this kind of situations cannot be carried out in the context of our simple approach, convincing arguments can be
provided to support the results discussed in the previous sections. Let us assume that our two clusters, with mass $M_1$ and $M_2$, are merging in a volume of average size $R_{sm}$ where the overdensity if $1 + \delta$ ($\delta = 0$ corresponds to matter density equal to the mean value). Clearly the overdense region must contain more mass than that associated with the two clusters, therefore for a top–hat overdensity at $z = 0$ we can write:

$$\frac{4}{3} \pi R_{sm}^3 \rho_{cr} \Omega_m (1 + \delta) = \xi (M_1 + M_2),$$

(3.17)

where $\xi > 1$ is a measure of the mass in the overdense region in excess of $M_1 + M_2$. In numbers, using $\Omega_m = 0.3$, this condition becomes:

$$(1 + \delta) = 2 \xi M_{15} R_{10}^{-3},$$

(3.18)

where $M_{15}$ is $M_1 + M_2$ in units of $10^{15}$ solar masses and $R_{sm} = 10$ Mpc $R_{10} h^{-1}$.

If the clusters are affected by the potential well of an overdense region with total mass $M_{tot}$, the maximum relative speed that they can acquire is $v_{max} \approx 2 \sqrt{GM_{tot}/R_{sm}}$. Note that this would be the relative speed of the two clusters if they merged at the center of the overdense region and with a head–on collision, therefore any other (more likely) configuration would imply a relative velocity smaller than $v_{max}$. In particular, the presence of the local overdensity might even cause a slow down of the two merging clusters, rather than a larger relative velocity. In numbers

$$v_{max} = 1.1 \times 10^8 \xi^{1/2} M_{15}^{1/2} R_{10}^{-1/2} \text{ cm/s.}$$

Using the usual expression for the sound speed in a cluster with mass $M_i$ we also get

$$c_s = 8.8 \times 10^7 M_{i,15}^{1/3} \text{ cm/s.}$$

Therefore the maximum Mach number that can be achieved in the i-th cluster is

$$M_{i,max} = 1.25 \xi^{1/2} M_{10}^{-1/2} M_{i,15}^{1/2} M_{i,15}^{-1/3}.$$  

(3.19)

As stressed in the previous sections, the Mach numbers which may be relevant for particle acceleration are $M > 3$, which implies the following condition on $\xi$:

$$\xi > 5.8 M_{15}^{-1} M_{i,15}^{2/3} R_{10},$$

(3.20)

that, when introduced in equation (3.18) gives:

$$(1 + \delta) > 11.6 R_{10}^{-2} M_{i,15}^{2/3}.$$  

(3.21)

Similar results may be obtained using the velocity distribution of dark matter halos as calculated in semi–analytical models (Sheth & Diaferio , 2001) and transforming this distribution into a pairwise velocity distribution, by adopting a suitable recipe.

The probability to have an overdensity $1 + \delta$ in a region of size $R_{sm}$ has the functional shape of a log–normal distribution, as calculated by Kayo, Taruya &
3.2. Merger trees and strength of the shocks in the intracluster medium

Table 3.1: Overdensity $\delta$ necessary for a cluster of mass $M_{i,15}$ to achieve a Mach number at least 3 in the collision with another cluster. Different smoothing radii $R_{10}$ are considered. The probability $P(\delta)$ of having the needed overdensity is reported.

<table>
<thead>
<tr>
<th>$M_{i,15}$</th>
<th>$R_{10}$</th>
<th>$(1 + \delta)$</th>
<th>$P(\delta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8</td>
<td>18.1</td>
<td>$7 \times 10^{-5}$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>11.6</td>
<td>$9 \times 10^{-5}$</td>
</tr>
<tr>
<td>1</td>
<td>1.2</td>
<td>8.1</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8</td>
<td>11.4</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>7.3</td>
<td>$10^{-3}$</td>
</tr>
<tr>
<td>0.5</td>
<td>1.2</td>
<td>5.1</td>
<td>$2 \times 10^{-3}$</td>
</tr>
<tr>
<td>0.1</td>
<td>0.8</td>
<td>3.9</td>
<td>$2 \times 10^{-2}$</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>2.5</td>
<td>$5 \times 10^{-2}$</td>
</tr>
<tr>
<td>0.1</td>
<td>1.2</td>
<td>1.7</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Suto (2001). Equation (3.21) gives the overdensity $1 + \delta$ necessary for a cluster of mass $M_i$ to achieve a Mach number at least 3 in the collision with another cluster in the same overdense region. In the table below we report the probabilities $P(\delta)$ as a function of $M_i$, evaluated following Kayo, Taruya & Suto (2001) for different sizes of the overdense region. These numbers must be interpreted as upper limits to the probability that a cluster of mass $M_i$ develops a merger shock with Mach number larger than three, since the probabilities refer to the configuration in which the relative velocity between the two clusters is maximized. For rich clusters, with masses larger than $5 \times 10^{14}$ solar masses (corresponding to X-ray luminosities $L_X > 4 \times 10^{44} \text{erg/s}$) the probability that the presence of a local overdensity may generate Mach numbers relevant for the nonthermal activity has been estimated to be $\sim 10^{-3}$ or smaller, suggesting that our two body approximation is reasonable, in particular for the massive clusters that are typically observed to have nonthermal activity. For smaller clusters, the probabilities become higher, indicating that the distribution of Mach numbers might have a larger spread compared with that illustrated in figure 3.1. Note however that for small clusters, even the two body approximation gives relatively high Mach numbers, provided the merger occurs with a bigger cluster, simply as a result of a lower temperature and a correspondingly lower sound speed.

Motivated by this argument, in the following we will continue to consider only binary mergers.

3.2.3 X–Ray observation of shock waves in the intracluster medium

Observationally, the strength of merger related shocks has been measured from high resolution X–ray images that allow one to evaluate the temperatures on the two

sides of a shock. One instance is provided by Cygnus A \((z = 0.057)\), where two subclusters with comparable masses \((\eta \sim 1)\) are merging. ASCA observations of this cluster have been used by Markevitch, Sarazin & Vikhlinin (1999) to determine the temperature upstream \((T_1 = 4 \pm 1 \text{ keV})\) and downstream \((T_2 = 8^{+2}_{-1} \text{ keV})\) of the shock, so that the compression factor \(r\) at the same shock can be determined from the Rankine-Hugoniot relations:

\[
\frac{1}{r} = \left[4 \left(\frac{T_2}{T_1} - 1\right) + \frac{T_2}{T_1}\right]^{1/2} - 2 \left(\frac{T_2}{T_1} - 1\right).
\]

(3.22)

Numerically we obtain \(r \sim 2.2 - 2.4\), corresponding to a Mach number \(M \sim 2\), roughly consistent with our predictions in figure 3.1. Other cases of current merger events provide further examples of moderately supersonic relative motion of the merging clusters (Krivonos et al., 2003; Markevitch et al., 2002; Markevitch & Vikhlinin, 2001). None of the observed shocks seems to have Mach numbers around \(5\), which appears in agreement with our findings. If the peak in Mach numbers were at \(5\) it would be statistically unfavourable to observe only Mach numbers smaller that this value.

3.3 Particle acceleration at merger shocks and related non-thermal emission

3.3.1 Particle spectra and related non-thermal emission

A shock with compression factor \(r\) and Mach number \(M\) can accelerate particles to a power law in momentum \(p^2 f(p) \propto p^{-\alpha}\), with slope \(\alpha\) related to the Mach number and compression factor by the following expressions:

\[
\alpha = \frac{r + 2}{r - 1} = 2 \left(\frac{M^2 + 1}{M^2 - 1}\right).
\]

(3.23)

The distribution function of the accelerated particles is normalized here by:

\[
4\pi \int_{p_{\text{min}}}^{p_{\text{max}}} dp E(p) p^2 f(p) = \eta_{\text{shock}} \rho u^2
\]

(3.24)

where \(E(p) = \sqrt{p^2 + m^2}\) and \(m\) is the mass of the accelerated particles, \(\eta_{\text{shock}}\) is an efficiency of acceleration, \(\rho\) and \(u\) are the density and speed respectively of the fluid crossing the shock surface. The minimum and maximum momenta \((p_{\text{min}}\) and \(p_{\text{max}}\)) of the accelerated particles are determined by the properties of the shock. In particular, \(p_{\text{max}}\) is the result of the balance between the acceleration rate and either the energy loss rate or the rate of escape from the acceleration region. Less clear is how to evaluate \(p_{\text{min}}\); the minimum momentum of the particles involved in the
3.3. Particle acceleration at merger shocks and related non-thermal emission

The particle acceleration process depends on the microphysics of the shock, which is very poorly known. Fortunately, most of the physical observables usually depend very weakly on $p_{\text{min}}$.

The basic points introduced above can be simply applied to the case of shocks originated in cluster mergers, where the approximate values of the parameters are known. In the following, we use these parameters to estimate the maximum energies attainable for electrons and protons as accelerated particles. The acceleration time, as a function of the particle energy $E$, can be written as

$$
\tau_{\text{acc}}(E) = \frac{3}{u_1 - u_2} D(E) \left[ \frac{1}{u_1} + \frac{1}{u_2} \right] = \frac{3D(E) r(r + 1)}{v^2} r - 1,
$$

(3.25)

valid for any choice of the diffusion coefficient $D(E)$, for which we consider two possible models. First we use the expression proposed in (Blasi & Colafrancesco, 1999):

$$
D(E) = 2.3 \times 10^{29} B^{-1/3}_\mu L^{2/3}_{20} E(\text{GeV})^{1/3} \text{cm}^2/\text{s},
$$

(3.26)

where $B_\mu$ is the magnetic field in microgauss and $L_{20}$ is the largest scale in the magnetic field power spectrum in units of 20 kpc. Here we assumed that the magnetic field is described by a Kolmogorov power spectrum.

In this case the acceleration time becomes:

$$
\tau_{\text{acc}}(E) \approx 6.9 \times 10^{13} B^{-1/3}_\mu L^{2/3}_{20} E(\text{GeV})^{1/3} v_8^{-2} g(r) \quad \text{s},
$$

(3.27)

where $v_8 = \frac{v}{10^8 \text{cm/s}}$ and $g(r) = r(r + 1)/(r - 1)$ and $v = u_1$.

For electrons, energy losses are dominated by inverse Compton scattering off the photons of the cosmic microwave background, provided the magnetic field is smaller than $\sim 3 \mu G$. The maximum energy of accelerated electrons is obtained by requiring $\tau_{\text{acc}} < \tau_{\text{loss}}$:

$$
E_{e,\text{max}}^e \approx 118 L^{1/2}_{20} B^{1/4}_\mu v_8^{3/2} g(r)^{-3/4} \text{GeV}.
$$

(3.28)

The relation between the compression ratio $r$ and the Mach number is

$$
r = \frac{8}{3} \frac{\mathcal{M}^2}{\mathcal{M}^2 + 2},
$$

(3.29)

valid for an ideal monoatomic gas. Here $\mathcal{M}$ is the Mach number of the unshocked gas, moving with speed $v_8$.

For protons, energy losses are usually not relevant and the maximum energy is clearly determined by the finite time duration of the merger event. Therefore the maximum energy for protons will be defined by the condition $\tau_{\text{acc}} < t_{\text{merger}}$, which gives

$$
E_{p,\text{max}}^p \approx 9 \times 10^7 L^{2}_{20} B^{6}_\mu v_8^6 g(r)^{-1/2} \text{GeV},
$$

(3.30)
for $t_{\text{merger}} \sim 10^9$ years. As a second possibility for the diffusion coefficient we assume Bohm diffusion, well motivated for the case of strong turbulence. In this case:

$$D(E) = 3.3 \times 10^{22} E / B_\mu \text{ cm}^2 / \text{s}. \quad (3.31)$$

For electrons we obtain:

$$E_{\text{max}}^e \approx 6.3 \times 10^4 B_\mu^{1/2} v_8 g(r)^{-1/2} \text{ GeV}, \quad (3.32)$$

while for protons

$$E_{\text{max}}^p \approx 3 \times 10^9 B_\mu v_8^2 g(r)^{-1} \text{ GeV}. \quad (3.33)$$

If $E_{\text{max}}^p$ becomes larger than $\sim 10^{10}$ GeV, energy losses due to proton pair production and photopion production on the photons of the microwave background become important and limit the maximum energy to less that a few $10^{10}$ GeV (Kang, Rachen & Biermann, 1997).

The relative abundance of electrons and protons at injection is an unknown quantity. Theoretically, there are plausibility arguments for having a small $e/p$ ratio, related to the microphysics of shock acceleration: protons resonate with Alfvén waves on a wide range of momenta, so that they can be efficiently extracted from the thermal distribution and injected into the acceleration engine. For electrons this is much harder. Low energy electrons do not interact with Alfvén waves and some other modes need to be excited (for instance whistlers) and sustained against their strong damping. A detailed discussion of the electron injection at non relativistic shocks can be found in McClements et al. (1997) and Levinson (1994) and references therein.

Another issue that contributes to the suppression of the injection of the electron component in a shock is the finite thickness of the shock, comparable with the Larmor radius of thermal protons (Bell, 1978a,b). Electrons can be injected in the shock accelerator only if their Larmor radius is larger than the thickness of the shock. For a proton temperature of $\sim 8$ keV, only electrons with energy larger than $\sim 5 - 10$ MeV can be injected in the acceleration box. This energy is much larger than the typical electron temperature in the intracluster medium. The value of $5 MeV$ can be adopted as a sort of low energy cutoff in the injection spectrum of electrons.

For each of these binary mergers, we evaluate the duration of the merger and the relative volumes that take part in it.

A prescription on how to define these two quantities is required since they affect the amount of energy injected in the cluster in the form of relativistic particles. Acceleration occurs at each one of the two shocks formed as a result of the merger. Once the relative velocity is calculated from equation (3.13), the duration of the merger event is defined as

$$\tau_{\text{mer}} \approx \frac{\tau_{\text{vir},1}}{V_r}$$
3.3. Particle acceleration at merger shocks and related non-thermal emission

Figure 3.3: Time-integrated average proton spectra resulting from all the mergers in a cluster. The three curves are obtained for $\Delta_m = 0.05$ (solid line), $\Delta_m = 0.1$ (dotted line) and $\Delta_m = 0.6$ (dashed line).

where $r_{\text{vir},1}$ is defined as the virial radius of the bigger cluster.

It is extremely important to keep in mind that the radiation from electrons directly accelerated at a shock is, in first approximation, solely dominated by the last merger events that the cluster suffered; on the other hand, for protons and the consequent secondary products, it is mandatory to study the full history of the cluster and account for all the shocks that traversed the cluster volume. Both acceleration of new particles and re-energization of previously accelerated particles need to be taken into account, as explained in §2.2.3.

Observations may be explained by a population of electrons (primaries or secondaries) injected in the cluster with a power law spectrum having a slope around $2.3 - 2.4$ (see for instance Blasi (2000)). In the discussion below we will refer exclusively to injection spectra, without explicitly accounting for the obvious steepening by one power in energy in the equilibrium spectrum of the radiating electrons (energy losses are dominated by inverse Compton scattering and synchrotron emission in the energy range of interest). In other words, if the injection spectrum of electrons
Figure 3.4: *Slope of the time-integrated proton spectra resulting from all the mergers in a cluster. The three curves are obtained for $\Delta_m = 0.05$ (solid line), $\Delta_m = 0.1$ (dotted line) and $\Delta_m = 0.6$ (dashed line)*

(as primaries or as secondary products of hadronic interactions) is (locally) a power law with slope $2.3 - 2.4$, the corresponding equilibrium spectra will have local slope $3.3 - 3.4$.

In the following we distinguish the two cases of primary electrons (directly accelerated at the shocks) and secondary electrons, whose spectrum at energies above a few GeV approximately reproduces the spectrum of the parent protons. We start with the case of secondary electrons, concentrating our attention upon the spectrum of the accelerated protons. We stress again that merger related shocks accelerate *new* protons and reaccelerate protons which were already confined in the parent clusters from previous times. The efficiency for particle acceleration at each shock is taken as a constant equal to 10%. For simplicity we assume that the accelerated particles are all confined in the cluster volume, although this may not be true at the highest energies for large diffusion coefficients. It seems however a good approximation for protons with the energies that we are interested in. The spectra that result from our calculations are plotted in figure 3.3. These curves are obtained averaging
3.3. Particle acceleration at merger shocks and related non–thermal emission

![Figure 3.5: Slope of the time-integrated proton spectra resulting from all the mergers in a cluster for $\Delta m = 0.05$ at different proton energies: 10 GeV (solid line), 100 GeV (dotted line) and 1 TeV (dashed line).](image)

the spectra of 500 clusters, each one followed in its merger history back in time to redshift $z = 3$. We consider the three cases $\Delta m = 0.6$ (dashed line), $\Delta m = 0.1$ (dotted line) and $\Delta m = 0.05$ (solid line).

It is clear that at high energy it is crucial to account for small mergers, since they are responsible for flatter spectra. In fact the flatter regions are also due to some level of reacceleration of pre-existing protons, confined in the intracluster volume. If only major mergers are considered ($\Delta m = 0.6$), the resulting spectrum is too steep to be of any relevance for the generation of the observed nonthermal radiation.

In figure 3.4 we plot the distribution of spectral slopes ($\alpha$ in equation (3.23)) for different choices of the threshold in mass ratio for mergers. Since the spectra are not power laws (see figure 3.3) it is convenient to plot the slopes at fixed energy, say 10 GeV (it is worth to recall that protons with this energy typically generate electrons of few GeV, which are the ones relevant for the production of radio halos). When major mergers are considered ($\Delta m = 0.6$), the typical slopes of the spectra of accelerated particles peak around 4.4, too steep to be relevant for nonthermal

Figure 3.6: Distribution of Mach numbers for the mergers occurred in the last one billion years for a cluster with present mass $10^{15}$ solar masses. The dashed line indicates the Mach number that corresponds to shocks able to accelerate electrons with a spectrum $E^{-2.4}$. The histogram is obtained from 500 different realizations of the cluster merger tree.

Radiation in clusters. When mergers between clusters with very different masses are considered, the situation improves, but still, even for $\Delta_m = 0.05$, the spectra remain too steep. An important point is that flat spectra, if any, are not obtained in major mergers but rather in those mergers that according to Fujita & Sarazin (2001) qualify as accretion events. Shocks related to major mergers are not strong enough to account for the observed spectra. The spread in the values of $\alpha$ in figure 3.4 reflects the fluctuations in the Mach numbers in figure 3.1, due to the distribution of formation redshifts in the simulation and to the stochasticity of the merger tree. Even accounting for these fluctuations, the strength of the shocks appears to be insufficient to generate the required spectra of accelerated particles.

In figure 3.5 we also plot the slope of the proton spectrum for $\Delta_m = 0.05$, at three energies, 10 GeV (solid line), 100 GeV (dotted line) and 1 TeV (dashed line). It is clear that the spectrum flattens at high energies, which is again in some...
disagreement with observations of radio halos, which seem to suggest a steepening of the radio spectrum towards its higher frequency end.

We now consider the acceleration of primary electrons as responsible for non-thermal radiation in clusters of galaxies. In this case, only recent mergers, occurred within about one billion years may be related to the observed radiation, due to the short lifetimes of relativistic electrons. Therefore, we simulate the merger tree of clusters with fixed present mass, limiting the simulations to one billion years far into the past. In fact, even one billion years is a time appreciably longer than the time for losses of relativistic electrons, so that our predictions have to be considered as optimistic. Mergers and related shocks have been treated as discussed in the previous sections.

Of the 500 clusters with mass $10^{15}$ solar masses that we simulated the merger tree of, about 30% suffered a merger during the last billion years. In figure 3.6 we plotted the Mach numbers of the shocks generated in the clusters that suffered at least one merger (actually it is rare to have more than one merger during the last billion years). The dashed line represents the value of the Mach number that would correspond to a slope of the accelerated electrons equal to 2.4, required to explain observations. Only 20% of the shocks have Mach numbers fulfilling this condition, so that in the end, about 6% of the 500 simulated $10^{15}$ solar masses clusters of galaxies have suffered a merger than may have generated non-thermal activity with Coma-like spectrum. However, due to the short diffusion length of relativistic electrons, the radio emission should be observable only close to the shock surfaces. Thus, primary electrons accelerated at merger shocks cannot generate radio emission on Mpc scale, but might produce sources whose morphology could resemble the so-called radio relics, observed in a number of X-ray bright clusters (Ensslin, 2003).
As seen in the previous chapter, shocks are driven by gravity during the process of large scale structure formation. Diffusive acceleration takes place at these shock surfaces and particles can be energized up to ultrarelativistic energies. These high energy particles are expected to produce gamma ray photons through their interactions with the intergalactic gas and with the background radiation fields. In this chapter, after a brief overview on gamma ray astronomy, we discuss the possibility of detecting forming clusters of galaxies at gamma ray energies with next generation instruments. Besides merger shocks, already considered in the previous chapter, we include in the present analysis also accretion shocks, which are expected to form at the cluster periphery due to continuum infall of material (see §1.3). Our predictions are specialized to the cases of the GLAST and AGILE satellites, for which a few tens of objects are expected to be detected. The capability of ground based Čerenkov telescopes to detect TeV photons from nearby forming clusters is also discussed. Finally, we evaluate the contribution of structure formation to the extragalactic diffuse gamma ray background, concluding that it adds up to at most 10% of the observed one.

4.1 Gamma ray astrophysics: a brief overview

The first complete survey of the gamma ray sky in the energy range from about 30 $MeV$ to 30 $GeV$ was performed by the EGRET telescope, on board of the Compton Gamma Ray Observatory (Nolan et al., 1992). The most evident feature in the all sky map is the presence of a strong diffuse emission coming from the galactic plane. This emission is the result of the interaction between galactic cosmic rays and matter or radiation fields.

Both relativistic protons and electrons can contribute to the diffuse emission. Galactic cosmic ray protons having mass $m_p$ and kinetic energy greater than the
4. Gamma ray emission during large scale structure formation

threshold value:

\[ E - m_p c^2 = 2m_\pi c^2 + \left( \frac{m_\pi}{2m_p} \right) m_\pi c^2 \sim 280 \text{MeV} \]  

(4.1)

can interact with the protons of the interstellar gas and produce neutral pions, which in turn decay into gamma ray photons:

\[ pp \to \pi^0 + \text{anything} \]  

\[ \pi^0 \to \gamma\gamma \]  

(4.2)

The resulting gamma ray spectrum shows a peak corresponding to the energy \(1/2m_\pi c^2\), while at higher energies the spectral shape is the same of the parent cosmic ray population, which is \( \propto E^{-2.75} \) for galactic cosmic rays (Dermer, 1986). Close to the peak energy, \( \pi^0 \) decay is the main contributor to the total emission from the galactic plane, while inverse Compton scattering from cosmic ray electrons dominates elsewhere (Moskalenko, Strong & Reimer, 2004).

The photon spectrum due to inverse Compton scattering can be easily calculated when the product of the electron Lorentz factor \( \gamma \) and the typical energy of the background photons \( \epsilon \) is much lower than \( m_p c^2 \). If this condition is satisfied, we can adopt the Thomson cross section \( \sigma_T \) to describe the scattering. Moreover, if \( \epsilon E_\gamma \ll (m_p c^2)^2 \) the energy of the \( \gamma \)-ray produced is on the average:

\[ E_\gamma = \frac{4}{3} \gamma^2 \epsilon \]  

(4.3)

Within the assumption that cosmic ray electrons have an energy distribution of the form \( I(E) = KE^{-\alpha} \), it is straightforward to calculate the spectrum of the upscattered photons, which is given by (Stecker, 1974):

\[ q(E_\gamma) = 4 \pi n_{ph} \sigma_T \int dE KE^{-\alpha} \delta \left( E_\gamma - \frac{4}{3} \epsilon \gamma^2 \right) \propto E_\gamma^{-\alpha/3} \]  

(4.4)

where \( n_{ph} \) is the number density of target photons and we have made the delta function approximation for the differential cross section:

\[ \sigma(E_\gamma | \epsilon, E) \sim \sigma_T \delta \left( E_\gamma - \frac{4}{3} \epsilon \gamma^2 \right) \]  

(4.5)

In the Galaxy, CMB and optical/IR photons from stars contribute to \( n_{ph} \) while cosmic ray electrons have a power law spectrum with \( \alpha \sim 3 \). So, the galactic gamma ray spectrum due to inverse Compton scattering is expected to be equal to \( q(E_\gamma) \propto E_\gamma^{-2} \), which at high energies is flatter than the proton related component.

Neutral pion decay and inverse Compton scattering are the most important mechanisms for gamma ray production not only in the Galaxy, but also in several astrophysical objects, including clusters of galaxies and shocks related to the process
of large scale structure formation, which are the subject of our research. Other less important contributions come from electron Bremsstrahlung and inverse Compton scattering of secondary electrons produced during the inelastic proton–proton interactions between cosmic rays and interstellar matter.

Apart from the galactic emission, EGRET detected ~ 300 point sources, roughly half of which still need to be identified with objects observed at other wavelengths. The identification of these sources is one of the most important open problems of gamma ray astronomy. We are interested here in the high latitude, presumably extragalactic sources. In this case, the majority of the identified sources are active galaxies belonging to the blazar class (Hartman et al., 1999). Despite the fact that several candidates have been proposed to explain the origin of the unidentified sources, including radio galaxies (Torres, 2004), clusters of galaxies (see next section for a detailed discussion) and high latitude molecular clouds (Walker, Mori & Ohishi, 2003), it is impossible to date to make firm claims on their nature.

Another important issue concerns the exact determination of the spectrum and flux level of the isotropic extragalactic background. This component can be revealed by subtracting the galactic contribution from the total gamma ray map. Thus, results depend much onto the adopted model of the Galactic background, which is not yet firmly established. Sreekumar et al. (1998) used the relation between the model predicted Galactic emission and the total diffuse emission to estimate the extragalactic background as the extrapolation to zero Galactic contribution. They considered photon energies in the range $100\text{MeV} - 30\text{GeV}$ and obtained a featureless power law spectrum containing roughly equal energy flux per decade in photon energy at the level of $\sim 1\text{keV}/(\text{cm}^2 \cdot \text{s} \cdot \text{sr})$.

Faint, undetected gamma ray blazars have been considered for a long time to be the main contributors to the extragalactic background (Padovani et al., 1993; Setti & Woltjer, 1994). Even if there is now a general consensus that blazars should contribute to the observed diffuse emission, it is not yet clear whether they can saturate it (Comastri, Di Girolamo & Setti, 1996; Stecker & Salamon, 1996) or not (Mucke & Pohl, 2000; Mukherjee & Chiang, 1999).

However, due to the uncertainty in the determination of the Galactic contribution at high latitude, very different scenarios might also be possible. In fact, Dar & De Rújula (2001) have proposed that observations can be also explained as the result of a population of relativistic electrons upscattering the microwave and starlight radiation to gamma ray energies through inverse Compton scattering. These electrons, located in the Galactic halo, would not be correctly accounted for in standard models of cosmic ray propagation in the Galaxy. As we will see in the following, also forming structure in the universe have been proposed to be the sources of the diffuse extragalactic background (Loeb & Waxman, 2000). This issue, together with the possibility to detect forming clusters in gamma rays, will be discussed in great detail in the next sections.
4.2 Gamma rays from large scale structures in the universe: review of past theoretical and observational results

As seen in §3.1, radio observations of galaxy clusters tell us that relativistic electrons, and presumably protons, are present in the intracluster medium (Feretti, 2003). From X-ray observations we also know that clusters contain a considerable amount of baryons in the form of a hot and diffuse gas (Sarazin, 1988). These two observational facts led Dennison (1980) to propose that clusters might be gamma ray sources, due to the inelastic interactions between cosmic ray protons and protons in the intergalactic medium (see equation 4.2).

Dar & Shaviv (1995) considered the same emission mechanism and, within the assumption of universality of the cosmic ray spectrum, proposed that the whole extragalactic diffuse gamma ray background could be originated in groups and clusters of galaxies. Some problems were identified in these calculations and discussed by Stecker & Salamon (1996b) and Berezinsky, Blasi & Ptuskin (1997). Moreover, Berezinsky, Blasi & Ptuskin (1997) made the important discovery that the hadronic cosmic ray component is confined within the cluster volume for cosmological times. This is a fact that needs to be taken into account for a correct determination of the expected gamma ray emissivity from clusters (Berezinsky, Blasi & Ptuskin (1997) and see §3.1.1 for a quantitative discussion of the issue).

A more realistic calculation of the extragalactic background due to clusters was carried out in (Colafrancesco & Blasi, 1998), where the authors concluded that not more than a few percent of the observed background of gamma rays could be accounted for in terms of hadronic interactions in clusters of galaxies. Colafrancesco & Blasi (1998) also estimated the gamma ray fluxes for single, nearby and rich clusters, finding these fluxes to be all below the EGRET sensitivity.

More recently, Loeb & Waxman (2000) have reproposed a connection between the extragalactic gamma ray background and clusters of galaxies. More correctly the connection should exist between the diffuse background and the process of hierarchical large scale structure formation. The claim is that the whole background can be explained in terms of inverse Compton scattering of ultrarelativistic electrons accelerated at shocks related to the process of structure formation. Shocks form naturally in the intracluster medium during the merger of two halos that generate a new bigger structure.

Following this earlier optimistic claim, Totani & Kitayama (2000) and Waxman & Loeb (2000) used an approach based on the Press–Schechter formalism (see §1.2) to predict the existence of a connection between forming clusters of galaxies and unidentified high latitude EGRET sources. This prediction seemed to be supported by some preliminary observational evidences for a temptative association between unidentified high latitude EGRET sources and Abell clusters (or cluster
4.3 Detectability of clusters with space telescopes

Moreover, a statistical detection of gamma ray emission from clusters of galaxies was claimed by Scharf & Mukherjee (2002). The statistical significance and physical plausibility of such an association was strongly questioned by Reimer et al. (2003). The authors, after an accurate analysis of the EGRET data for a complete sample of objects, reached the important conclusion that we still have to await the first observational evidence for the high-energy gamma-ray emission of galaxy clusters. Only upper limits for the gamma ray fluxes from nearby and X-ray bright clusters can be given (Reimer, 1999; Reimer et al., 2003).

4.3 Detectability of gamma rays from forming clusters of galaxies with space–borne gamma ray telescopes

As discussed in many previous papers, gamma rays can be generated in clusters of galaxies mainly through the following physical processes:

1. inverse Compton scattering of electrons accelerated at merger and accretion shocks or injected in the intracluster medium by AGNs or normal/starburst galaxies;

2. decay of neutral pions produced in inelastic $pp$ collisions;

3. inverse Compton scattering of secondary electrons from the decay of charged pions produced in inelastic $pp$ collisions;

4. Bremsstrahlung emission of relativistic primary and secondary electrons;

with the first two mechanisms dominating the total emission (Blasi, 2001).

Due to cosmic ray confinement, protons are stored in the intracluster medium for cosmological times, and in principle can produce secondary products continuously (Berezinsky, Blasi & Ptuskin, 1997; Colafrancesco & Blasi, 1998), namely with no time correlation with phenomena like mergers of two clusters. While this process is of great importance and can be responsible for appreciable gamma ray emission, it is also quite uncertain because the amount of protons diffusively stored in the intracluster medium is determined by the history of the cluster (see §3.1.1). In the following we discuss only three of the elements that contribute to this uncertainty:

**Mergers as accelerators**

Different clusters have different merger histories, namely different merger trees. As stressed above, most mergers produce weak shocks, unable to efficiently energize cosmic ray particles. Depending on the cluster, there may be several strong shocks formed when mergers between clusters with very different masses take place. The
energy density in protons trapped within the cluster that will eventually result from several merger events occurring during its history depends on the number and strength of the shocks that accelerated those protons. Since there are only a few strong shocks per merger tree, the spectrum of accelerated protons at the high energies relevant for gamma ray emission can vary a lot for different clusters.

Mergers as re-accelerators

While most electrons accelerated during a merger lose most of their energy before the next merger event, protons are simply stored in the intracluster medium and each new merger related shock re-energizes the particles trapped until that time, besides accelerating new particles. Again, the merger history affects this component in a crucial way. Moreover, the spectrum of the re-energized particles depends on the spectrum of the protons accelerated and re-energized by previous mergers. The spectrum of the re-accelerated particles also depends upon the minimum energy that the protons possess. These effects may change the spectrum of the accelerated particles trapped in the intracluster medium even by orders of magnitude, in particular at the highest energies, where the spectrum depends crucially on the amount of energy that crossed the few strong shocks developed during the merger history of the cluster. A corresponding uncertainty affects the secondary products of cosmic ray interactions in the intracluster medium, namely gamma rays and electrons (positrons).

Additional sources

The storage properties that clusters of galaxies exhibit for the proton component make them very sensitive to all nonthermal events occurring within the cluster. In particular, acceleration processes in radio galaxies in cluster cores pollute the intracluster medium with accelerated protons which have no connection with shocks related to structure formation.

All these uncertainties can be seen as pieces of physics of the cosmic ray acceleration in clusters of galaxies that require further investigation. In the present work on the other hand we adopt a conservative attitude and restrict our attention to the gamma ray emission that is solely the result of inverse Compton scattering of relativistic electrons, accelerated at merger and accretion shocks, against photons of the CMB radiation. Any contribution to the gamma ray emission from proton interactions can only increase the gamma ray fluxes derived below.

4.3.1 Shock acceleration during structure formation

Shocks related to structure formation may serve as cosmic ray accelerators. In this section we estimate the maximum energy achievable by electrons accelerated in one of these shocks.

The acceleration time, necessary to energize a particle to energy $E$ is given by
4.3. Detectability of clusters with space telescopes

(Lagage & Cesarsky, 1983):

$$\tau_{\text{acc}}(E) = \frac{3}{u_1 - u_2} \left[ \frac{D_1(E)}{u_1} + \frac{D_2(E)}{u_2} \right], \quad (4.6)$$

where $u_1$ and $D_1(E)$ ($u_2$ and $D_2(E)$) are the fluid speed and diffusion coefficient of the particles upstream (downstream). From the physical point of view, $\tau_{\text{acc}}$ is the sum of the residence times of particles upstream (index 1) and downstream (index 2). For well behaved diffusion coefficients, larger magnetic field strengths correspond to lower values of the diffusion coefficient, namely particles diffuse more slowly when the magnetic field is larger. This immediately suggests that the diffusion coefficient upstream is larger than the same quantity (at the same particle energy) downstream, so the residence time of the particles upstream is larger than the residence time in the downstream region. The acceleration time can therefore be approximated as follows:

$$\tau_{\text{acc}}(E) \approx \frac{3D_1(E)r}{u_1^2(r-1)} = \frac{4D_1(E)}{u_1^2} \left\{ \frac{\mathcal M^2}{\mathcal M^2 - 1} \right\}, \quad (4.7)$$

where we introduced the compression factor $r = u_1/u_2$ and we used the relation between $r$ and the Mach number $\mathcal M$, $r = 4\mathcal M^2/(\mathcal M^2 + 3)$. It is worth noticing that the acceleration time depends in first approximation only on quantities that refer to the upstream fluid, where the magnetic field has not been affected yet by the passage of the shock, and preserves therefore its pre-shock structure. One may argue that the diffusion coefficient is dominated by the scattering of particles on magnetic field fluctuation lengths comparable to the Larmor radius of the particles. In this case, a reasonable approximation could be to assume a Böhm diffusion coefficient, in the form $D_2(E) \approx (1/3)cE/eB_2$. It is likely that these fluctuations are in fact created by the accelerated particles themselves. In the upstream region however it is possible that the diffusion coefficient is determined by the magnetic field structure pre-existing the shock transit. In principle the momentum and magnetic field dependence of the two fields upstream and downstream may be different.

For simplicity and convenience we assume here that the diffusion coefficient on both sides is Böhm-like. Moreover, since the term in brackets in equation (4.7) varies between 1.8 and 1 when the Mach number changes between 1.5 and infinity, we neglect this term for the purpose of estimating the acceleration time, therefore the acceleration time becomes:

$$\tau_{\text{acc}} \approx 0.3B_\mu^{-1}E(\text{GeV}) \left( \frac{u_1}{10^8 \text{cm/s}} \right)^{-2} \text{years}. \quad (4.8)$$

We stress again here that the magnetic field $B_\mu = (B/\mu G)$ and velocity $u_1$ refer to the unshocked medium. Here we focus our attention on electrons. Their energy losses are dominated by inverse Compton scattering, with loss rate (Rybiki &
4. Gamma ray emission during large scale structure formation

Lightman (1979):

\[
\left( \frac{dE}{dt} \right)_{ICS} = 4 \frac{\sigma_T c U_{CMB} \gamma^2}{3 T_c} \tag{4.9}
\]

and time scale:

\[
\tau_{ICS} = \frac{E}{\left( \frac{dE}{dt} \right)} = 10^9 E(\text{GeV})^{-1} \text{ years}, \tag{4.10}
\]

where \(U_{CMB}\) is the energy density in the CMB radiation and \(\gamma\) is the electron Lorentz factor. Equating the acceleration and losses time scales we obtain an estimate of the maximum energy of accelerated electrons in case of Böhm diffusion:

\[
E_{\text{max}} \approx 57 B_{\mu}^{1/2} \left( \frac{u_1}{10^8 \text{cm/s}} \right) \text{ TeV}. \tag{4.11}
\]

For electron energies up to \(\sim 400\) TeV the inverse Compton scattering off the CMB photons occurs in the Thomson regime, therefore the maximum energy of the radiated photons is simply:

\[
E_{\gamma,\text{max}} = \frac{4}{3} \left( \frac{E}{m_e c^2} \right)^2 \epsilon_{CMB} \approx 7.5 B_{\mu} \left( \frac{u_1}{10^8 \text{cm/s}} \right)^2 \text{ TeV}. \tag{4.12}
\]

As shown by Blasi (2001) and Gabici & Blasi (2003b, 2004), if the diffusion coefficient in the unshocked medium is not Böhm-like, then the maximum energy of the electrons may be appreciably lower than found in equation (4.11), and insufficient to upscatter the CMB photons to gamma ray energies. In the following we assume that the maximum energy is well defined by equation (4.11).

The value of the magnetic field to adopt in our calculations depends crucially on whether we are describing particle acceleration at merger related shocks, or at accretion shocks. The former have relatively low Mach numbers (Gabici & Blasi, 2003) and occur in the virialized regions of clusters of galaxies, where the magnetic field is expected to be in the \(\mu\)G range. Accretion shocks propagate in a cold unshocked medium (Bertschinger, 1985; Gabici & Blasi, 2003b), where the magnetic field is expected to set at the cosmological value, for which only upper limits are available. The field in the intracluster medium could be the result of magnetic pollution from sources within the cluster (Kronberg, Lesch & Hopp, 1999), being therefore unrelated to the magnetic field in the intergalactic medium outside clusters. Alternatively, the intracluster magnetic field could result from the compression of the external field, if the latter is of cosmological origin. The energy of the photons radiated by the accelerated electrons through inverse Compton scattering is above \(E_\gamma\), whenever the magnetic field \(B_{\mu}\) is larger than \(1.3 \times 10^{-3} (u_1/10^8 \text{cm/s})^{-2} (E_\gamma/10 \text{GeV}) \mu\text{G}\). This bound is certainly satisfied by the intracluster magnetic field, not necessarily by the poorly known field in the intergalactic medium outside clusters, for which upper limits of \(10^{-9} - 10^{-11} \mu\text{G}\) have been found, depending on the field structure.
4.3. Detectability of clusters with space telescopes

(Blasi, Burles & Olinto, 1999). On the other hand this constraint has to be taken with caution for the case of accretion shocks, because in a more realistic situation, as suggested by numerical simulations, high Mach number shocks are located close to filaments (Miniati et al., 2000), where a substantial compression of cosmological fields might have occurred. We assume in the following that the electrons accelerated at both merger and accretion shocks are energetic enough to radiate gamma ray photons up to the $\sim$ TeV region.

The energy distribution of the accelerated particles per unit time is a power law in energy of the form $q(E) \propto E^{-\alpha}$, with slope $\alpha$ given by $\alpha = 2(\mathcal{M}^2 + 1)/(\mathcal{M}^2 - 1)$ where $\mathcal{M}$ is the relevant Mach number. Electrons accelerated at large scale shocks are injected in the intergalactic medium, where they loose their energy mainly via inverse Compton scattering on a timescale which is much shorter than the lifetime of the shock. As a consequence, the balance between energy losses and continuous injection of newly accelerated particles, drives the electrons toward their time independent equilibrium distribution, with a spectrum one power steeper than the injected one. The number of photons emitted per unit energy and unit time can be easily calculated by:

$$n_{eq}(E_\gamma) = q(E) \tau_{ICS} \left( \frac{-dE}{dt} \right) \frac{dE}{dE_\gamma} \frac{1}{E_\gamma} \propto E_\gamma^{-\frac{\alpha}{2}-1} \tag{4.13}$$

where $E_\gamma$ is the gamma ray photon energy.

### 4.3.2 Source counts for merging and accreting clusters

We consider first merger shocks whose strength is a function of the masses of the clusters involved in the merger, and more specifically of the ratio of the masses (see §3.2). As discussed above, merger shocks are often weak, following the distribution given in figure 3.2. Flat electron spectra, necessary to generate an appreciable gamma ray emission above 100 MeV, are obtained at strong shocks that form only during mergers between clusters with very different masses. In the approach presented by Gabici & Blasi (2003) and reviewed in chapter 3, at each merger event two shocks are generated, propagating in each one of the merging clusters. The Mach numbers $\mathcal{M}$ of both shocks can be calculated from the relative velocity of the clusters and if the sound speed in each cluster is known. This allows us to estimate the spectral slope of the accelerated particles.

The rate at which energy is channelled into nonthermal electrons in merger related shocks can be evaluated as

$$L_e^{\text{mer}} = \eta_2 q_{b,i} v^3 S_i \tag{4.14}$$

where $q_{b,i}$ is the baryonic density of the $i-th$ cluster, $S_i$ is the surface of the shock that propagates in the $i-th$ cluster and $v$ is the merger relative velocity.
4. Gamma ray emission during large scale structure formation

The gamma ray luminosity due to inverse Compton scattering of the accelerated electrons is calculated according with equation (4.13).

The number of observable merging clusters is (Gabici & Blasi, 2004):

\[ N_{\text{mer}}(F_{\text{lim}}) = \int_0^1 \frac{dV}{dM} \int_{M_{\text{min}}}^{M_1} dM_1 n(M_1, z) \times \]

\[ \times \int_{M_{\text{min}}}^{M_1} dM_2 R(M_1, M_1 + M_2, z) \Delta t_{\text{mer}} \vartheta [F_{\text{lim}} - F_{\gamma}(M_1, M_2, z)], \]  

(4.15)

where \( n(M, z) \) and \( R(M_1, M_1 + M_2, z) \) are the Press–Schechter mass distribution and the merger rate given in §3.2.1,2.3, \( F_{\text{lim}} \) is the telescope sensitivity, \( F_{\gamma}(M_1, M_2, z) \) is the gamma ray flux we receive when a cluster with mass \( M_1 \) merges with a cluster with mass \( M_2 \) at a redshift \( z \), \( \Delta t_{\text{mer}} = R_2/v \) is the duration of the merger event, \( M_{\text{min}} = 10^{13} M_\odot \) is the typical mass for galaxy groups and \( \vartheta \) is the Heaviside step function.

We consider here also the accretion shocks, forming at the cluster periphery due to the secondary infall of matter onto an already formed object (Bertschinger, 1985). For simplicity, let us assume that the shock is located exactly at the virial radius \( R_v \) and that the velocity of the accreting matter at the shock is the free fall velocity. The accretion shock, by definition, propagates in a cold (non–virialized) medium, therefore its Mach number may be very high, although the typical speed of the shock is of the same same order of magnitude as the merger speed of two clusters. The exact value of the Mach number depends on the temperature of the medium before entering the overdense virialized region. If we take for such temperature the range of values \( T = 10^4 - 10^6 \) K, and a typical shock speed of \( \sim 10^8 \) cm/s, the corresponding Mach numbers range between 10 and 100. These Mach numbers, being much larger than unity, correspond to spectra of accelerated particles which are \( \sim E^{-2} \), with a cutoff at the maximum energy. If the intergalactic medium were pre–heated before the gravitational collapse, these Mach numbers could be lower (Totani & Inoue, 2002).

The energy per unit time converted at the shock into nonthermal electrons can be written as

\[ L_{\text{e}}^{\text{acc}} = \frac{1}{2} \eta_\text{e} (1 + z)^3 \frac{v_{\text{ff}}^3 R_v^2}{4 \pi} \propto M_5^{5/3} \]  

(4.16)

where \( \eta_\text{e} \) is the shock acceleration efficiency for electrons, fixed here at 5%, \( v_{\text{ff}} = (2GM/R_v)^{1/2} \) is the free fall velocity of the gas at the virial radius \( R_v \) and \( \eta_\text{b} = \Omega_\text{b}\rho_\text{cr} \) is the present mean baryonic density of the universe, if \( \Omega_\text{b} = 0.045 \) is the baryon fraction and \( \rho_\text{cr} \) is the critical density. The secondary infall just described is a simplification of the mass flow through large scale shocks in the filamentary structures seen in N–body simulations. Although the geometry is different, the total energy crossing the shock per unit time and per unit surface should not be very different from the same quantity calculated for spherical inflow.
4.3. Detectability of clusters with space telescopes

Figure 4.1: Number of accreting (solid line) and merging (dashed line) clusters with gamma ray flux greater than $F$. The vertical lines represent the GLAST, AGILE and EGRET sensitivity for point sources.

It is important to keep in mind that the accretion of gas onto the cluster only contributes a small fraction ($\sim 10\%$) of the total final mass, which is instead mainly built up through mergers with other massive clusters.

The differential gamma ray luminosity, expressed in photons per unit energy per unit time, is easily calculated as inverse Compton scattering of the CMB photons, and is proportional to $L_{\text{acc}}$ given above. It follows that, for a given redshift, the gamma ray luminosity is a growing function of the cluster mass, so that the number of accreting objects observable by a gamma ray telescope with sensitivity $F_{\text{lim}}$ can be written as follows (Gabici & Blasi, 2004):

$$N_{\text{acc}}(F_{\text{lim}}) = \int_0^\infty dz \frac{dV}{dz} \int_{M(F_{\text{lim}}, z)}^\infty dM n(M, z),$$  \hspace{1cm} (4.17)

where $dV$ is the comoving volume in the redshift region between $z$ and $z + dz$, $n(M, z)$ is given by equation (1.53) and $M(F_{\text{lim}}, z)$ is the mass of a cluster accreting at redshift $z$ whose flux is $F_{\text{lim}}$. 

61
4. Gamma ray emission during large scale structure formation

The aim of this section is to determine the log $N$ – log $S$ distribution of clusters of galaxies as gamma ray sources, for the cases of mergers and accretion. The results of our calculations are shown in figure 4.1 in the form of number of clusters with gamma ray emission at energies in excess of 100 MeV larger than some value $F$ on the x-axis: the solid and dashed lines refer to accretion and mergers respectively. The vertical lines correspond to the GLAST, AGILE and EGRET sensitivities for point sources, as indicated in the plot.

The fact that the two curves in figure 4.1 overlap almost exactly is the result of a combination of many factors: each merger is on average much more luminous that accretion, because the energy involved is larger and the time during which it is released is shorter ($\sim 10^9$ years for a merger versus the cluster lifetime for accretion). On the other hand, most of the times the merger event does not result into gamma ray emission, because the shocks involved are too weak. The slight deviation from a power law at low fluxes in figure 4.1 is the consequence of the cosmological effects in a non-euclidean geometry. The change of slope appears at larger fluxes for the mergers (dashed line) than for accretion. This is a result of the larger luminosity per merger that allows one to detect a bright merger from larger distances but also the consequence of the strong evolution in the merger luminosity. This latter effect is easy to explain: for a minimum cluster mass of $\sim 10^{15} M_\odot$, the flat spectra are obtained only when a merger occurs with a cluster of mass $\sim 10^{15} M_\odot$; these clusters are already on the tail of the Press-Schechter distribution at the present cosmic time, while they had not been formed yet at much earlier times. Despite the larger luminosity of the mergers (on average), it is interesting to notice that the gamma ray emission for the merger case is limited to the merger period or shortly after, while the gamma ray emission associated to accretion is continuous, although less intense.

Our calculations show that $\sim 50$ clusters should be detected in gamma rays above 100 MeV by GLAST, equally distributed between merging and accreting clusters. AGILE on the other hand should be able to detect $\sim 10 - 20$ objects. Moreover, we predict that no cluster should have been detected by EGRET, in perfect agreement with observations (Reimer et al., 2003).

The results summarized in figure 4.1 are compared with the sensitivity of EGRET, GLAST and AGILE as obtained for point sources. This seems justified at least for gamma rays with energy around 100 MeV: the gamma ray emission region in clusters is approximately a few degrees wide and the point spread function of GLAST at 100 MeV is $\sim 4^\circ$, while it reduces to $0.1^\circ$ at 10 GeV (see http://www-glast.slac.stanford.edu/software/IS/glast(lat_performance.htm). Similar arguments hold even for the AGILE angular resolution (details can be found at http://agile.mi.iasf.cnr.it/Homepage/performances_3.shtml).
4.4 The diffuse gamma ray background from large scale structure formation

The superposition of the emission from single clusters results in a diffuse background. Using the formalism developed above, we can now estimate the contribution from forming clusters to the extragalactic diffuse background radiation. For merging clusters, the flux of gamma radiation (in units of \( \text{phot cm}^{-2}\text{s}^{-1}\text{sr}^{-1}\text{GeV}^{-1} \)) is then given by (Gabici & Blasi, 2003b):

\[
I_\gamma(E_\gamma) = \frac{c}{4\pi H_0} \int_0^{z_{\text{max}}} dz \frac{1}{S(z)} \int_0^\infty dM \ n(M, z) \times \int_0^M dM' \mathcal{R}(M, M + M', z) Q_\gamma(E_\gamma(1 + z), M, M') \Delta t_{\text{mer}}(M, M'),
\]

(4.18)

where \( S(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda} \), \( \mathcal{R}(M, M + M', z) \) is the merger rate between clusters of masses \( M \) and \( M' \) at redshift \( z \), and \( \Delta t_{\text{mer}}(M, M') \) is the duration of the merger between two clusters with given masses \( M \) and \( M' \), as defined in Gabici & Blasi (2003). For each merger between two clusters with mass \( M \) and \( M' \), \( Q_\gamma(E_\gamma(1 + z), M, M') \) is the rate of gamma ray production from the accelerated electrons.

The diffuse flux from accreting clusters can be estimated in a similar way (Gabici & Blasi, 2003b):

\[
I_\gamma(E_\gamma) = \frac{c}{4\pi H_0} \int_0^{z_{\text{max}}} dz \frac{1}{S(z)} \int_0^\infty dM n(M, z) Q_\gamma(E_\gamma(1 + z), M, z).
\]

(4.19)

where \( Q_\gamma(E_\gamma, M, z) \) is the gamma ray production rate in \( \text{phot s}^{-1}\text{GeV}^{-1} \).

The diffuse flux of gamma radiation from mergers (dashed line) and from accretion (dotted line) is plotted in figure 4.2, where an acceleration efficiency (for electrons) of 5% is assumed. The observed background is the shaded region (Sreekumar et al., 1998). In the same figure we plot for comparison the predictions by Loeb & Waxman (2000) (solid line), who also adopted an acceleration efficiency equal to 5%. The meaning of the dash-dotted line will be explained below.

Three conclusions are evident:

1) the flux of gamma radiation from both mergers and accretion is a factor \( \sim 10 \) smaller than the observed background and smaller than the flux predicted by Loeb & Waxman (2000) and Totani & Kitayama (2000), by the same factor. The discrepancies with these previous estimates will be discussed in detail in the next section. An acceleration efficiency of the order of 50% should be adopted in order to reproduce observations. This would be unreasonable for electrons as accelerated particles, and would violate our initial assumption of shock acceleration in the linear regime (no backreaction of the accelerated particles on the shock);
4. Gamma ray emission during large scale structure formation

Figure 4.2: Diffuse gamma ray emission from structure formation. The shaded area is the result EGRET observations. The dashed line is the result of our calculations for mergers while the dotted line is the flux of gamma rays from accretion. The dash-dotted line assumes a minimum mass of the merging halos of $10^{13} M_\odot$.

2) the gamma ray diffuse flux from mergers is at the same level as that due to accretion (secondary infall);

3) all predicted spectra are approximately power laws with slopes between 2 and 2.1.

The flux of diffuse gamma rays due to accretion (secondary infall) has a spectrum which is exactly $E^{-2}$ because the Mach number of the accretion shock is always much larger than unity. It is somewhat surprising that the diffuse gamma ray background contributed by electrons accelerated at the accretion shock is comparable with that produced in merger events (the latter, as stressed above, is dominated by mergers between clusters with $M'/M \ll 1$, that in some literature are indeed defined as accretion events). In this respect however some additional discussion is required: for a cluster with mass $10^{14} M_\odot$ a mass ratio $M'/M \sim 10^{-2}$ corresponds to a substructure with mass $10^{12} M_\odot$, comparable with the mass of our galaxy. This clearly does not make physical sense. Galaxies move within the intracluster medium without their
medium being shocked. It is more likely that a bow shock is formed in front of
the galaxy, due to the internal pressure of the galactic medium (Stevens, Acreman
& Ponman, 1999). Actually simulations show that large galaxies penetrating the
intracluster medium of a rich cluster can even be completely stripped of their gas
content (Stevens, Acreman & Ponman, 1999). This suggests that a low mass
cutoff should be imposed in the calculation of the gamma ray diffuse background
from cluster mergers. The dash-dotted line in figure 4.2 has been obtained by
considering only structures with virial masses larger than $10^{13} M_{\odot}$, corre-
spending to galaxy groups. The diffuse background of gamma rays is a factor $\sim 10$
in this case and is slightly steeper in spectrum. This happens because the main
contribution comes from mergers between clusters with masses $M_{\text{min}} = 10^{13} M_{\odot}$
and $M = M_{\text{min}}/10^{-2} \approx 10^{15} M_{\odot}$. Clusters with masses as large as $10^{15} M_{\odot}$ are
already on the tail of the Press-Schechter distribution even at $z = 0$, therefore the
corresponding contribution to the diffuse background is suppressed. On this basis,
the dash-dotted line in figure 4.2 is the most realistic estimate of merger shocks to
the diffuse gamma ray background, amounting to $\sim 1\%$ of the observed one. This
result agrees with the estimate given by Berrington & Dermer (2003). On the other
hand the strong shocks associated to accretion of matter onto a cluster may generate
a gamma ray background as large as that plotted as a dotted line in figure 4.2, and
this contribution remains at the level of $\sim 10\%$ of the observed background. This
may be considered as the most realistic prediction of the contribution of clusters of
galaxies to the diffuse extragalactic gamma ray background.

4.5 Comparison with other theoretical results

Since our findings are somewhat different from at least some previous results of
other authors, some comments and comparisons are required. The most important
point in our approach is the fact that the Mach number of the shocks is estimated
in a self-consistent way, as explained in §3.2 (Gabici & Blasi, 2003). The correct
estimate of the shock Mach number is of crucial importance if one is interested in
the gamma ray emission from forming structures. This can be better clarified by
using figure 4.3. The upper panel shows the average normalized energy flux per unit
time through the merger related shocks of a cluster with mass $10^{15} M_{\odot}$ (solid line),
$10^{14} M_{\odot}$ (dotted line) and $10^{13} M_{\odot}$ (dashed line) at redshift $z = 0$, as a function
of the mass of the merging subcluster (here $M'/M \leq 1$ is the ratio between the
masses of the two subclusters). The curves represent the energy flux contributed by
mergers with mass ratio larger than $M'/M$. The energy flux sums up to $\sim 70 - 80\%$
of the total for subcluster masses larger than $\sim 0.1 M$. This implies that the energy
flux that crosses the shocks formed during mergers of the cluster with mass $M$ and
subclusters with masses smaller than $0.1 M$ is rather small ($20 - 30\%$). In other
words, the energy flux is dominated by major mergers.
Figure 4.3: Upper panel: normalized energy flux per unit time through the merger related shocks of a cluster with mass $10^{15}M_{\odot}$ (solid line), $10^{14}M_{\odot}$ (dotted line) and $10^{13}M_{\odot}$ (dashed line) at redshift $z = 0$ with clusters with mass larger than $M'/M$.

Middle panel: Same as above for a cluster of mass $10^{14}M_{\odot}$ at redshifts $z = 0$ (solid line), $z = 0.5$ (dotted line) and $z = 1$ (dashed line). Lower panel: normalized energy flux through the merger related shocks that contribute to the diffuse gamma ray emission above 100 MeV, for the same halos as in the upper panel.

In the second panel of figure 4.3 the energy flux is plotted for a cluster of fixed mass of $10^{14}M_{\odot}$ at three redshifts, $z = 0$ (solid line), $z = 0.5$ (dotted line) and $z = 1$ (dashed line). The same conclusions explained above hold here.

The third panel is the most interesting: it represents the normalized energy flux through the merger related shocks that contribute to the diffuse gamma ray emission above 100 MeV, for clusters of masses as labelled in the upper panel, at $z = 0$. It is immediately clear that most of the contribution to the gamma ray emission is provided by mergers with small mass ratios, $M'/M \leq 10^{-2}$, namely the ones having the largest Mach numbers (see Fig. 3.1).

Summarizing, while most of the energy flows through shocks associated to major mergers, the energy flux that contributes to the gamma ray background is the one
that crosses strong shocks, occurring when a large cluster encounters a subcluster with \( \sim 0.01 \) times the mass of the larger cluster.

Strong accretion shocks are not considered in the above discussion. They can accelerate particles with flat spectra that can appreciably contribute to the gamma ray emission. However, the accreted mass accounts only for \( \sim 10\% \) of the total mass of the cluster. Since the free fall velocity at the accretion shock is of the same order of magnitude of the relative velocity during a merger, we can conclude that only a small fraction of the total kinetic energy flows across these shocks.

### 4.5.1 Semi–analytical models

Totani & Kitayama (2000) and Loeb & Waxman (2000) developed a Press–Schechter based method similar to the one presented here in order to make predictions about the number of clusters detectable in gamma rays with energy above 100 MeV with future gamma ray telescopes. The authors claimed that at least a few tens of clusters should have been visible to EGRET. GLAST, on the other hand, was predicted to be able to detect more than a few thousands of such objects.

In the work by Totani & Kitayama (2000) all the shocks were assumed to be strong. As a consequence, the spectrum of the accelerated particles was always taken to be \( \propto E^{-2} \). As discussed above, this assumption leads to incorrect results if applied to merger shocks (Gabici & Blasi, 2003). Totani & Kitayama (2000) overestimated the gamma ray emission by orders of magnitude (despite the fact that the fraction of the energy crossing the shock which is converted to nonthermal electrons is assumed to be the same as here, namely 5%), and its spectrum does not reflect the real strength of the shocks developed during mergers of clusters of galaxies. This concept can be rephrased in another way: at fixed total energy flux crossing the merger shocks, and at fixed efficiency of electron acceleration, the nonthermal energy is distributed among accelerated particles in a different way if the spectrum of nonthermal electrons is taken to be \( E^{-2} \) rather than as the one derived from the correct evaluation of the shocks strength. The clusters that we predict to be detectable with AGILE and GLAST in gamma rays above 100 MeV all have flat spectra, which means that the gamma ray emission selects the few mergers that happen to have a flat spectrum of accelerated electrons (note that for Mach numbers in the range 1.5 – 4, the slope of the electron spectrum spans the range \( \alpha \sim 2 – 5 \)), in addition to the less luminous nearby accreting clusters that always generate flat spectra.

Similar arguments apply to the calculations by Waxman & Loeb (2000): the formation of a cluster was exemplified there as a spherical collapse, in which \( \sim 5\% \) of the kinetic energy crossing the accretion shock was converted into nonthermal electrons with spectrum \( E^{-2} \). From the energetic point of view, this approach is roughly equivalent to that by Totani & Kitayama (2000). However, as discussed above, only a relatively small fraction of the mass in a cluster is expected to be
accumulated through accretion, being mergers with other clusters the main physical process for mass build-up. As stressed above, the spectrum of the particles that the energy is channelled into is the key point in understanding whether one can expect appreciable gamma ray fluxes from clusters of galaxies.

The same considerations made above apply also for the estimate of the diffuse background presented by Loeb & Waxman (2000). According to their order of magnitude calculation, the whole extragalactic background can be explained by the inverse Compton scattering emission from electrons accelerated at strong shocks. Their assumption on shock strength explains part of the difference of more than one order of magnitude between their results and those presented here. Another ingredient introduced in our work and neglected by Loeb & Waxman (2000) is the redshift dependence of the gamma ray emissivity. Neglecting this dependence leads to an overestimate of the diffuse flux. These facts were also pointed out in a more recent work (Keshet et al., 2003), in which the background was reevaluated and there seems to be there a closer agreement with the conclusions of our calculations.

To conclude, our results are in perfect agreement with the recent evidence for the lack of an association between the unidentified high latitude EGRET sources and rich Abell clusters (Reimer et al., 2003). Moreover, Strong, Moskalenko & Reimer (2004) have recently re-evaluated the isotropic extragalactic gamma ray background, obtaining a different spectrum and a lower flux with respect to the earlier estimate by Sreekumar et al. (1998). All these observational facts are in disagreement with the semi-analytical models we have reviewed here.

4.5.2 Numerical simulations

Predictions of the gamma ray emission from large scale structure formation can also be done by means of numerical simulations.

In his simulations, Miniati (2002, 2003) considered the gamma ray emission from both protons and electrons accelerated at shocks. He adopted a thermal leakage recipe (see §6.4) to describe the injection of particles at shocks. In this approach, the fraction of thermal particles which starts to be accelerated is a fixed quantity, while the fraction of the shock thermal energy which is converted into cosmic rays is an increasing function of Mach number. For strong shocks, the efficiency can be as high as ~ 40% or more but, despite this fact, the spectra of the accelerated particles have been calculated in the test-particle approximation. Inverse Compton scattering from electrons, which is the dominant mechanism in the outer regions of clusters, was found to dominate the total contribution to the diffuse extragalactic emission. The fraction of the extragalactic background above 100MeV generated by cosmic ray accelerated at intergalactic shocks was found to be less than 30% and a flux level slightly higher than the EGRET sensitivity was predicted for the Coma cluster. However, a direct comparison with our results is difficult, because of the different assumptions made on the efficiency of the shocks in accelerating electrons.
4.5. Comparison with other theoretical results

Figure 4.4: Cumulative number of forming clusters of galaxies with flux above 100 MeV greater than $F$ (left panel). The solid (dashed) line refers to our prediction for accreting (merging) clusters. Our estimate of the contribution from accreting cluster to the extragalactic background is plotted in the right panel (solid line), together with the EGRET observational estimate (shaded region). Circles in both panels refer to the results from the simulation by Keshet et al. (2003). Figure from (Gabici, 2004).

and because of the higher number of strong shocks found in the simulation with respect to our semi–analytical approach (see §3.2.1 for a detailed discussion).

In the simulation by Keshet et al. (2003) the only considered mechanism for gamma ray production is inverse Compton scattering emission from electrons accelerated at shocks. Spectra of accelerated electrons are normalized assuming that a fraction equal to the 5% of the thermal shock energy is converted into relativistic electrons. Since these are the same assumptions we have made above, it is straightforward to compare their results with our findings. In figure 4.4 we plot our results superimposed with the results from the numerical simulation by Keshet et al. (2003) (open circles), finding that they are in very good agreement. The slope of the cumulative distribution of gamma ray bright clusters predicted by simulations is slightly steeper than the Press–Shechter based estimate (Keshet et al., 2003), but the predicted number of detectable objects is roughly the same in both approaches.
4.6 Clusters of galaxies as TeV sources

In §3.3 we showed that, within the assumption of Bohm diffusion, electrons can be diffusively accelerated at merger and accretion shocks, and their maximum energy can be as high as several tens of TeV. As a consequence, they can upscatter CMB photons up to TeV energies, if the magnetic field in the vicinity of the shock is in the range $0.1 - 1 \mu G$.

In this section we address the issue of detectability of the gamma rays with energy above 100 GeV from clusters of galaxies with ground based Cherenkov detectors. The calculations of the gamma ray emission for merging and accreting clusters have already been described in the previous sections. The only new ingredient that needs to be included here is the absorption of high energy gamma rays due to pair production in the universal photon background, in particular the infrared background (IRB). Pair production starts to be important when the energy in the center of mass of a photon–photon scattering equals twice the electron mass. The net result of the propagation of high energy gamma rays with energy in excess of a few hundred GeV is an absorption cutoff for sources more distant that the pathlength for pair production.

4.6.1 Absorption of extragalactic TeV photons by infrared photons

A very high energy gamma ray photon can interact with a soft photon in the CMB or in the infrared background. The interaction leads to an electron–positron pair production and for this reason photon spectra of extragalactic sources show a cutoff at high gamma ray energy (Gould & Schreder, 1966). Consider the interaction between a gamma–ray photon with energy $E(z) = E_0(1 + z)$ and a soft photon with energy $\epsilon(z) = \epsilon_0(1 + z)$. If $\theta$ is the angle between the directions of the two photons in the lab frame, pair production is expected above the threshold energy:

$$E_0\epsilon_0(1 + z)^2(1 - \cos \theta) > 2(mc^2)^2$$

with cross section given by (Jauch & Rohrlich, 1955):

$$\sigma[E(z), \epsilon(z), x] = 1.2510^{-25}(1 - \beta^2) \left[ 2\beta(\beta^2 - 2) + (3 - \beta^4) \ln \left( \frac{1 + \beta}{1 - \beta} \right) \right] cm^2$$

where $x = (1 - \cos \theta)$ and $\beta^2 = 1 - 2(mc^2)^2/[E_0\epsilon_0x(1 + z)^2]$. For a $\sim 1 TeV$ photon the cross section is maximized when the soft photon energy is equal to (Stecker, De Jager & Salamon, 1992):

$$\epsilon(E) \sim \frac{2(mc^2)^2}{E} \sim 0.5 \left( \frac{1 TeV}{E} \right) eV$$

70
which falls in the infrared band (~2μm). If the spectrum of the infrared background is described by \( n(e)de \), the optical depth for attenuation of a gamma–ray photon of present energy \( E_0 \) emitted by a source at a redshift \( z_s \) is (Stecker, 1999):

\[
\tau(E_0, z_s) = c \int_0^{z_s} \frac{dt}{dz} \int_0^{z_0} dx \frac{x}{2} \int_{\epsilon_1}^{\infty} d\epsilon_0 n[\epsilon_0(1+z)] \sigma(s)
\]

(4.23)

with \( \epsilon_1 = (2m^2c^4) / [E_0 x(1+z)^2] \) and \( s = 2E_0\epsilon x(1+z) \).

One of the main problem in evaluating \( \tau \) is that the infrared background is poorly measured in the energy range of interest. A fitting formula for the opacity was proposed by De Jager & Stecker (2002), who used an empirical model to estimate the extragalactic infrared background (Malkan & Stecker, 2001). Once the opacity is known, it is possible to calculate the attenuation factor for an object at redshift \( z \): this is simply given by \( \exp[-\tau(E, z)] \).

### 4.6.2 Detectability of clusters with Čerenkov telescopes

In figure 4.5 we plot the results of our calculations for the high energy gamma ray spectrum generated from a Coma-like cluster of galaxies at 100 Mpc distance for the case of merger and accretion. The effect of the gamma ray absorption in the infrared background is illustrated by the difference between the solid lines (with absorption) and dashed lines (without absorption).

From top to bottom, the lines refer to three different cases: 1) a merger between two clusters with masses \( 10^{15}M_\odot \) and \( 10^{13}M_\odot \); 2) an accreting cluster with mass \( 10^{15}M_\odot \) with a magnetic field at the shock in the upstream region \( 0.1\mu G \); 3) an accreting cluster with mass \( 10^{15}M_\odot \) with a magnetic field at the shock in the upstream region \( 0.01\mu G \).

The thick solid lines represent the sensitivities for a generic Cherenkov telescope as calculated in (Aharonian et al., 1997). These results are obtained considering an array of imaging atmospheric Cherenkov telescopes (IACT) consisting of \( n \) cells, each consisting of a 100 × 100 m² quadrangle with four ‘100 GeV’ class IACTs in its corners. The two thick curves in figure 4.5 represent the minimum detectable fluxes for point sources (lower curve) and an extended 1° wide source (upper curve) for an exposure time of 1000 hours. The exposure time here is defined as the product between the observation time and the number of cells that form the array. For instance, an exposure of 1000 hours can be achieved with a 100 hours observation performed by an array consisting of 10 cells. If the angular resolution of the Cherenkov experiment is taken to be 0.1 degrees, most clusters of galaxies, both merging and accreting, are diffuse sources for these experiments. The angular size of a cluster at distance \( D \) and with an emitting region of size \( R = 1 \) Mpc is \( \alpha \approx 1° / (D/100 \) Mpc). For accretion, the shock surface at which acceleration occurs is located approximately at the virial radius, that for a Coma-like cluster is \( \sim 3 \) Mpc, therefore the emission
Figure 4.5: Gamma ray emission in the 100 GeV - 10 TeV region. The thick solid lines represent the sensitivities of a IACT for point sources (lower curve) and extended sources (upper curve). The predicted gamma ray fluxes from a Coma-like cluster at a distance of 100 Mpc with and without absorption of the infrared background are plotted as dashed and solid lines respectively.

is even more extended than for a merger, and with angular size that approaches or even exceeds the aperture of the Cherenkov telescopes (a few degrees).

For the optimistic conditions that figure 4.5 refers to, the predicted flux of gamma rays between 100 GeV and a few hundred GeV for a merger between two clusters with mass ratio $10^{-2}$ is only slightly above the sensitivity of a Cherenkov telescope for a 1000 hours exposure. The predicted fluxes from accretion are unobservable. The sensitivities of the telescopes might be improved to some extent, compared with the curves in figure 4.5 by simulating off-center showers and different triggering modes of the telescopes in a cell. It appears however that the perspectives for detection of TeV gamma radiation from inverse Compton scattering of ultrarelativistic electrons in clusters of galaxies are not very promising.
Chapter 5

Turbulent Alfvénic reacceleration: protons and electrons

The origin of the non–thermal radio emission from clusters of galaxies is still an open problem. The difficulty in explaining the extended radio halos arises from the combination of their $\sim Mpc$ size and the relatively short lifetime of the radio emitting electrons. Indeed, the diffusion time necessary for the radio electrons to cover such distances is orders of magnitude larger than their radiative lifetimes.

To solve this problem, Jaffe (1977) proposed that relativistic electrons could be continuously re–energized through some in situ reacceleration mechanism. This possibility was studied more quantitatively by Schlickeiser et al. (1987) who successfully reproduced the integrated radio spectrum of the radio halo in the Coma cluster. In the framework of the in situ reacceleration model, Harris et al. (1980) first suggested that cluster mergers might provide the energetics necessary to reaccelerate the relativistic particles. An alternative scenario was put forward by Dennison (1980), who suggested that relativistic electrons may be produced in situ by inelastic proton–proton collisions through production and decay of charged pions. However, this model has serious difficulties in reproducing the fine radio properties of the radio halos (Brunetti, 2003).

Major merger events during the hierarchical assembly of a cluster are extremely energetic events. They can provide the total power required to form a radio halo if a small fraction of the gravitational energy is converted into relativistic electrons. This argument is also supported by the observational evidence for an association between merging system and radio halos (Buote, 2001). Merger related shock waves can accelerate particles via Fermi mechanism (see §2) but, as discussed in chapter 3, such shocks are typically weak and consequently unlikely to result in appreciable non–thermal activity (Gabici & Blasi, 2003; Berrington & Dermer, 2003).

Reacceleration of a population of relic electrons by turbulence powered by major mergers is suitable to explain the very large scale of the observed radio emission and is also a promising possibility to account for the fine radio structure of the diffuse
emission (Brunetti et al., 2001a,b). There are a number of possibilities to channel the energy of the turbulence in the acceleration of fast particles, namely via Magneto-Sonic waves, via magnetic Landau damping (Kulsrud & Ferrari, 1971), via Lower Hybrid waves (Eilek & Weatherall, 1999) or via Alfvén waves. Since Alfvén waves are likely to be able to transfer most of their energy into relativistic particles, they have received much attention in the last few years. In this framework for instance Ohno, Takizawa & Shibata (2002) developed a time-independent model for the acceleration of the relativistic electrons expected in radio halos through magnetic turbulence. The authors studied the acceleration of continuously injected relativistic electrons by Alfvén waves with a power law spectrum and applied this model to the case of the radio halo in the Coma cluster. More recently, Fujita, Takizawa & Sarazin (2003) studied the effect of Alfvénic acceleration of relativistic electrons in clusters of galaxies. These authors invoked the Lighthill theory to establish a connection between the large scale fluid turbulence and the radiated MHD waves. The electron and MHD-wave spectra adopted by Fujita, Takizawa & Sarazin (2003) are obtained via a self-similar approach by requiring that the spectra are described by two power laws.

These approaches have two intrinsic limitations: the first one is in the assumption, mentioned above, that all spectra are time-independent and that the turbulence spectrum is a power law. The second is that they neglect, as all other previous approaches did, the effect of relativistic hadrons in the intracluster medium: it is well known that the interaction of the Alfvén waves with relativistic particles is, in general, more effective for protons than for electrons (Eilek, 1979). It is also well known that the presence of a significant energy budget in the form of relativistic particles can significantly affect the spectrum of the Alfvén waves through damping. In fact, this damping occurs even on the thermal protons in the intracluster medium, another effect which was never included in previous calculations.

The calculations presented here provide a self-consistent time-dependent treatment of the non-linear coupling of Alfvén waves, relativistic electrons, thermal and relativistic protons. The results previously appeared in the literature can be obtained as special cases of our very general approach, which is in principle applicable to scenarios other than clusters of galaxies.

Throughout this chapter, numerical values are given assuming an Hubble constant $h = 0.5$.

5.1 Origin and spectrum of the relativistic particles in the intracluster medium

As seen in the previous chapters, relativistic particles can be injected in the intracluster medium through several different mechanisms. Cosmic ray sources in clusters of galaxies include large scale shocks, AGNs and normal/starburst galaxies...
5.1. Origin and spectrum of the relativistic particles in the intracluster medium

Figure 5.1: Electron spectrum at $z = 0$ injected at as a single burst at $z = 0.5, 0.3, 0.1, 0.01$ (left to right) adopting an injection spectrum $\propto p^{-2.5}$ and maximum Lorentz factor $10^4$. Typical values for the intracluster gas density and magnetic field strength are assumed.

(Berezinsky, Blasi & Ptuskin, 1997). Since most of these scenarios imply power law spectra of relativistic particles, in the following we will restrict our calculation to this case. It is worth recalling that transport effects and energy losses modify the shape of these spectra, that are not expected to be power laws at later times.

5.1.1 Electrons

Relativistic electrons with momentum $p = m_e c \gamma$ in the intracluster medium loose energy through Coulomb collisions (Sarazin, 1999):

$$\left( \frac{dp}{dt} \right)_i = -3.310^{-29} n_{th} \left[ 1 + \frac{\ln(\gamma/n_{th})}{75} \right]$$

(5.1)
and synchrotron and inverse Compton emission:

\[ \left( \frac{dp}{dt} \right)_{rad} = -4.810^{-4} p^2 \left( \frac{B_{\mu G}}{3.2} \right)^2 \frac{\sin^2 \theta}{2/3} + (1 + z)^4 \]  

(5.2)

where \( n_{th} \) is the number density of the thermal plasma, \( B_{\mu G} \) is the magnetic field strength and \( \theta \) is the pitch angle of the emitting electrons. Radiative losses dominate for Lorentz factors \( \gamma \gg 100 \) while Coulomb interactions are important at lower energies. Electrons with energies in the intermediate range \( \gamma = 100-500 \) accumulate for a few billions years in clusters before thermalize.

Figure 5.1 shows the spectra at \( z = 0 \) of electrons injected as a single burst at redshifts \( z = 0.5, 0.3, 0.1, 0.01 \) with an initial power law momentum distribution \( p^{-2.5} \) with maximum Lorentz factor \( \gamma = 10^4 \). Calculations have been carried out solving the equation:

\[ \frac{\partial N_e(p, t)}{\partial t} = \frac{\partial}{\partial p} \left[ \left( \frac{dp}{dt} \right)_{rad} + \left( \frac{dp}{dt} \right)_{i} N_e(p, t) \right] \]  

(5.3)

where typical values for the intracluster gas number density \( n_{th} = 10^{-3} \text{cm}^{-3} \) and magnetic field strength \( B = 1 \mu G \) are assumed. It is evident from figure 5.1 that most electrons injected at redshift \( z \gtrsim 0.2 \) get thermalized.

### 5.1.2 Protons

Relativistic protons are diffusively confined for cosmological times in the cluster volume and the energy loss time due to inelastic collisions with protons in the intracluster gas is greater than the Hubble time (see §3.1.1). As a consequence, they simply accumulate in clusters and their spectrum is unaffected by energy losses (Berezinsky, Blasi & Ptuskin, 1997; Völk, Aharonian & Breitschwerdt, 1996). On the contrary, the spectrum at mildly and sub–relativistic energies can be significantly modified by Coulomb interactions. In fact, protons which are more energetic than the thermal electrons, namely protons with velocity \( \beta > \beta_e = (3/2m_e/m_p)^{1/2}\beta_e \) (\( \beta_e \) here is the velocity of the thermal electrons, \( \beta_e \approx 0.18(T/10^8 K)^{1/2} \) lose energy according to (Schlickeiser, 2002):

\[ \frac{dp}{dt} \simeq -1.7 \times 10^{-29} \left( \frac{n_{th}}{10^{-3}} \right) \frac{\beta}{x_m^3 + \beta^3} \]  

(5.4)

where \( x_m = \left( \frac{3\sqrt{2}}{5} \right)^{1/3} \beta_e \).

The time evolution of the proton spectrum is described by the following equation:

\[ \frac{\partial N_p(p, t)}{\partial t} = \frac{\partial}{\partial p} \left[ \left( \frac{dp}{dt} \right) N_p(p, t) \right] + Q_p(p), \]  

(5.5)

76
5.2 Alfvenic reacceleration of relativistic particles

Figure 5.2: Present–epoch spectrum of the cosmic ray protons continuously injected (with $p > 0.1 m_p c$) in the intracluster medium starting from $z_i$. The spectra are plotted for $s = 2.4$ (dotted lines) and 3.0 (solid lines) and for $z_i = 0.1, 0.5, 1$ (from bottom to top). Calculations are carried out assuming $n_{th} = 10^{-3} cm^{-3}$.

Figure 5.2 shows the present day spectrum of protons if a time independent continuous injection $Q(p) \propto p^{-s}$ is assumed. Different values for the slope $s$ and for the redshift at which injection starts have been considered. Spectra show a large modifications at low energy due to Coulomb losses.

5.2 Alfvenic reacceleration of relativistic particles

Alfvén waves efficiently accelerate relativistic particles via resonant interaction. The condition for resonance between a wave of frequency $\omega$ and wavenumber projected along the magnetic field $k_\parallel$, and a particle of type $\alpha$ with energy $E_\alpha$ and projected velocity $v_\parallel = v_\mu$ is (Melrose, 1968; Eilek, 1979):

$$\omega - \nu \frac{\Omega_\alpha}{\gamma} - k_\parallel v_\parallel = 0$$

(5.6)
where, in the quasi parallel case \( k_\perp << m_\alpha \Omega_\alpha/p \), \( \nu = -1 \) and \( \nu = 1 \) for electrons and protons respectively.

The dispersion relation for Alfven waves in an isotropic plasma with both thermal and relativistic particles can be written as (Barnes & Scargle, 1973):

\[
\omega^2 = \frac{k_\parallel^2 v_A^2 (1 + \eta_1)}{1 + \frac{4}{3} \eta_2 + v_A^2 c^{-2} (1 + \eta_1)}
\]

(5.7)

where

\[
\eta_1 = \frac{n_{p,rel}}{n_{th}} + \frac{m_e}{m_p} \left( 1 + \frac{n_{p,rel}}{n_{th}} \right)
\]

(5.8)

and

\[
\eta_2 = \frac{n_{e,rel} m_e < \gamma_e > + n_{p,rel} m_p < \gamma_p >}{n_{th} m_p}
\]

(5.9)

Since the number density (and possibly the energy density as well) of the thermal component in the intracluster medium is considerably larger than the corresponding non-thermal component, one can show that \( \eta_1 << 1 \) and \( \eta_2 << 1 \) so that the dispersion relation becomes \( \omega \simeq |k_\parallel| v_A \). Combining the dispersion relation of the waves with the resonant condition, one can derive the resonant wavenumber, \( k_{res} \), for a given momentum \( p = m v \gamma \) and pitch angle cosine \( (\mu) \) of the particles:

\[
k_{res} \sim |k_\parallel| = \frac{\Omega_\alpha m}{p} \frac{1}{(1 \pm \frac{\mu}{v})},
\]

(5.10)

where the upper and lower signs refer to protons and electrons respectively. In an isotropic distribution of waves and particles, the particle diffusion coefficient in momentum space is given by (Eilek & Henriksen, 1984):

\[
D_{pp}(p, t) = \frac{2\pi^2 e^2 v_A^2}{c^3} \int_{k_{min}}^{k_{max}} W_k(t) \frac{k}{k^2} \left[ 1 - \left( \frac{v_A}{c} + \frac{\Omega m}{pk} \right)^2 \right] dk,
\]

(5.11)

where the minimum wavenumber (maximum scale length) of the waves interacting with particles is given by:

\[
k_{min} = \frac{\Omega_\alpha m}{p} \frac{1}{(1 \pm \frac{\mu}{v})}
\]

(5.12)

and \( k_{max} \) is given by the largest wavenumber of the Alfven waves, limited by the fact that the frequency of the waves cannot exceed the proton cyclotron frequency, namely \( \omega < \Omega_p \). It follows that \( k_{max} \sim \Omega_p/v_A \) or \( k_{max} \sim \Omega_p/v_M \), \( v_M \) being the magnetosonic velocity (here we assume \( k_{max} \sim \Omega_p/v_M \)).
5.2. Alfvenic reacceleration of relativistic particles

In the simple case of a power law spectrum of the MHD waves, \( W(k) \propto k^{-w} \) (for \( k_o < k < k_{max} \)), Equation (5.11) would give:

\[
D_{pp}(p) = A_w (\delta B)^2 v_A^2 \left( \frac{P}{B} \right)^w \left( 1 \pm \frac{v_A}{c} \right)^w \times \left\{ \frac{1}{w} \pm \frac{v_A/c}{w + 1} - \frac{(v_A/c)^2}{w(w + 1)} \right\}
\]

(5.13)

where

\[
\delta B^2 = 8\pi \int W_k dk,
\]

(5.14)

and

\[
A_w = \frac{\pi e^{2-w} w - 1}{2 e^{3-w} w + 2} k_o^{w-1}
\]

(5.15)

The first order expansion (for \( v_A/c << 1 \)) of equation (5.13) is the diffusion coefficient generally used in most recent theoretical papers on electron acceleration in galaxy clusters (Ohno, Takizawa & Shibata, 2002; Fujita, Takizawa & Sarazin, 2003).

From equation (5.10) one can see that the momentum of electrons or protons which can resonate with waves with a given wavenumber \( k \) depends on the pitch angle cosine \( \mu \).

The minimum momentum of the electrons for which resonance with waves of a given wavenumber \( k \) can occur is:

\[
p_{min} = m_e k \left( \Omega_e + v_A k \right),
\]

(5.16)

which, in the relativistic limit becomes:

\[
p_{\text{rel}}^{min} = \frac{\Omega_e m_e}{k} \frac{1}{1 - \frac{v_A}{c}}.
\]

(5.17)

Since the wavenumber of Alfven waves in a plasma is limited by \( \omega < \Omega_p \), from equation (5.16), one has that the minimum momentum of the electrons which can resonate with Alfven waves is:

\[
p_{min} = p_{th} \frac{v_A}{v_{th}} \left( \frac{m_p}{m_e} + 1 \right),
\]

(5.18)

which, in general, gives \( p_{min} >> p_{th} \), \( p_{th} = m_e v_{th} \) being the momentum of the thermal electrons. It follows the well known result that thermal electrons cannot resonate with Alfven waves (Hamilton & Petsosian, 1992).

This important limitation of Alfven waves as particle accelerators forces us to consider the situation in which a relic population of relativistic electrons exists in the intracell cluster medium, and no electron acceleration from the thermal background is taken into account.
The situation is different for protons. In fact, in this case, the minimum momentum which may resonate with waves having wavenumber \( k \) is:

\[
p_{\text{min}} = \frac{m_p}{k} (\Omega_p - v_A k).
\] (5.19)

which in the relativistic limit reduces to:

\[
p_{\text{min}}^{\text{rel}} = \frac{\Omega_p m_p}{k} \frac{1}{1 + \frac{v_A}{c}}.
\] (5.20)

From equation (5.19), one has that the minimum momentum of the protons which can resonate with Alfven waves in units of the momentum of the thermal protons is:

\[
p_{\text{min}} = p_{\text{th}} \frac{v_A}{v_{\text{th}}} \left( \frac{\Omega_p}{\omega} - 1 \right).
\] (5.21)

Since \( \omega < \Omega_p \), this basically means that thermal protons can efficiently resonate with Alfven waves (Hamilton & Petrosian, 1992).

### 5.2.1 From fluid turbulence to Alfven waves: the Lighthill mechanism

We assume that fluid turbulence is present in the cluster volume with a power spectrum

\[ W_f(x_f) = W^{\omega}_f x_f^{-m} \] (5.22)

in the range \( x_{f,\text{min}} < x_f < x_{f,\text{max}} \), where \( x_{f,\text{min}} \) is the wavenumber corresponding to the maximum scale of injection of the turbulence and the maximum wavenumber is that at which the effect of fluid viscosity starts to be important and it is of the order of \( x_{f,\text{max}} \sim x_{f,\text{min}} (R)^{-3/4} \) (Landau & Lifshitz, 1982), \( R \) being the Reynolds’ number.

For Kolmogorov turbulence we have \( m = 5/3 \), while for Kraichnan turbulence (Kraichnan, 1965) one has \( m = 3/2 \).

Here we investigate the connection between the fluid turbulence that we start with and the MHD waves that we use as particle accelerator. Fluid turbulence can radiate MHD modes (Kato, 1968) via the Lighthill process. A fluid eddy may be thought of as radiating MHD waves in the mode \( j \) at a wavenumber \( k = (v_j(x)/v_j) x_f \), where \( v_j \) is the velocity of the \( j \)-mode wave. The MHD modes are expected to be driven only for \( x > x_T \), \( x_T \) being the wavenumber at which the transition from large-scale ordered turbulence to small-scale disordered turbulence occurs. Following previous works in the literature (Eilek & Henriksen, 1984; Fujita, Takizawa & Sarazin, 2003), we adopt the Taylor wavenumber as an estimate of this transition scale, namely:

\[
l_T = \frac{2\pi}{x_T} \sim \left[ < v_{1,i}^2 > / \left( \frac{\partial v_{1,i}}{\partial x_i} \right)^2 > \right]^{1/2} \sim l_o (15/R)^{1/2},
\] (5.23)
where the Reynolds number is given by \( R = l_0 v_l / \nu_K \), and \( \nu_K \) is the kinetic viscosity. The fraction of the fluid turbulence radiated is small for all but the larger eddies, near the Taylor scale, and thus the Lighthill radiation can be expected to not disrupt the fluid spectrum. More specifically, the energy rate radiated via the Lighthill mechanism into waves of mode \( j \) and wavenumber \( k \) is given by (Eilek & Henriksen, 1984):

\[
I_j(k) = I_{j,0} \left( \frac{k}{x_T} \right)^{-y_j},
\]

where

\[
y_j = (2n_j + 3) \frac{m - 1}{3 - m}\]

\[
I_{j,0} \sim \frac{2n_j(1-m) + 6 - 4m}{3 - m} \rho v_j^3 \left( \frac{\mathcal{E}_i}{\rho v_j^2 R} \right)^{\frac{3+2n_j}{3-m}},
\]

\( \mathcal{E}_i \sim \rho v_j^2 \) is the energy density of the fluid turbulence, and

\[
R = \frac{x_0 W_f(x_0)}{x_T W_f(x_T)}
\]

Here, \( n_j = 0 \) for Alfvén and slow magnetosonic waves, and \( n_j = 1, 2 \) for fast magnetosonic waves. In the following we concentrate on Alfvén waves, in which case \( y_j = 3/2 \) (\( y_j = 1 \)) for a Kolmogorov (Kraichnan) spectrum of the fluid turbulence.

### 5.2.2 Basic equations and time evolution

In our calculations we assume for simplicity that Alfvén waves propagate isotropically in the cluster volume and we assume \( k \approx |k|| \). The spectrum of Alfvén waves driven by the fluid turbulence evolves as a result of wave–wave and wave–particle coupling. In particular, the wave–particle involves the thermal and relativistic particles, in the way explained in the previous section. The combination of these processes produces a modified, time–dependent spectrum of Alfvén waves, \( W_k(t) \), which can be calculated by solving the continuity equation (Eilek, 1979):

\[
\frac{\partial W_k(t)}{\partial t} = \frac{\partial}{\partial k} \left( D_{kk} \frac{\partial W_k(t)}{\partial k} \right) - \sum_{i=1}^{n} \Gamma^i_k W_k(t) + I_k(t).
\]

The first term on the right hand describes the wave–wave interaction, with diffusion coefficient \( D_{kk} = k^2 / \tau_s \). \( \tau_s \) is the spectral energy transfer time, that for a given wavelength is given by \( \tau_s \sim \tau_{NL}^2 / \tau_3 \) (Zohu & Matthaeus, 1990), where \( \tau_{NL} = \lambda / \delta v \) is the non–linear eddy–turnover time (\( \delta v \) is the rms velocity fluctuation at \( \lambda \)) and \( \tau_3 \) is the time over which this fluctuation interacts with other fluctuations of similar size.

In the framework of the Kolmogorov phenomenology, the Alfvén crossing time \( \tau_A = \lambda / v_A \) largely exceeds \( \tau_{NL} \), therefore fluctuations of comparable size interact in
5. Turbulent Alfvénic reacceleration: protons and electrons

one turnover time, namely $\tau_3 \sim \tau_{NL}$. In the Kraichnan phenomenology, $\tau_A < \tau_{NL}$, therefore convection limits the duration of an interaction and $\tau_3 \sim \tau_A$. Since the velocity fluctuation, $\delta v$, is related to the rms wave field, $\delta B$, through the relation $\delta v^2/v_A^2 = \delta B^2/B^2$, the diffusion coefficient is given by (Miller & Roberts, 1995):

$$D_{kk} \simeq v_A \left\{ \begin{array}{ll}
  k^{7/2} \left( \frac{W_k(t)}{2W_B} \right)^{1/2}, & \text{(Kolmogorov)} \\
  k^4 \left( \frac{W_k(t)}{2W_B} \right), & \text{(Kraichnan)}
\end{array} \right.,$$

(5.29)

in the Kolmogorov and Kraichnan phenomenology, respectively.

The second term in equation (5.28) describes the damping with the relativistic and thermal particles in the intracluster medium. In the case of nearly parallel wave propagation ($k_\perp << m\Omega/p, k \approx |k_\parallel|$) and isotropic distribution of particles of type $\alpha$, the cyclotron damping rate for Alfvén waves is as given by Melrose (1968):

$$\Gamma_\alpha^c(t) = -\frac{4\pi^3 e^2 v_A^2}{kc^2} \int_{p_{\min}}^{p_{\max}} p^2 (1 - \mu_\alpha^2) \frac{\partial f_\alpha(p,t)}{\partial p} dp = \frac{\pi^2 e^2 v_A^2}{kc^2} \int_{p_{\min}}^{p_{\max}} (1 - \mu_\alpha^2) \left( \frac{2N_\alpha(p,t)}{p} - \frac{\partial N_\alpha(p,t)}{\partial p} \right) dp,$$

(5.30)

where, for relativistic particles, one has:

$$\mu_\alpha^{rel} = \frac{v_A}{c} \pm \frac{\Omega_\alpha m_\alpha}{pk},$$

(5.31)

while for sub-relativistic particles:

$$\mu_\alpha^{th} = \frac{v_A m_\alpha}{p} \pm \frac{\Omega_\alpha m_\alpha}{pk}.$$

(5.32)

Here the upper and lower signs are for negative and positive charged particles respectively. Lacombe (1977) showed that the damping rate for isotropic Alfvén waves are well within a factor of $\sim 3$ of that calculated for nearly parallel wave propagation, therefore we are justified to use equation (5.30) in our calculations.

The third term in equation (5.28) describes the continuous injection of Alfvén waves as radiated by the fluid turbulence through the Lighthill mechanism. From equation (5.26) with $n_j = 0$ one has:

$$I_k \simeq 2 \left| \frac{3 - 2m}{3 - m} \right| \rho v_A^3 \left( \frac{v_A^2}{u_A^2 R} \right)^{\frac{3}{3 - m}} \times k^{-3 + \frac{m}{3 - m}}.$$

(5.33)

5.2.3 Turbulent cascade

The equation that describes the turbulent cascading without accounting for damping processes and wave injection is the following:

$$\frac{\partial W_k(t)}{\partial t} = \frac{\partial}{\partial k} \left( D_{kk} \frac{\partial W_k(t)}{\partial k} \right).$$

(5.34)
5.2. Alfvénic reacceleration of relativistic particles

Figure 5.3: **Left Panel**: Time evolution of the spectrum of Alfvén waves injected in a single burst at a given scale. The spectra are plotted for $10^{14}$ (solid line), $5 \times 10^{15}$ (dotted line), $3 \times 10^{16}$ (dashed line), and $3.2 \times 10^{16}$ s (dot-dashed line) after the injection event. In the calculations, a Kolmogorov diffusion coefficient is adopted. The temperature of the gas, the magnetic field and the gas density are $T = 10^8$ K, $B = 1 \mu G$, and $n_{th} = 10^{-4}$ cm$^{-3}$ respectively. **Right Panel**: Time evolution of the cascade-time scale estimated for the same parameters and with the corresponding line-styles as in the Left Panel.

The cascade timescale at a given wavelength is $\tau_{kk} \sim k^2/D_{kk}$, and using equation (5.29):

\[
\tau_{kk}(l) \simeq \frac{2 \times 10^8 \text{yr}}{B_{\mu G}} \left( \frac{l_{100}}{\text{kpc}} \right) \left( \frac{n_{th}}{10^{-3}} \right)^{\frac{1}{2}} \begin{cases} 
\sqrt{2} \left( \frac{\delta B_{>k}}{B} \right)^{-1} & \text{(Kolmogorov)} \\
2 \left( \frac{\delta B_{>k}}{B} \right)^{-2} & \text{(Kraichnan)}
\end{cases}
\]

(5.35)

where we define $\delta B_{>k} \sim \sqrt{8\pi kW_k}$. It is worth noticing that the cascade timescale in the Kolmogorov regime does not depend on the value of the magnetic field strength. We also notice that in both Kolmogorov and Kraichnan regimes, the cascade timescale depends on the scale of the waves and on the density of the thermal plasma. In particular, $\tau_{kk}$ is smaller in low density regions.

In figure 5.3 (left panel) we plot the evolution of the spectrum of a population of waves injected at a given scale as obtained from equation (5.34) in the Kolmogorov phenomenology (see caption for details); the broadening of the wave distribution at
scales larger than the injection scale clearly shows the effect of stochastic wave–wave diffusion. In qualitative agreement with equation (5.35), it is clear from figure 5.3 that the developing rate of wave-wave cascade from larger to smaller scales increases with decreasing scale. In figure 5.34 (right panel) we show the cascade time-scale calculated for the spectra (and corresponding times) used in the left panel. As a general remark, we find that for typical conditions of the intracluster medium, the wave–wave time scale below 1 pc, namely on the scale relevant for wave–particle interaction, is considerably shorter than $10^7$ yr.

### 5.2.4 Damping processes

Alfvén waves can be damped in their interaction with thermal and relativistic protons and relativistic electrons, so that the global damping time can be written as a sum of three different contributions:

$$
\tau_d = \left( \sum_{j=1}^{3} \Gamma_k^j \right)^{-1}.
$$

(5.36)

For typical conditions in the intracluster medium, the damping time on the thermal proton gas is $< 10^5$ sec (but the process is efficient only for $k/k_{\text{max}} > 0.1$). The damping time on the relativistic component (especially protons) is usually $> 10^8$ sec.

The time scale for the development of the wave-wave cascade depends on the wave-wave diffusion coefficient, $D_{kk}$, and thus on the energy density of the waves. Given a spectrum of injection of waves per unit time, $I_k$, one simple possibility to estimate the cascade time scale and thus to compare it with the time scale of the damping processes is to use the spectrum of the waves under stationary conditions and without damping processes, namely

$$
W_k \sim \frac{1}{k} \left\{ \left( \frac{B_k^2 L_k^4}{4\pi v_A^3} \right)^{1/3}, \text{ (Kolmogorov)} \right\}
$$

(5.37)

The wave-wave time scale is therefore given by:

$$
\tau_s = k^2 D_{kk}^{-1} \sim \frac{1}{k} \left\{ \left( \frac{B_k^2}{4\pi} \right)^{1/3} / \left( \frac{v_A^2}{3} I_k^{1/3} \right), \text{ (Kolmogorov)} \right\}
$$

(5.38)

A comparison between the time scales of the damping processes and of the wave-wave cascade is given in figure 5.4 for typical values of the parameters (see caption). We assume that the energy in relativistic protons and electrons is respectively $\sim 1$ and $0.1$ per cent of the total thermal energy. Protons are injected starting from
5.2. Alfvénic reacceleration of relativistic particles

Figure 5.4: Comparison between cascade (solid straight line) and damping time-scales on thermal gas (dashed line), relativistic protons (solid line) and relativistic electrons (dotted line). Calculations are carried out assuming a Kolmogorov diffusion coefficient, and adopting $T = 10^8 K$, $n_{th} = 10^{-3} cm^{-3}$ and $d(\delta B)^2/dt = 3.310^{-15} \mu G^2$.

$z = 1$ with a spectrum having slope $s = 2.2$. Figure 5.4 shows that the time scale due to the damping with the thermal pool is considerably shorter than the cascade time scale for $k/k_{max} >> 0.1$ so that a break or a cutoff in the spectrum of the waves is expected at large wavenumbers. However, the most important result illustrated in figure 5.4 is that, if a relatively large number of relativistic protons is present in the intracluster medium, the resulting damping time scale can become comparable with or shorter than the wave-wave cascade time scale. This means that, at the corresponding wavenumbers, the spectrum of the waves is modified by the effect of the dampings and therefore that a power law approximation for the spectrum of the MHD waves cannot be achieved. We also note that the effect of the damping due to relativistic protons is particularly evident at those wavenumbers that can exhibit a resonance with the bulk of the relativistic electrons in the intracluster medium (those with $\gamma \sim 200 - 1000$) and thus that this effect may have important consequences for the acceleration of the relativistic electrons.
In fact, as already mentioned, the damping of Alfven waves on the relativistic protons modifies the spectrum of the waves and therefore indirectly affects the acceleration of electrons. The damping of the waves at a given wavenumber basically depends on the number of protons with momentum that can resonate with such waves. At fixed number of relativistic protons with supposedly a power law spectrum $N(p) \propto p^{-s}$, the damping rate at wavenumbers corresponding to $p >> p_{\text{low}}$ ($p_{\text{low}}$ being the minimum momentum in the proton spectrum) decreases with increasing $s$.

The efficiency of the damping in galaxy clusters is further reduced in the case of steep spectra since, in this case, Coulomb losses affect the bulk of protons.

Figure 5.5: Time scale for damping of Alfven waves on relativistic protons. Different curves refer to proton injection spectra with slopes $s = 3.8, 3.2, 2.6, 2.2$ (from top to bottom).

These effects are illustrated in figure 5.5, where the cascade time-scale is compared with the damping time for different injection spectra containing the same number of protons.
5.3 Quasi stationary solutions

For a given model for the injection of the waves in the intracluster medium, the spectra of electrons, protons and waves can be calculated from the following equations (Brunetti et al., 2004):

\[
\begin{align*}
\frac{\partial N_e(p,t)}{\partial t} &= \frac{\partial}{\partial p} \left[ N_e(p,t) \left( \frac{dp}{dt} \bigg|_{\text{rad}} + \frac{dp}{dt} \bigg|_i - \frac{2}{p} D_{pp} \right) \right] + \frac{\partial}{\partial p} \left[ D_{pp} \frac{\partial N_e(p,t)}{\partial p} \right] \\
\frac{\partial N_p(p,t)}{\partial t} &= \frac{\partial}{\partial p} \left[ N_p(p,t) \left( \frac{dp}{dt} \bigg|_i - \frac{2}{p} D_{pp} \right) \right] + \frac{\partial}{\partial p} \left[ D_{pp} \frac{\partial N_p(p,t)}{\partial p} \right] + Q_p(p,t) \\
\frac{\partial W_k(t)}{\partial t} &= \frac{\partial}{\partial k} \left( D_{kk} \frac{\partial W_k(t)}{\partial k} \right) - \sum_{i=1}^{n} \Gamma_k^i W_k(t) + I_k(t)
\end{align*}
\]

The spectra of electrons, protons and waves, as discussed above, result from a coupling between all these components: the spectrum of the waves develops in time due to the turbulent cascade until damping becomes efficient and particle acceleration occurs. It is worth noticing that the time scales for the processes of damping and cascading are quite different from those related to particle losses and transport. While the wave spectrum develops over \( \sim 10^7 \) sec, particle acceleration occurs on time scales of \( \sim 10^{14} \) sec, and we are interested in following the particle evolution for a typical time of \( \sim 10^{15} \) sec.

The clear difference among these time scales suggests that we use a quasi stationary approach, in which it is assumed that at each time-step the spectrum of the waves approaches a stationary solution and that this solution changes with time due to the evolution of the spectrum of the accelerated electrons and protons. In other words, at any time-step we solve the set of three differential equations with \( \partial W/\partial t = 0 \). Intermittent injection of turbulence in the intracluster medium may occur on time-scales \( \geq 10^7 - 10^8 \) yrs which are much longer than the time-scales of damping and cascading, and thus the quasi–stationary approach discussed above remains applicable.

In the following we will confine our attention to the case in which the amount of energy injected in the form of turbulence is smaller than the thermal energy of the intracluster medium. As a consequence, we can safely assume that the thermal distribution of protons in the intracluster medium is not appreciably affected by the interaction with waves.

5.3.1 The spectrum of Alfvén Waves

The shape of the spectrum of waves at any time is determined by the damping of these waves, mainly on protons. The proton spectrum in turn changes because of acceleration, and backreacts upon the spectrum of waves: this implies that even for a time-independent rate of injection of waves, the strength of the damping rates and the spectrum of the MHD waves are expected to change with time.
Turbulent Alfvénic reacceleration: protons and electrons

Figure 5.6: Temporal evolution of the spectrum of the Alfvén waves at times $2 \times 10^{15}$, $5 \times 10^{15}$, $8 \times 10^{15}$, $10^{16}$, and $1.5 \times 10^{16}$ sec after the beginning of the acceleration (from top to bottom). The calculations are carried out for a Kolmogorov spectrum of the fluid turbulence (i.e., $y_j = 3/2$), a Kolmogorov diffusion coefficient $d(\delta B)^2/dt = 3.3 \times 10^{-15} (\mu G)^2/s$, $T = 10^8 K$, $n_{th} = 10^{-3} cm^{-3}$, $B = 0.5 \mu G$, $\mathcal{E}_e = 0.001 \times \mathcal{E}_{th}$, $\mathcal{E}_p = 0.005 \times \mathcal{E}_{th}$, $s = 3.2$, $z_i = 1.0$ and $p_{inj} > 0.1 m_p c$. The Taylor scale is at $k \sim 10^{-5} k_{max}$.

In particular, the damping rate due to relativistic protons increases with time, as a consequence of the fact that most of the energy injected in MHD waves is channelled into relativistic protons.

A relevant example of the time evolution of the spectrum of waves is illustrated in figure 5.6: as expected, the energy associated with MHD waves which contribute to the acceleration of the bulk of the relativistic electrons ($k/k_{max} \sim 10^{-1} - 10^{-3}$) decreases with time. In addition to the general finding that the spectrum of the Alfvén waves evolves with time, here we also point out that:

a) the spectrum is not a simple power law: it has a different slope at different $k$ and the curvature of the spectrum also changes with time.
5.3. Quasi stationary solutions

b) the spectrum has a low-\(k\) cutoff due to the maximum injection scale, close to the Taylor scale;

c) the spectrum has a high-\(k\) cutoff generated by the damping with the thermal particles.

5.3.2 Electron acceleration

The process of electron acceleration via Alfvén waves has been extensively investigated in the literature. Eilek & Henriksen (1984) showed that a self-similar solution for the spectra of electrons and waves can be found if these spectra are both required to be power laws in energy and wavenumber respectively. In this case, the slopes of the electron (\(\delta\)) and of the waves (\(\omega\)) spectrum are related by \(\delta = 6 - \omega\).

As pointed out above, the assumption of power law spectra is usually not fulfilled. In fact, more often a stationary solution is found in the form of a pile up spectrum \(N(p) \propto p^2 \exp\{-p/p_c\}\) (Borowsky & Eilek, 1986).

In the general case considered here, the electron spectrum may be even more complex due to the fact that the spectrum of the waves is not a power law and the whole evolution is time-dependent. Note that the initial stage of reacceleration of relic relativistic electrons (i.e. \(\gamma \sim 100 - 1000\) electrons) is mainly affected by the competition between Coulomb losses and acceleration due to the Alfvén waves, while later stages, of further acceleration to the highest allowed energies, are limited by radiative losses.

In figure 5.7 we plot the time scale for electron acceleration, \(\sim p^2/2D_{pp}\), and compare it with the time scale of the energy losses of electrons for different proton spectra injected in the intracluster medium (see caption). For steep proton spectra, the electron acceleration is more efficient because less energy gets channeled into the proton component. In general, hard proton spectra make the acceleration of electrons to Lorentz factors \(\gamma > 10^3\) relatively difficult.

If this is true in the initial stage of evolution of the system, it becomes increasingly less so at later times: after about \(\sim 0.5 - 0.7\) Gyr, relativistic protons have accumulated enough of the waves energy that the damping of the waves becomes even more efficient and further acceleration of electrons is prevented.

The continuous backreaction between waves and protons creates a sort of wave-proton boiler that in a way is self-regulated.

If the injection of fluid turbulence is intermittent on time scales of the order of the cooling time of electrons with Lorentz factors \(\gamma \sim 10^3 - 10^4\), then the effect of the wave-proton boiler on the electron acceleration may be reduced. The reason for this is that for a given reacceleration rate, the accumulation of energy in the form of relativistic protons requires longer times and the electron acceleration remains efficient for \(\sim 1\) Gyr.
5. Turbulent Alfvénic reacceleration: protons and electrons

Figure 5.7: Acceleration time-scale (thin lines) and time-scale for energy losses (thick lines) for relativistic electrons as a function of the Lorentz factor. From bottom to top, the acceleration time-scales are calculated at $2 \times 10^{15}$, $5 \times 10^{15}$, $10^{16}$, and $2 \times 10^{16}$ sec after the beginning of the acceleration. The following values of the parameters are adopted: $d(\delta B)/dt = 3.3 \times 10^{-15} (\mu G)^2/s$, $T = 10^8 K$, $n_{th} = 10^{-3} \text{cm}^{-3}$, $B = 0.5 \mu G$, and $\mathcal{E}_e = 0.001 \times \mathcal{E}_{th}$. **Left Panel:** $\mathcal{E}_p = 0.2 \times \mathcal{E}_{th}$, $s = 2.0$ and $z_i = 1.0$; **Central Panel:** $\mathcal{E}_p = 0.025 \times \mathcal{E}_{th}$, $s = 3.0$ and $z_i = 1.0$; **Right Panel:** $\mathcal{E}_p = 0.002 \times \mathcal{E}_{th}$, $s = 4.0$ and $z_i = 1.0$;

5.3.3 Proton acceleration

Alfvénic acceleration of thermal and relativistic protons is extensively studied in the literature, in particular as applied to the case of solar flares (Miller & Roberts, 1995; Miller, Guessoum & Ramaty, 1990).

We consider here the extension of these calculations to the case of Alfvénic acceleration in clusters of galaxies.

If the energy injected in Alfvén waves is significantly larger than that stored by the relativistic protons at the beginning of the acceleration phase, then the spectrum of protons is expected to be considerably modified. Under these conditions, we illustrate, in 5.8, the evolution of the spectrum of the relativistic protons. The figure clearly shows that the spectrum flattens and develops a bump.

The prominence of this bump increases with time as the energy absorbed by relativistic protons also increases. Moreover, the bump moves toward larger momenta of the particles during the acceleration time.

The presence of a bump in the proton spectrum may be of some importance in the calculation of the spectrum of the high energy secondary electrons which are expected to be produced during hadronic collisions. We find that, under typical conditions in the intracluster medium and assuming an energetics of the Alfvén waves considerably larger than that of the initial proton population, relevant bumps are produced at energies in excess of 200 GeV which should be visible in the spectrum.
5.3. Quasi stationary solutions

Figure 5.8: Time evolution of the spectrum of cosmic ray protons as a function of $p$. From bottom to top, the curves are obtained at times $2 \times 10^{15}$, $5 \times 10^{15}$, $8 \times 10^{15}$, $10^{16}$, and $1.5 \times 10^{16}$ sec after the beginning of acceleration. The initial proton spectrum is plotted as a dashed line. The values of the parameters are as in figure 5.6.

of secondary electrons at $\gamma > 10^5$. We also point out that proton acceleration is strongly reduced at some momentum $p_{\text{max}}$ (see figure 5.8) which corresponds to the momentum at which the resonance condition is satisfied for wavenumber corresponding to the maximum injection scale of the Alfvén waves. A maximum energy of the injected secondary electrons is expected as well.

5.3.4 The Wave–Proton Boiler

One of the most important results of our investigation is the quantitative treatment of the backreaction of the accelerated protons on the waves and in turn on the electrons. Qualitatively, given typical conditions in the intracluster medium, we can identify three main temporal stages of the acceleration process:

1) Cascading stage:
5. Turbulent Alfvénic reacceleration: protons and electrons

Figure 5.9: Evolution of the proton acceleration time-scale during the acceleration phase as a function of $p$. From bottom to top the curves refer to $2 \times 10^{15}$, $5 \times 10^{15}$, $10^{16}$, and $2 \times 10^{16}$ sec after the beginning of the acceleration stage. The following values of the parameters have been adopted: $d(\delta B)^2/dt = 3.3 \times 10^{-15}(\mu G)^2/2$, $B = 0.5\mu G$, $T = 10^8 K$, $n_{th} = 10^{-3} cm^{-3}$, $E_e = 0.001 \times E_{th}$, $E_p = 0.0025 \times E_{th}$ and $s = 3.0$.

For a non negligible rate of energy injection in the form of Alfvén waves, the cascade time is shorter than the damping time. This remains true up to some critical wavenumber, which depends on energetics and spectrum of protons, where damping starts to be relevant. If such a wavenumber is larger than about $10^{-2} k_{max}$, then enough energy is left in the form of waves at the scales which may resonate with relic relativistic electrons. In this case electrons are effectively re-energized.

2) Stage of proton backreaction:

Once the Alfvén waves start to accelerate electrons and protons to higher energies, the spectrum of protons and electrons becomes harder and the fraction of the energy stored in non-thermal particles starts to be large enough to make damping more severe. As a consequence, the rate of electron acceleration is
reduced.

3) *End of acceleration:*

At the beginning of the acceleration phase, the bulk of protons is located at supra-thermal or trans-relativistic energies. It takes a few $10^8$ yrs, however, for these protons to be energized to higher energies, as illustrated in figure 5.9, where we plot the acceleration time scale of relativistic protons (we chose a relatively steep injection spectrum). From figure 5.9 we see that after about 0.5 – 0.7 Gyr the acceleration time scale has increased by about one order of magnitude. At this point the acceleration stage of protons and electrons can be considered as concluded, unless the injection of turbulence occurs intermittently.

After the end of the third stage, the electrons cool due to radiative and Coulomb losses, while the Alfvén acceleration is only able to prevent the thermalization of these particles maintaining their Lorentz factor around $\gamma \sim 100 - 1000$.

### 5.4 Non-thermal emission from galaxy clusters

#### 5.4.1 Cluster mergers and turbulence

Major mergers are among the most energetic events in the Universe. Cluster mergers involve a collision of at least two subclusters with relative velocity of $\sim 1000 – 2000$ km s$^{-1}$. During these events, a gravitational energy in excess of $10^{63}$ erg is released through the formation of shock waves. Numerical simulations show that cluster mergers can generate relatively strong turbulence in the intracluster medium (Ricker & Sarazin, 2001). The bulk of the turbulence is most likely injected on scales $\geq 100$ kpc during the motion of the subclusters. Afterwards this turbulence eventually cascades toward smaller scales. As discussed in §5.2.1, when the turbulent cascade reaches scales close to the Taylor scale, a fraction of the energy flux of the fluid turbulence can be transferred to MHD waves which in turn can accelerate fast particles.

For simplicity, we assume here that the bulk of the fluid turbulence in a given point of the cluster volume is injected at the scale $l_\alpha$, for a time $\tau_i$, of the order of the time necessary for the subclump to cross the scale $l_\alpha$:

$$\tau_i(\text{Gyr}) \sim 0.3\xi\left(\frac{l_\alpha}{300}\right)\left(\frac{v_c}{10^3}\right)^{-1},$$

where $v_c$ is the velocity of the subclump in the host cluster and $\xi$ is a parameter of the order of a few. Within these assumptions, the injection rate of energy in the form of fluid turbulence is given by:

$$F_f \sim \frac{2.3 \times 10^{-27}}{\xi}\left(\frac{n_{th}}{10^{-3}}\right)\left(\frac{T}{10^8}\right)^{-1}\left(\frac{l_\alpha}{300}\right)\left(\frac{v_c}{10^3}\right)\xi_{th},$$

(5.40)
5. Turbulent Alfvenic reacceleration: protons and electrons

where $\mathcal{E}_{th}$ is the local energy density of the intracluster medium in the form of thermal gas and $\mathcal{E}_t$ is that in the form of turbulence. The bulk of the fluid turbulence at the scale $l_o$ then cascades toward smaller scales producing a spectrum of the fluid turbulence that we write as $W_f(x) \propto x^{-m}$. Assuming a Kolmogorov phenomenology for the wave–wave diffusion in $k$–space, the time scale for the cascade can be estimated from equation (5.38) with $I_{x_o} \sim x_o^{-1} F_f$:

$$\tau_s(\text{Gyr}) \sim 0.2 \left( \frac{l_o}{300} \right) \left( \frac{10^8}{T} \right)^{1/3} \left( \frac{10^3}{v_c} \right)^{1/3} \xi^{1/3} \left( \frac{0.1 \mathcal{E}_t}{\mathcal{E}_{th}} \right)^{1/3}. \quad (5.41)$$

In our simple approach, this is the time delay between the merger event and the development of the turbulence at small scales (and thus the production of MHD waves). In the case of Alfven waves in the framework of a Kolmogorov phenomenology, the power injected in waves is (equation 5.33):

$$P_A = \int I(k) dk \simeq \frac{1.5 \times 10^{-29}}{R_{16}^{1/2}} \left( \frac{v_f}{400} \right)^4 B_{\mu G}^{-1} \left( \frac{n_{th}}{10^{-3}} \right)^2 \left( \frac{l_o}{300} \right)^{-1}. \quad (5.42)$$

All the quantities involved in the calculation of the power radiated in the form of Alfven waves can be relatively well modelled. The only parameter which is very difficult to estimate is the value of the Reynolds number, $R = l_o v_f / \nu_K$, due to the large uncertainties in the value of the kinetic viscosity $\nu_K = u_p \lambda_{eff} / 3$ ($u_p$ is the thermal velocity and $\lambda_{eff}$ the effective mean free path of protons). For transverse drift of protons in a magnetic field $B$, the mean free path is given by $\lambda_{eff} \sim \lambda_p^2 / \lambda_c$ (Spitzer, 1962) with $\lambda_p$ and $\lambda_c$ being the proton gyroradius and the mean free path due to Coulomb collisions respectively. In this case the kinetic viscosity is given by (Fujita, Takizawa & Sarazin, 2003):

$$\nu_K = 1.3 \times 10^5 \left( \frac{n_{th}}{10^{-3}} \right) \left( \frac{T}{10^8} \right)^{-1/2} \left( \frac{\ln \Lambda}{40} \right) B_{\mu G}^{-2}. \quad (5.43)$$

The resulting Reynolds number is:

$$R \sim 2.8 \times 10^{26} \left( \frac{l_o}{300} \right) \left( \frac{v_f}{400} \right) \left( \frac{10^{-3}}{n_{th}} \right) \left( \frac{T}{10^8} \right)^{1/2} B_{\mu G}^2 \left( \frac{40}{\ln \Lambda} \right), \quad (5.44)$$

which is extremely large and, most likely, should be considered as an upper limit, due to the assumption of transverse drift of protons on the magnetic field lines. In general, diffusion of the protons along the magnetic field lines can substantially increase the value of $\lambda_{eff}$ and thus reduce $R$. The Reynolds number has been roughly estimated in a number of astrophysical situations, being $\sim 10^7$ in the solar wind (Grappin et al., 1982), $\sim 10^{11}$ in the extragalactic radio jets imaged with the VLA (Henriksen, Bridle & Chan, 1982), $\sim 10^{14}$ in the solar corona (Ofman &
5.4. Non-thermal emission from galaxy clusters

Aschwanden (2002) and $10^{16}$ for the hot phase of the local interstellar medium (Armstrong, Rickett & Cordes, 1981). Following this last estimate, in our modelling we take $R \sim 10^{16}$ but also stress that, due to the poor dependence of $P_A$ on $R$, an uncertainty in $R$ by six orders of magnitude implies only one order of magnitude change in $P_A$.

In our simple model for the merger, it is easy to show that the condition $P_A \ll F_f$ is satisfied, therefore the application of the Lightill scheme in our calculations appears to be self-consistent.

The power radiated in the form of Alfvén waves strongly depends on the velocity of the eddies in the fluid turbulence. In addition, assuming that the energy of the fluid turbulence is a fraction of the local thermal energy (i.e., $E_t \sim \rho v_f^2$, and $v_f = \text{const}$) and that the largest scale of the spectrum of turbulence, $l_o$, does not depend on the position in the cluster volume, we notice that the injected power in Alfvén waves increases with increasing the number density of the intracluster medium and with decreasing the strength of the local magnetic field. For instance, assuming a typical scaling law for the magnetic field in the cluster $B \propto n_{th}^{2/3}$, it would be $P_A \propto n_{th}^{5/6}$, namely the power injected in the form of Alfvén waves is expected to be slightly larger in the central regions of the cluster.

A second important quantity in our calculations is the largest scale of the spectrum of the Alfvén waves, $k_T \sim x_T v_f(x_T)/v_A$. It is convenient to parametrize this quantity in terms of the wavenumber corresponding to the minimum scale of the MHD waves, $k_{\max} \sim \Omega_p/v_M$, namely:

$$\frac{k_T}{k_{\max}} = \frac{R_{16}^{1/3} (n_{th}/10^{-3})^{1/2} (\frac{300}{l_o}) B_{\mu G} (\frac{v_f}{100}) (\frac{T}{10^8})^{1/2}}{2.4 \cdot 10^7}.$$ (5.45)

Given the resonance condition (equations 5.17 and 5.20), the presence of a maximum scale in the wave spectrum, $l_T \sim 2\pi/k_T$, implies a limit for the energy of the particles that can be efficiently accelerated:

$$E_{res}(\text{TeV}) \leq 24 \times \left( \frac{k_{\max}/k_T}{6 \cdot 10^6} \right) \left( \frac{T}{10^8} \right)^{1/2} \times R^{-1/3} n_{th}^{-1/2} l_o B^2 v_f^{-1}. \quad (5.46)$$

This limit provides a relatively good estimate in the case of relativistic protons, which are basically loss-free particles, while the maximum energy of the electrons is driven by the competition between acceleration and loss terms.

5.4.2 Constraining the model parameters

In this Section we derive some constraints on the physical conditions in the intracluster medium in order to obtain the reacceleration efficiency necessary to allow the production of the observed non-thermal emission.
From equation (5.42) it is clear that a key parameter in the calculation of the efficiency of particle acceleration in our approach is given by:

\[
\Psi = \left( \frac{v_f}{400} \right)^4 \left( \frac{R}{10^{16}} \right)^{-1/6} \left( \frac{L_o}{300} \right)^{-1},
\]

(5.47)

which is very sensitive to the velocity of the eddies in the fluid turbulence. A constraint on this velocity can be obtained by requiring that the energy injected in the form of fluid turbulence \((E_t \sim x_0 W(x_0) \sim m_p n_{th} v_f^2)\) does not exceed the thermal energy. From equations (5.39) and (5.40) we obtain:

\[
\left( \frac{v_f}{400} \right) \leq 3 \left( \frac{T}{10^8} \right)^{1/2} \left( \frac{E_t}{E_{th}} \right)^{1/2}.
\]

(5.48)

An additional constraint to the parameter \(\Psi\) comes from requiring that the energy stored in the form of relativistic protons is small enough to allow for efficient electron acceleration.

As pointed out above, after an acceleration time of the order of 0.5-0.7 Gyr, relativistic protons get a large fraction of the energy previously in the form of Alfvén waves. We limit ourselves to cases in which \(P_A \cdot \tau_i < 0.1 \varepsilon_{th}\). From equations (5.39) and (5.42) one has:

\[
\Psi \leq \frac{16}{\xi} B \mu G \left( \frac{n_{th}}{10^{-3}} \right)^{-\frac{1}{2}} \left( \frac{T}{10^8} \right) \left( \frac{L_o}{300} \right)^{-1}.
\]

(5.49)

Finally, we are interested in the production of relatively long-living (e.g., for \(> 0.3\) Gyr) non-thermal phenomena, in order to have a chance of observing them in some clusters. From equation (5.39) we can set a limit on the parameter \(\xi\):

\[
\xi > \left( \frac{v_c}{1000} \right) \left( \frac{L_o}{300} \right)^{-1}.
\]

(5.50)

The maximum energy, \(\gamma_{max}\), of the accelerated electrons is obtained by balancing energy losses and energy gains.

In figure 5.10 (left panel) we plot \(\gamma_{max}\) as a function of \(\Psi\), for different acceleration times, \(\Delta T\), for a given set of values of the parameters defining the environment of cluster cores. In the right panel we plot the same quantity for different values of the energy density in the form of relativistic protons, \(E_p\). The injection spectrum of protons is taken as a power law with slope 2.2.

In order to obtain \(\gamma_{max} >> 1000\), needed to explain the synchrotron emission at GHz frequency as well as the inverse Compton hard X-ray photons, we are forced to require that \(\Psi \geq 1\) and \(E_p \leq 3\% E_{th}\).

Such a stringent limit is the consequence of the effective damping of Alfvén waves upon the relativistic proton component, which inhibits the acceleration of electrons.

In the periphery of the cluster, where the magnetic field is expected to be lower, the conditions to obtain high energy electrons are less stringent. It remains true
5.4. Non–thermal emission from galaxy clusters

Figure 5.10: Maximum energy attainable for relativistic electrons during the acceleration (only systematic terms are considered here) plotted as a function of $\Psi$. **Left Panel:** $\gamma_{\text{max}}$ at $3 \times 10^{15}$ (dashed line) and $10^{16}$ sec (solid line) after the beginning of the acceleration stage. The values of the parameters are: $B = 1 \mu G$, $n_{\text{th}} = 10^{-3} \text{cm}^{-3}$, $E_e = 0.001 \times E_{\text{th}}$, $E_p = 0.003 \times E_{\text{th}}$, $s = 2.2$, and $z_i = 1.0$. **Right Panel:** $\gamma_{\text{max}}$ at $10^{16}$ sec from the beginning of the acceleration stage for $E_p = 0.003 \times E_{\text{th}}$ (solid line) and $E_p = 0.03 \times E_{\text{th}}$ (dashed line). Other parameters are as in the Left Panel case.

however that no more than a few percent of the thermal energy of the cluster can be in the form of relativistic protons with $p > 0.1 m_e c$ if we want to interpret the observed non–thermal phenomena as the result of radiative processes of high energy electrons accelerated via Alfvén waves.

A steeper injection spectrum of protons, containing the same energy density, makes the constraints found above even more stringent; however this would imply a very large energy injected in relativistic protons in the intracluster medium.

On the other hand, at given proton number density, a steeper spectrum contains less relativistic particles, which would allow for more efficient electron acceleration.

Clearly, a crucial parameter in the modeling of the non–thermal phenomena in galaxy clusters is the strength of the magnetic field in the intracluster medium. There is still debate on whether this field is of several $\mu G$ or rather fractions of $\mu G$: on one hand, if the hard X–ray excess is interpreted as a result of inverse Compton emission of relativistic electrons, the required volume averaged field is $\leq 0.2 - 0.4 \mu G$ (Fusco–Femiano et al., 1999; Fusco–Femiano , 2000). On the other hand, Faraday rotation measurements of the radiation coming from cluster radio sources seem to require a magnetic field strength of $\sim 4 - 8 \mu G$ (Clarke, Kronberg & Böringer , 2001). A number of possibilities to reconcile these two predictions have been proposed in the literature, based on either choosing more realistic spectra of
Figure 5.11: Synchrotron cut-off frequency as a function of the magnetic field strength in the case of $n_{th} = 10^{-3} \text{cm}^{-3}$ (empty symbols) and $n_{th} = 2 \times 10^{-4} \text{cm}^{-3}$ (filled symbols). The results are reported after $7 \times 10^{15} \text{sec}$ of reacceleration. In the calculations the following values of the parameters have been used: $d(\delta B)^2/dt = 3.3 \times 10^{-15} (\mu \text{G})^2/\text{s}$, $T = 10^8 \text{K}$, $\mathcal{E}_e = 0.001 \times \mathcal{E}_{th}$, $\mathcal{E}_p = 0.01 \times \mathcal{E}_{th}$, $s = 2.2$ and $z_i = 1.0$.

It is also worth recalling that Faraday rotation measures do not provide a direct measurement of the magnetic field, but rather an estimate of a quantity which is a (not necessarily trivial) convolution of the component of the magnetic field parallel to the line of sight, of the electron density and that accounts for the topology of the magnetic field: Newman, Newman & Rephaeli (2002) showed how the assumption of a single-scale magnetic field leads to an overestimate of the field strength extracted from rotation measures. Other authors have further discussed the influence of the power spectrum of the magnetic field topology in the intracluster medium on the rotation measures (Ensslin & Vogt, 2003; Vogt & Ensslin, 2003; Govoni et al., 2003).
5.4. Non–thermal emission from galaxy clusters

2003).

In order to illustrate the effect of these uncertainties on the conclusions inferred from our calculations, we evaluated the synchrotron cut–off frequency for a cluster with temperature $T = 10^8$ K and with a relic electron and proton energy densities chosen as $\mathcal{E}_e = 0.001 \times \mathcal{E}_\text{th}$, $\mathcal{E}_p = 0.01 \times \mathcal{E}_\text{th}$ (protons have a spectrum with power index $s = 2.2$; see caption of figure 5.11 for additional information). The calculation is carried out for a dense region with $n_\text{th} = 10^{-3} \text{cm}^{-3}$ (empty symbols) and for a low density region with $n_\text{th} = 2 \times 10^{-4} \text{cm}^{-3}$ (filled symbols). Given the shape of the spectrum of the accelerated electrons, a synchrotron cut–off at $\geq 300$ MHz is required to account for the synchrotron radiation observed in the form of radio halos.

Our conclusion, based on figure 5.11, is that in high density regions ($n_\text{th} \sim 10^{-3} \text{cm}^{-3}$) Alfvénic reacceleration of relic electrons cannot be an efficient process for $B \gg 4 \mu$G and for $B \ll 0.5 \mu$G. These constraints become less stringent in the case of low density regions.

5.4.3 A simplified models for Radio Halos and Hard X–ray emission

The best evidence for the diffuse non–thermal activity in clusters of galaxies is provided by the extended synchrotron radio emission observed in about $\sim 30$ massive clusters of galaxies (Feretti , 2003). A recent additional evidence supporting the existence of relativistic electrons is given by the hard X–ray tails in excess to the thermal emission discovered by BeppoSAX and RXTE in the case of a few galaxy clusters (Fusco–Femiano et al., 1999; Fusco–Femiano , 2000). The possibility that hard X–ray tails are due to inverse Compton scattering of the CMB photons is intriguing as, in this case, radio and hard X–ray radiations would be emitted by roughly the same electron population and thus the combination of radio and hard X–ray data would allow us to infer an estimate of the volume–averaged magnetic field strength and of the energy density of relativistic electrons. Additional pieces of evidence for non–thermal phenomena are the so called radio Relics and the EUV excesses whose origin may however be not directly connected to that of radio halos and hard X–ray emission. Therefore these phenomena will not be considered here. For a recent review on these arguments the reader is referred to the works by Ensslin (2003); Bowyer (2003) and references therein.

In this section we apply the formalism described in previous sections in order to show that for the conditions realized in the intracluster medium, Alfvénic reacceleration of relic electrons may generate the observed radiation, provided the energy content in the form of relativistic protons is not too large. For simplicity we neglect here the production and reacceleration of secondary products of proton interactions. Our simple model for the intracluster medium assumes a $\beta$–model (Cavaliere & Fusco–Femiano , 1976) for the radial density profile of the thermal gas in the
Figure 5.12: **Temporal evolution of the accelerated electron spectra after 0, 10^{14}, ..., 5 \times 10^{15}, 7 \times 10^{15}, 10^{16}, and 1.2 \times 10^{16} \text{sec from the beginning of the acceleration stage. The following values of the parameters have been used: } d(\delta B)^2/dt = 1.6 \times 10^{-15}(\mu G)^2/s, B = 0.5\mu G, T = 10^8 K, n_{th} = 10^{-3} \text{cm}^{-3}, E_e = 0.001 \times E_{th}, E_p = 0.01 \times E_{th}, s = 2.2 \text{ and } z_i = 1.0.**

Intracluster medium, in the form

\[ n_{th}(r) = n_{th}(r = 0)\left(1 + \left(\frac{r}{r_c}\right)^2\right)^{-3\beta/2}, \tag{5.51} \]

where \(r_c\) is the core radius and we adopted \(\beta = 0.8\). The magnetic field is assumed to scale with density according with flux conservation:

\[ B(r) = B(r = 0)\left(\frac{n_{th}(r)}{n_{th}(r = 0)}\right)^{2/3}. \tag{5.52} \]

Based on the constraints given in §5.4.2 we adopt \(B(r = 0) \sim 0.5 - 4\mu G\).

Finally, we assume that the ratio between the energy density of the relic relativistic particles (at the beginning of the acceleration phase) and that of the thermal plasma is constant with distance throughout the cluster volume:

\[ E_{p[e]} = E_{th} \eta_{p[e]}, \tag{5.53} \]
5.4. Non-thermal emission from galaxy clusters

Figure 5.13: Spectra of electrons accelerated for \(1.2 \times 10^{16}\) sec at different distances from the cluster center: \(r = 0.3, 0.6, 0.9, 1.2, 1.4, 2.1 r_c\) (from top of the diagram). The central values assumed in the calculations are: \(n_{th}(0) = 1.5 \times 10^{-3} \text{cm}^{-3}\), \(B(0) = 1.5 \mu \text{G}\), \(d(\delta B(0))^2/dt = 2.2 \times 10^{-15} (\mu \text{G})^2/\text{s}\) with the scaling laws given in Sect. 6.3, \(r_c = 400 \text{ kpc}\).

where \(\eta_{p[e]}\) is a free parameter (\(\eta < 0.1\)).

For simplicity we also assume that the maximum injection scale of the turbulence, the Reynolds number and the velocity of the turbulent eddies are independent of the location within the cluster volume.

Using the scaling relationship for the magnetic field, equation (5.52), and the expression for the injection power in the form of Alfvén waves, equation (5.42), we obtain:

\[
P_A(r) = P_A(r = 0) \left( \frac{n_{th}(r)}{n_{th}(r = 0)} \right)^{5/6}.
\]

(5.54)

The spectrum of the relativistic electrons in the core region is plotted in figure 5.12. We can see that the bulk of relativistic electrons, initially at \(\gamma \sim 10^2\), can be energized up to \(\gamma \sim 10^4\) for a relatively long time.

Equation (5.54) indicates that, in our simple approach, the power injected in
the form of Alfven waves decreases with increasing distance from the cluster center. Since radio halos have a considerable size, it is needed to check that our model provides enough energy in the outskirts of clusters.

In figure 5.13 we plot the electron spectra at different distances from the cluster center (see caption). What emerges from this figure is that at large distances the effect of acceleration is even stronger than in the central region and the electron spectra peak at slightly higher energies than in the core. This is due to the fact that in the outskirts the damping rate is reduced more than the rate of injection of turbulence.

The synchrotron emissivity roughly scales as $J_{\text{syn}} \propto N_e \gamma_{\text{max}}^2 B^2 \propto n_{\text{th}}^{7/3} \gamma_{\text{max}}^2$. Such a relatively soft dependence of the emissivity on the radius allows for a relatively broad synchrotron brightness profile.

More specifically, the synchrotron emissivity can be written as:

$$J_{\text{syn}}(\nu,t) = \frac{\sqrt{3}e^3 B}{mc^2} \int_\nu d\nu \int_0^\pi dp \sin^2 \theta N(p,t) F\left(\frac{\nu}{\nu_c}\right), \quad (5.55)$$

where the synchrotron Kernel is given by (Rybik & Lightman, 1979):

$$F\left(\frac{\nu}{\nu_c}\right) = \frac{\nu}{\nu_c} \int_{\nu/\nu_c}^{\infty} K_\frac{3}{2}(y)dy, \quad (5.56)$$
and \( \nu_c = (3/4\pi)p^2eB \sin \theta/(mc)^3 \).

The emissivity due to inverse Compton scattering off the CMB photons is given by (Blumenthal & Gould, 1970):

\[
J_{ic}(\nu_1, t) = \left( 2\pi r_o m^2 c^2 \right) ^2 h\nu_1^2 \int d\nu \int dp \frac{N(p, t)p^{-4}}{\exp(\frac{h\nu}{k_B T_c}) - 1} \times \\
\left[ 1 + 2 \ln \left( \frac{\nu_1 m^2 c^2}{4p^2 \nu} \right) + \frac{4p^2 \nu}{m^2 c^2 \nu_1} - \frac{m^2 c^2 \nu_1}{2p^2 \nu} \right],
\]

(5.57)

where \( T_c = 2.73(1 + z) \) is the temperature of the CMB photons.

The corresponding synchrotron and inverse Compton spectra are plotted in figure 5.14 at different times, for a central magnetic field \( B(r = 0) \simeq 1\mu G \). The spectra are compared with that observed for the radio halo in the Coma cluster: an initial energy density in the relic relativistic electrons of the order of \( 5 \times 10^{-5}E_{th} \) is required to account for the data. In figure 5.14 we also report the luminosity of the EUV excess in the Coma cluster as an upper limit since it is commonly accepted that the origin of the EUV excess is not directly related to the same electron population responsible for the radio and possibly for the HXR emission (Bowyer, Berghöfer & Korpela, 1999; Ensslin, Lieu & Biermann, 1999; Atoyan & Völk, 2000; Brunetti et al., 2001b; Tsay, Hwang & Bowyer, 2002). Here we stress that figure 5.14 does not show the best fit to the data but just a comparison between data and time evolution of the emitted spectra resulting from the very simple scaling of the parameters described above. On the other hand, it should also be stressed that the time evolution of the synchrotron and inverse Compton spectra reported in figure 5.14 is generated via the first fully self-consistent calculation of particle acceleration in galaxy clusters. Thus provided that the energy of relativistic protons in galaxy clusters is not larger than a few percent of the thermal energy, figure 5.14 proves the possibility to obtain the observed magnitude of the non-thermal emission in these objects via Alfvénic acceleration.

In passing we note that the integrated radio spectrum at \( \sim \) GHz frequencies is steeper than the inverse Compton spectrum in the hard X-ray band independently of the time slice. This seems in nice agreement with some temptative evidences published in some recent literature (Fusco-Femiano, 2000).
5. Turbulent Alfvénic reacceleration: protons and electrons
Chapter 6

Future perspectives: 
non–liner shock acceleration 
during structure formation

Shock waves play an important role during the formation of large scale structures in the universe, being the main responsible for the heating of the hot intergalactic gas. It is now strongly believed that these shocks can also accelerate efficiently particles via first order Fermi acceleration, but to date there is not any conclusive and direct evidence supporting this fact (see chapters 3,4). To this respect, future gamma ray observations, both from space and from Earth, will be of crucial importance. In fact, according to our predictions (see chapter 4), gamma rays emitted by particles accelerated at strong large scale shocks, located at the periphery of clusters and around filaments, should be detectable by next generation instruments. Moreover, due to their short diffusion length, electrons are supposed to radiate all their energy via inverse Compton scattering very close to the sites in which they are accelerated. For this reason, gamma rays from electrons energized via diffusive acceleration during structure formation are expected to trace the large scale structure of strong shock waves in the universe.

However, all the predictions made about the detectability of such gamma ray photons, rely on the assumption that shocks are unmodified by the pressure of cosmic rays they accelerate. This assumption is satisfied only if a very small fraction of the thermal particles crossing the shock is injected in the suprathermal distribution and accelerated up to relativistic energies. Since shocks are believed to be efficient accelerators, more realistic calculations should include the backreaction of cosmic rays onto the shock structure. This inclusion makes shocks complex non–linear systems and the spectra of accelerated particles are supposed to be no longer simple power laws.

In principle, particle spectra can be calculated analytically solving the diffusion–advection equation for cosmic rays, coupled with the hydrodynamical equations
that describe the dynamics fluid flow. The Euler equation has to be opportunely modified by adding a term representing the cosmic ray pressure. However, all the attempts made to solve these equations revealed the presence of multiple solutions, corresponding to very different particle spectra. This make these approaches hardly applicable to some practical astrophysical situation, since all the solutions found are physically plausible.

In this chapter we discuss the issue of non–linear shock acceleration, with particular attention to the approaches based on the analytical solution of the diffusion–advection equation. The problem of multiple solutions will be discussed, and a possible way to avoid it will be proposed. Finally, we make some speculations on the consequences that non–linear effects might have on the non–thermal activity during large scale structures formation.

6.1 Non–linear shock acceleration: a semi–analytical approach

In chapter 2 we have reviewed the most important theoretical aspects of diffusive shock acceleration. All the results have been obtained treating cosmic rays as test particles having negligible pressure. Such an approach is somewhat unsatisfactory, since it requires a very small fraction $\eta << 10^{-4}$ of the particles crossing the shock to be converted into cosmic rays, while collisionless shock waves are believed to accelerate particles much more efficiently. Evidence for such an high efficiency comes from direct observations of the Earth bow shock (Ellison, Möbius & Paschmann, 1990) and more indirectly from radio emission from supernova remnants (Reynolds & Ellison, 1992). Moreover, also numerical simulations seem to suggest high values for the efficiency parameter, which is found to be $\eta \sim 10^{-3}$ (Ellison, Berezhko & Baring, 2000; Kang, Jones & Gieseler, 2002).

Besides numerical simulations, there are other different approaches to study the properties of shocks modified by the backreaction of the accelerated particles. One is the two fluid approach, in which cosmic rays are treated as a relativistic fluid (Drury & Völk, 1981). This model allows one to derive the thermodynamical properties of modified shocks but do not provide any information about the spectrum of the accelerated particles. This problem can be avoided solving analytically the diffusion–advection equation for cosmic rays accelerated at the shock, coupled with the fluid equations describing the behaviour of the thermal gas. Promising results have been recently obtained in the framework of this analytical approach (Malkov, 1997; Blasi, 2002, 2004).

Both the fluid and the analytical approaches are characterized by the puzzling appearance of multiple solutions. It has been suggested that the shock behaves like a self–regulating system settling on the critical point (Malkov, Diamond & Völk, 2000) but it is also possible that the multiple solutions are simply a consequence of
6.1. Non-linear shock acceleration: a semi-analytical approach

some of the assumptions of the models (Blasi, 2004).

In fact, as we will show in the following, the use of an appropriate recipe to
describe in a self-consistent manner the injection of particles at shocks might help
to solve this problem.

We review here the results presented in (Blasi, 2002, 2004), where the test-
particle assumption has been relaxed and the diffusion–advection equation has been
solved in the most general case in which the cosmic ray pressure is not negligible
with respect to the thermal gas pressure.

In the case of a one-dimensional shock the diffusive transport of particles is
described by the equation:

$$\frac{\partial}{\partial x} \left[ D \frac{\partial f(x,p)}{\partial x} \right] - u \frac{\partial f(x,p)}{\partial x} + \frac{1}{3} \frac{du}{dx} \frac{\partial f(x,p)}{\partial p} + Q_0(p) \delta(x) = 0 \quad (6.1)$$

where stationarity ($\partial f/\partial t = 0$) has been assumed. The equation can be integrated
with the help of the quantity (Blasi, 2002, 2004):

$$u_p = u_1 - \frac{1}{f_0} \int_{-\infty}^{0} dx \frac{du}{dx} f(x,p) \quad (6.2)$$

which represents the average fluid velocity experienced by particles with momentum
$p$ while diffusing upstream away from the shock. Here, $u_1$ is the gas velocity just
upstream of the shock and $f_0$ is the cosmic ray distribution function at the shock.
The main difference with respect to the linear case is that accelerated particles
experience a spatially variable advection velocity in the upstream region, due to
the cosmic ray pressure that slows down the fluid. Since the diffusion coefficient in
general depends on the particle momentum, particles with different momenta feel
a different velocity and a different compression factor, higher at high energies for
well-behaved diffusion coefficients. The meaning of the function $u_p$ can be better
understood in this way: particles with momentum $\hat{p}$ can diffuse over a distance
$x_{\hat{p}}$ away from the shock. Up to this distance, the particle distribution function
$f(x,\hat{p})$ can be considered roughly constant, while it is strongly suppressed at greater
distances. This means that:

$$u_{\hat{p}} \sim u_1 - \int_{-x_{\hat{p}}}^{0} dx \frac{du}{dx} \sim u(x_{\hat{p}}) \quad (6.3)$$

which is the fluid speed at the point where the particles with momentum $\hat{p}$ reverse
their motion and come back to the shock.

It is straightforward to integrate equation (6.1) following the same procedure
adopted for the derivation of equation (2.20). The solution can be written in the
implicit form (Blasi, 2004):

$$f_0(p) = \int_p^{p_0} \frac{d\hat{p}}{\hat{p}} \frac{3}{u_{\hat{p}} - u_2} \left[ u_0 f_{-\infty}(\hat{p}) + Q_0(\hat{p}) \right] \exp \left\{ \int_{\hat{p}}^{p_0} \frac{dp'}{p'} \frac{3}{u_{p'} - u_2} \left[ u_{p'} + \frac{1}{3} \frac{du_{p'}}{dp'} \right] \right\} \quad (6.4)$$
6. Future perspectives: non-linear shock acceleration during structure formation

where \( u_0 \) and \( u_2 \) are the upstream infinity and downstream fluid velocities respectively, and \( f_{-\infty} \) is the distribution function of the seeds particles. In the case of monochromatic injection with momentum \( p_{\text{inj}} \) we can write:

\[
Q_0(p) = \frac{n_{\text{gas}} u_1}{4 \pi p_{\text{inj}}^2} \delta(p - p_{\text{inj}}) \tag{6.5}
\]

where \( n_{\text{gas}} \) is the gas density of particles immediately upstream and \( \eta \) is an efficiency representing the fraction of the particles crossing the shock which starts to be accelerated. It is useful to introduce the dimensionless velocity \( U(p) = u_p/u_0 \) and the quantities \( R_{\text{tot}} = u_0/u_2 \) and \( R_{\text{sub}} = u_1/u_2 \) which represent the total compression factor and the compression factor of the modified shock (subshock). Equation (6.4) can be rewritten as (Blasi, 2004):

\[
f_0(p) = \frac{3R_{\text{tot}}}{R_{\text{tot}} U(p) - 1} \int_0^p \frac{dp'}{p'} f_{-\infty}(p') \exp \left\{ - \int_{p'}^{p} \frac{dp''}{p''} \frac{3R_{\text{tot}} U(p'')}{R_{\text{tot}} U(p'') - 1} \right\} + \frac{3R_{\text{sub}}}{R_{\text{tot}} U(p) - 1} \eta \frac{n_{\text{gas}}}{4 \pi p_{\text{inj}}^2} \exp \left\{ - \int_{p_{\text{inj}}}^{p} \frac{dp''}{p''} \frac{3R_{\text{tot}} U(p'')}{R_{\text{tot}} U(p'') - 1} \right\} \tag{6.6}
\]

The solution is known if the dimensionless velocity \( U(p) \) is known. Note that \( U(p) \) in turn depends on the distribution function \( f_0(p) \), and this fact reflect the non-linearity of the problem.

The dynamics of the fluid is modified by the presence of cosmic rays and it is described by the continuity and Euler equations:

\[
\begin{align*}
\rho_0 u_0 &= \rho_p u_p \tag{6.7} \\
\rho_0 u_0^2 + P_{g,0} + P_{\text{CR},0} &= \rho_p u_p^2 + P_{g,p} + P_{\text{CR},p} \tag{6.8}
\end{align*}
\]

where \( \rho \) is the gas density, \( P_g \) and \( P_{\text{CR}} \) are the gas and cosmic ray pressures. The subscripts 0 and \( p \) refer to quantities evaluated upstream infinity and where the fluid velocity is \( u_p \), respectively. The cosmic ray pressure at a given point upstream of the shock is the sum of two different contributions: the pressure of the adiabatically compressed seed cosmic rays coming from upstream infinity and the pressure of cosmic rays accelerated or reaccelerated at the shock.

If the diffusion coefficient is an increasing function of momentum, the typical distance that a particles with momentum \( p \) can move away from the shock is \( \sim D(p)/u_p \), which is an increasing function of momentum. This means that, from the all particles accelerated or reaccelerated at the shock, the only ones that contribute to the cosmic ray pressure at \( x_p \) are those with momentum greater than \( p \). The validity of this statement depends obviously on how strongly the diffusion coefficient depends on the particle momentum. We assume here that the momentum dependence of the diffusion coefficient is strong enough to allow us to write the total cosmic ray pressure at a generic point \( x_p \) as:

\[
P_{\text{CR},p} = P_{\text{CR},0} \left( \frac{u_0}{u_p} \right)^{\gamma_{\text{CR}}} + \hat{P}_{\text{CR}}(p) \tag{6.9}
\]

108
with the contribution of particles from the shock given by:

\[ \tilde{P}_{CR}(p) \sim \frac{4\pi}{3} \int_{p}^{p_{\text{max}}} dp p^3 v(p) f_0(p) \]  

(6.10)

where \( \gamma_{CR} = 4/3 \) is the cosmic ray adiabatic index, \( v(p) \) is the velocity of a particle with momentum \( p \) and \( p_{\text{max}} \) is the maximum momentum of the accelerated particles.

The gas pressure follows the well known adiabatic relation:

\[ P_{g,p} = P_{g,0} \left( \frac{u_0}{u_p} \right)^{\gamma_g} \]  

(6.11)

and its value far upstream can be written as a function of the shock Mach number \( M_0 \):

\[ P_{g,0} = \frac{\rho_0 u_0^2}{\gamma_g M_0^2} \]  

(6.12)

Here \( \gamma_g = 5/3 \) is the adiabatic index for a perfect gas.

Making use of equations (6.9–6.12), the Euler equation becomes (Blasi, 2004):

\[ p \frac{dU}{dp} \left[ 1 - \frac{\gamma_{CR} \xi_{CR}}{\gamma_g M_0^2} U^{-\gamma_{CR}+1} - \frac{1}{\gamma_g M_0^2} U^{-\gamma_g+1} \right] = \frac{4\pi}{3\rho_0 u_0^2} p^4 v(p) f_0(p) \]  

(6.13)

where the parameter \( \xi_{CR} \) quantifies the relative weight of the cosmic ray pressure compared with the gas pressure at upstream infinity: \( \xi_{CR} = P_{CR,0}/P_{g,0} \). The distribution function \( f_0 \) appearing on the right hand side of the equation depends on \( U(p) \) as in equation (6.6). Therefore, equation (6.13) is a integro–differential nonlinear equation for \( U(p) \).

Once the dimensionless fluid velocity \( U(p) \) is known, we can use equation (6.6) to evaluate the distribution function of the accelerated particles \( f_0(p) \). However, in order to do so, we need a relation between the two compression factors \( R_{\text{sub}} \) and \( R_{\text{tot}} \). It is easy to show that this relation can be written as (Blasi, 2004):

\[ R_{\text{tot}} = M_0^{\gamma + 1/2} \left[ \frac{(\gamma_g + 1) R_{\text{sub}}^{\gamma_g} - (\gamma_g - 1) R_{\text{sub}}^{\gamma_g+1}}{2} \right]^{\gamma_g+1}. \]  

(6.14)

### 6.2 The problem of multiple solutions

Equation (6.13) can be used to derive the function \( U(p) \) for any choice of \( R_{\text{sub}} \) and \( R_{\text{tot}} \). The boundary condition at \( p = p_{\text{inj}} \) is fully determined by the values of the compression factors, being \( U(p_{\text{inj}}) = u_1/u_0 = R_{\text{sub}}/R_{\text{tot}} \). The behaviour of \( U(p) \) at greater momenta can be obtained by means of equation (6.13). However, the only solutions having physical meaning are those having \( U(p_{\text{max}}) = 1 \), because at distances from the shock greater than \( x_{\text{pmax}} \), there are not accelerated particles contributing to the pressure and the fluid flow must be unperturbed.
6. Future perspectives: non-linear shock acceleration during structure formation

Figure 6.1: Velocity at upstream infinity in units of $u_0$ as a function of the total compression factor. Only particles injected at the shock from the thermal pool are considered. Different curves refers to different assumptions for the parameter $\eta$. Figure from Blasi (2004).

In figure 6.1, we plot $U(p_{\text{max}})$ as a function of $R_{\text{tot}}$ for a shock having Mach number $M_0 = 150$, as obtained solving equation (6.13). Different curves refer to different values of the efficiency parameter $\eta$. A maximum momentum equal to $10^5mc$ has been assumed and only particles accelerated at shock from the thermal pool have been considered. The injection momentum is taken as $10^{-3}mc$. Solutions are found at the intersections between the solid curves and the dashed straight line $U(p_{\text{max}}) = 1$. For low efficiency $\eta \sim 10^{-5}$, only the linear solution, corresponding to $R_{\text{tot}} \sim R_{\text{sub}} \sim 4$ exists. When the efficiency becomes greater than a few $10^{-4}$, other two solutions appear in addition to the linear one. Both these solutions are characterized by values of the total compression factor well above the standard value 4, derived from linear theory. If very high values for the efficiency are considered (e.g. $\eta \gg 10^{-3}$), only one, strongly modified solution exists. The value of the parameter $\eta$ at which the critical behaviour appears and disappears depends on the maximum momentum. In general, for an higher maximum momentum, the multiple solutions appear in correspondence to lower values for the efficiency.

It is possible to show that the appearance of multiple solutions is a very common fact. Multiple solutions are found if very different values of the parameters such as injection momentum, maximum momentum, Mach number and relative pressure of seeds particles are adopted (Blasi, 2004).
6.3 Comparison with an alternative approach

In the approach proposed by Blasi (2002, 2004) and described in the previous sections, particle spectra and velocity profiles of the fluid flow can be obtained without specifying a functional form for the diffusion coefficient \( D(p) \). The only condition to be satisfied is that this coefficient must be a rapidly increasing function of momentum. Within this assumption, the spectrum of the accelerated particles does not depend on the properties of the diffusion coefficient close to the shock. This fact makes the approach appealing because, even if it is commonly believed that the diffusion coefficient should be an increasing function of momentum, the exact determination of its functional form is an extremely difficult task.

The choice of a Böhm like diffusion coefficient, well motivated in presence of strong turbulence, seems plausible close to a shock surface. In this case, the diffusion coefficient simply scales linearly with particle energy. However, different scalings are also possible and one may wonder within which assumptions made on diffusion, the approach presented here ceases to be reliable. To this purpose, we make use of an alternative approach to the problem in which the diffusion coefficient needs to be explicitly specified (Malkov, 1997) and we compare the results coming from the two different methods.

Malkov (1997) considered a shock propagating in the positive-\( x \) direction and suggested to use the flow potential:

\[
\Psi = \int_0^x dx' u(x')
\]

(6.15)
as a new independent spatial variable instead of \( x \). Using the flow potential it is possible to define an integral transform of the flow profile as follows:

\[
\hat{U}(p) = \frac{1}{u_1} \int_{0^-}^{\infty} \exp \left[ -\frac{q(p)}{3D(p)} \Psi \right] du(\Psi)
\]

(6.16)

where \( q(p) = -d \ln f_0 / d \ln p \) is the spectral index of the particle distribution function and \( D(p) \) is the diffusion coefficient, which is assumed to be independent from position. An integral equation for \( \hat{U}(p) \) can be derived by applying the integral transform (6.16) to the \( x \)-derivative of the Euler equation (Malkov, Diamond & Völk, 2000):

\[
\hat{U}(t) = \left( \frac{R_{\text{sub}} - 1}{RR_{\text{sub}}} \right) \frac{\nu}{p_{\text{inj}}} \int_{p_{\text{inj}}}^{\text{max}} d\hat{p} \frac{\hat{p}}{\sqrt{\hat{p}^2 + 1}} \left[ 1 + \frac{q(p)D(p)}{q(\hat{p})D(\hat{p})} \right]^{-1} \times
\]

\[
\times \frac{\hat{U}(p_{\text{inj}})}{\hat{U}(\hat{p})} \exp \left[ -\frac{3}{RR_{\text{sub}}} \int_{p_{\text{inj}}}^{\hat{p}} d\ln p' \right]
\]

(6.17)

where \( R = R_{\text{tot}}/R_{\text{sub}} \) is the precursor compression factor and \( \nu \) is the injection parameter (Malkov, Diamond & Völk, 2000):

\[
\nu = \frac{4\pi mc^2}{3} \frac{\rho_0 u_0^4}{\rho_{\text{inj}} f_0(p_{\text{inj}})}
\]

(6.18)
6. Future perspectives: non–linear shock acceleration during structure formation

Figure 6.2: Downstream particle spectrum $f(p)p^4$ (upper panels) and fluid flow speed $U(p)$ (lower panels) for a multiple solution relative to a shock with Mach number 150, upstream temperature $10^4$ K and minimum and maximum momentum $10^{-3}$ and $10^5 mc$. The three solutions are characterized (left to right) by values of the compression factor $R = 15.30, 3.94, 1.05$. Solid lines are obtained following Blasi (2004), while dashed and dotted lines are obtained solving the system of equations (6.14,6.17,6.19) adopting two different diffusion coefficients (Böhm and Kolmogorov respectively).

related to the compression factor by means of the following equation (Malkov, Diam-
mond & Völk, 2000):

$$\nu = p_{inj} \left(1 - \frac{1}{R}\right) \left\{ \int_{p_{inj}}^{p_{max}} dp \frac{p}{\sqrt{p^2 + 1}} \frac{\dot{U}(p_{inj})}{U(p)} \exp \left[ -\frac{3}{R_{sub}R} \int_{p_{inj}}^{p} d\ln p' \right] \right\}^{-1} \quad (6.19)$$

Equations (6.14,6.17,6.19) form a close system and can be solved numerically.

In order to compare the results from the two different approaches we consider a shock having Mach number $M_0 = 150$ and upstream temperature $T = 10^4 K$. We set the value of the injection and maximum momenta equal to $10^{-3} mc$ and $10^5 mc$ respectively and we adopt an efficiency equal to $\eta = 10^{-4}$. Using the approach proposed by Blasi (2004), we find a triple solution, characterized by the values of the compression factor $R \sim 15.30, 3.94, 1.05$. Then, we adopt these three values for the compression factor to solve the system of equations (6.14,6.17,6.19) for different choices of the diffusion coefficient. In figure 6.2 we plot the particle spectrum (upper panels) and the dimensionless fluid velocity (lower panels) for each one of the three solutions. Note that the spectra corresponding to high values of the compression
6.4. Thermal leakage as a recipe for injection: preliminary results

The injection of particles into the cosmic ray population at a shock can be understood only considering the complex non-linear interactions between suprathermal particles, MHD waves and background thermal plasma. Suprathermal particles streaming against the background plasma can generate MHD waves which in turn can strongly scatter particles, inhibiting them from leaking upstream. As a consequence, only a small fraction of particles can swim upstream against the waves and be further accelerated via Fermi mechanism. The efficiency of the injection process regulates the degree of shock modification.

Due to its intrinsic complexity, the injection process is often parametrized by means of an injection momentum $p_{\text{inj}}$, representing the minimum momentum of the particles that can take part to the acceleration process, and an efficiency $\eta$, which fix the fraction of the thermal particles that are injected in the cosmic ray population. These two quantities are often treated as free parameters (Malkov, 1997; Malkov, Diamond & Völk, 2000; Blasi, 2002, 2004).

Another possibility is to adopt the thermal leakage model to describe the injection (Ellison & Eichler, 1984; Kang & Jones, 1995). In this model, the post shock gas is assumed to be completely thermalized at a temperature $T_2$. Protons in the tail of the Maxwellian distribution can recross the shock and escape back upstream if their velocity is high enough to allows them not to be trapped by waves. Those are the
6. Future perspectives: non–linear shock acceleration during structure formation

Figure 6.3: $U(p_{\text{max}})$ versus total compression factor for a shock according to Blasi (2004). A fraction $\sim 10^{-4}$ of the downstream particles are supposed to recross the shock and to start to be accelerated. Different lines refer to different values for the maximum momentum: (top to bottom) $p_{\text{max}} = 5 \times 10^{10}, 10^7, 10^5, 10^3, 10^2 mc$. The Mach number is 100 (left panel) and 1000 (right panel).

It is easy to implement such a recipe in the calculations presented in §6.1. Results are shown in figure 6.3, where the function $U(p_{\text{max}})$ has been calculated within the assumption that a fraction $\sim 10^{-4}$ of the particles are able to recross the shock from downstream to upstream. Two values for the shock Mach number, 100 (left panel) and 1000 (right panel), have been considered. Different curves refer to different values for the maximum momentum; from top to bottom, $p_{\text{max}} = 5 \times 10^{10}, 10^7, 10^5, 10^3, 10^2$ in units of $mc$. The most striking feature of this plot is the disappearance of multiple solutions. Moreover, the solution is found to be strongly modified ($R_{\text{tot}} \gg 4$), with the level of modification increasing with the maximum momentum. Therefore, these preliminary results seem to suggest that protons which are injected in the accelerator. Usually, in this model, the injection momentum is set to a few times the thermal momentum:

$$p_{\text{inj}} = \epsilon \sqrt{2mkT_2} \quad (6.20)$$

with $\epsilon$ tuned in order to fit observational (Ellison, Möbius & Paschmann, 1990) and theoretical (Ellison & Eichler, 1984) studies of diffusive shock acceleration. It is important to stress that the parameter $\epsilon$ fixes the value of both injection momentum and efficiency, which are no longer free parameters but are connected in a physically motivated way.
6.4. Thermal leakage as a recipe for injection: preliminary results

Figure 6.4: Example of multiple solutions. The efficiency is the same as in figure 6.3. The Mach number is 1415 and a very low value for the maximum momentum \( p_{\text{max}} = 150mc \) has been adopted.

The appearance of multiple solutions might be due to an unphysical choice of the couple of parameters \((p_{\text{max}}, \eta)\), whose values should instead be appropriately fixed to describe in a more self-consistent way the process of injection.

In fact, multiple solutions can appear even after the introduction of the thermal leakage model for the injection. However, they are confined in a region of parameter space which is usually of poor interest for the majority of the astrophysical situations considered. An example of a multiple solution is given in figure 6.4, where a shock with Mach number 1415 is considered. The parameter \( \epsilon \) is fixed as in figure 6.3, while a very low value for the maximum momentum \( p_{\text{max}} \sim 150mc \) has to be adopted in order to find a critical behaviour of the system.

However, our preliminary results seem to suggest that multiple solutions might be the artifact of some of the assumptions used in the analytical approach. Other assumptions, such as the request for stationarity and the fact that the role of the self-generated waves on the diffusion coefficient is not taken into account, should be
6. Future perspectives: non-linear shock acceleration during structure formation

relaxed in order to have a fully self-consistent treatment of the problem.

6.5 Possible implications for non-thermal phenomena during large scale structure formation

In chapter 4 we discussed the perspective for the observation of forming clusters of galaxies with both space telescopes and ground-based Čerenkov arrays. We considered the inverse Compton gamma ray emission from relativistic electrons accelerated at shocks associated with structure formation. Spectra of the accelerated electrons have been calculated in the test-particle approximation, namely, neglecting the backreaction of the accelerated particles onto the shock structure. Such an approach is somewhat unsatisfactory, because the behaviour of shocks is believed to differ from the predictions of linear theory even if a very small fraction \( \eta \gtrsim 10^{-4} \) of the thermal particles is injected into the cosmic ray population. Moreover, large scale accretion shocks are strong, and can accelerate particles up to extremely high energies. Both these facts suggest that these shocks might be strongly modified by cosmic ray pressure.

The dominant contribution to the cosmic ray pressure at shocks comes from protons, while electrons are likely to be energetically subdominant, for the reasons explained in §3.3.1. However, electrons feel the velocity flow profile in the upstream region as modified by the cosmic ray pressure and, for this reason, their spectra is expected to be also modified and to mimic the proton spectra at energies greater than \( \sim 1\text{GeV} \).

Modified spectra are no longer power laws and, at high energies, they are expected to be flatter than the spectra predicted from linear theory. This fact might modify our predictions about the number of detectable clusters in gamma rays. In particular, flatter spectra should produce an enhancement of the TeV emission, making the possibility of detecting forming clusters with ground based Čerenkov telescopes more appealing.

Also the predictions of the contribution from forming structures to the extragalactic gamma ray background could be modified by the inclusion of non-linear effects. This could be of great interest especially in connection with the recent reassessment of the extragalactic background to a lower value and to a different spectral shape (Strong, Moskalenko & Reimer, 2004).

Moreover, if the injected spectra of the relativistic protons are flatter than the canonical \( E^{-2} \) behaviour, a high fraction of the non-thermal energy can be stored in very high energy particles, for which the confinement argument suggested by Berezinsky, Blasi & Ptuskin (1997) might not apply. If this is the case, particle diffusion should be important and it might be necessary to reconsider the confinement hypothesis in this new framework.

The inclusion of the effects of cosmic ray pressure might also have interesting
6.5. Implications for non–thermal phenomena during structure formation

consequences for several topics related to structure formation. For example, if non–
linear effects are found to be important in shock acceleration, then the heating of
the gas could be strongly suppressed. Moreover, under the confinement hypothesis,
the cosmic ray energy density in the intracluster medium is an increasing function of
time. If the acceleration efficiency is high enough, this energy density could become
comparable with the thermal energy density, making cosmic rays important for the
dynamical evolution of clusters. All these effects have always been neglected in
previous studies and, in our view, it will be important to evaluate their relevance
for a better understanding of the “thermal–side” of structure formation.
6. Future perspectives: non-linear shock acceleration during structure formation
Conclusions

It is now commonly believed that structures in the universe are assembled hierarchically. According to this picture, more massive objects are the result of the merging of smaller ones, which have been formed at higher redshift. During mergers, a huge amount of gravitational energy is dissipated at large scale shock waves, which are the main responsible for the heating of the intracluster gas. Due to their hot temperatures, these shocks are bright in X-rays and can be directly imaged by means of high resolution observations performed by new generation space telescopes (Forman et al., 2002).

A possible link between the process of large scale structure formation and the presence of non-thermal activity in the intracluster medium was suggested a few decades ago, in order to explain the presence of a diffuse synchrotron radio emission from a few X-ray bright clusters of galaxies (Harris et al., 1980). Such a connection was proposed on the basis of simple considerations about the energy required to power the observed radio halos, which is only a very small fraction of the total energy released during a major merger event.

Radio halos are now observed from ~30% of the X-ray brightest clusters (Feretti, 2003), and additional evidences for the presence of relativistic electrons in the cluster volume come from hard X-ray (Fusco-Femiano et al., 2003) and extreme UV (Bowyer, 2003) observations. Moreover, the connection between non-thermal activity and cluster mergers has recently received observational support (Buote, 2001). However, the mechanism through which particles are accelerated up to relativistic energies is still not well understood.

It is well known that collisionless shock waves can extract particles from the thermal pool and accelerate them up to relativistic energies, through the so-called first order Fermi mechanism (Bell, 1978a; Blandford & Ostriker, 1978). For this reason, large scale shock waves associated to structure formation have been considered for a long time good candidates to accelerate the electrons responsible for the observed radio emission (Sarazin, 1999; Blasi, 2001; Fujita & Sarazin, 2001). Relativistic protons can also be accelerated at shocks, and they can produce electrons as secondary products via inelastic collisions with the protons of the interstellar medium. Synchrotron emission from secondary electrons has been considered also as a possibility to explain radio observations (Dennison, 1980; Blasi & Colafrancesco, 1999).
Conclusions

In chapter 3, we proposed a novel approach to the study of particle acceleration at merger related shocks. In the framework of the so-called extended Press–Schechter formalism (Lacey & Cole, 1993), we simulated several possible realizations of the merger trees of rich clusters and, for each merger event, we estimated the Mach number of the related shocks by using an analytical model to describe the dynamics of the collision. The most important result obtained, is that shocks related to major mergers are weak, having a Mach number of order unity. We find that relatively strong shocks, with Mach number \( \geq 3 \), form only during minor mergers, which are collisions between clusters with very different masses. This fact has crucial implications for the non-thermal emission from clusters, since the slopes of the spectra of the accelerated particles are solely determined by the value of the Mach number. Due to the weakness of merger related shocks, spectra of relativistic particles in the intracluster medium are expected to be very steep and unlikely to reproduce the observed spectral features of radio halos.

Promising alternative models have been proposed to explain radio observations, based on the turbulent reacceleration of relic electrons present in the intracluster medium (Schlickeiser et al., 1987; Brunetti et al., 2001a; Fujita, Takizawa & Sarazin, 2003). MHD turbulence is expected to be generated in the intracluster medium during mergers, while possible sources for the relic electron population are active and normal galaxies in the cluster and shocks in the intracluster medium. In chapter 5 we considered for the first time the most general situation in which MHD waves, relic electrons, relativistic and thermal protons are present in the cluster volume. Both wave–wave and wave–particle interactions are considered in a fully time dependent and self-consistent way. We find that present radio and hard X-ray observations of non-thermal radiation from clusters of galaxies are well described within this approach, provided the fraction of relativistic hadrons in the intracluster medium is smaller than \( 5 - 10\% \).

Besides mergers, clusters can increase their mass also through a continuous accretion of matter. This type of accretion is also called secondary infall, meaning that matter accretes on a potential well which has already been formed (Bertschinger, 1985). An accretion shock forms roughly at the cluster virial radius and it carries the information of the virialization of the inner region of the cluster. The accretion shock is strong, because it propagates in the cold, non virialized intergalactic medium.

Our Press–Schechter based approach has been extended in order to take into account the effects of these strong shocks. In chapter 4, we have shown that the total kinetic energy flux through shocks related to structure formation is dominated by the contribution from shocks related to major mergers. This means that only a small fraction of the gravitational energy released during structure formation is dissipated at strong shocks, which are associated with minor mergers or accretion events. However, these shocks have high Mach numbers and thus they can accelerate particles with flat spectra, whose contribution to the total non-thermal emission is
likely to be dominant at very high energies.

For this reason, we have calculated the gamma ray emission expected from forming clusters, due to inverse Compton scattering from electrons accelerated at merger and accretion shocks. Another important contribution to the total gamma ray emission might come from the decay of neutral pions, produced by the hadronic cosmic ray component, which is likely to be present in the intracluster medium. As explained in §4.3, this contribution is quite uncertain, due to the fact that protons are diffusively confined in the cluster volume for cosmological times (Berezinsky, Blasi & Ptuskin, 1997). As a consequence, the proton related gamma ray emission from a given cluster depends on its whole merger history, and similar clusters might emit gamma rays at a different flux level due to their different history. Thus, in chapter 4, we have adopted a conservative approach and consider only the gamma rays which are the result of inverse Compton emission of relativistic electrons. Any contribution to the total gamma ray emission from cosmic ray protons can only increase the fluxes we derived.

In chapter 4, we calculated the log $N - \log S$ distribution of merging and accreting clusters of galaxies and compared the expected fluxes with the sensitivities of EGRET, AGILE and GLAST. We found that no cluster was expected to be detected by EGRET, confirming a recent analysis of the archive EGRET data from the direction of several Abell clusters, for which only upper limits on the gamma ray emission could be inferred (Reimer et al., 2003). Our calculations significantly disagree with previous results found by Totani & Kitayama (2000) and Waxman & Loeb (2000), and this is due to their incorrect assumptions made about the strength of the shocks.

The perspectives for detection appear to be more promising for future space-borne gamma ray telescopes, such as AGILE and GLAST. Our calculations show that $\sim 10 - 20$ clusters should be detected by AGILE and $\sim 50$ clusters should appear in GLAST data, equally shared between merging and accreting clusters if an efficiency of $\sim 5\%$ (in energy) can be achieved in accelerating electrons at the shocks.

In principle, the gamma ray emission due to inverse Compton scattering of ultrarelativistic electrons can extend to a few tens of TeV, so that clusters might be seen in future ground based Cherenkov experiments such as VERITAS and HESS. Unfortunately, we found that only nearby ($\sim 100$ Mpc) merging clusters of galaxies are expected to be detectable, while the detection of accreting clusters appears more problematic due to their extended size.

In chapter 4, we also evaluated the contribution of structure formation to the extragalactic diffuse gamma ray background, concluding that, despite some recent optimistic claims (Loeb & Waxman, 2000), this contribution adds up at most $\sim 10\%$ of the observed background.

As pointed out in chapter 6, strong shocks are believed to efficiently accelerated particles. The effects of cosmic ray pressure have never been considered in the
Conclusions

calculations of the spectra of the particles accelerated at large scale shocks. We suggest that these modifications might have important implications on the expected gamma ray fluxes from clusters at both GeV and TeV energies. Moreover, if the acceleration efficiency is high enough, the cosmic ray energy density might become comparable with the thermal energy density of the intracluster medium, making relativistic particles important for the dynamical evolution of clusters.

All these facts seem to suggest that future gamma ray observations, both from space and ground based telescopes, will certainly constitute one of the most important advances in our knowledge of the non-thermal activity in large scale structure, but they could also improve our understanding of the process through which large scale structures in the universe are assembled.
Acknowledgments

First of all I would like to thank my thesis adviser, Pasquale Blasi. I have learnt a lot of physics talking and working with him. During these three years, we had innumerable lively discussions about several topics, including clusters of galaxies, food, diffuse radiation backgrounds, books, ultrahigh energy cosmic rays, politics, very low energy cosmic rays, jazz music (you are wrong, Pasquale!), old neutron stars, shocks, alcohol, starburst galaxies, gamma ray bursts, guitar chords, molecular clouds, TV programs, particle acceleration, supernova remnants, blues music and many many other stuff... I enjoyed them.

Thanks!

Many thanks to Prof. Franco Pacini, for his support and advices, and to all the members (past and present) of the High Energy Astrophysics Group in Arcetri for several interesting discussions.

Thank (Many thanks!) to Gianfranco Brunetti for a productive ongoing collaboration and for his precious help in solving lots of problems.

I also want to thank Prof. Giancarlo Setti and Luigina Feretti, who (together with Gianfranco) supervised my earlier work on clusters of galaxies when I was an undergraduate student in Bologna.

Moreover, I enjoyed instructive discussions with several people about issues related to my thesis work. These discussions helped me a lot in improving what I was doing. For this reason I wish to thank Tom Jones (shock waves during structure formation), Antonaldo Diaferio (effects of local overdensities onto the pairwise cluster velocities), Felix Aharonian (gamma rays from clusters and Cherenkov observations of extended sources), Don Ellison (non-linear shock acceleration) and Stephane Barland\footnote{These discussions were characterized by the constant presence of several cold beers, and this made them much more pleasant. I also have to thank Stephane for his .vimrc file!} (non-linear systems).

Thanks to all the people who shared with me the room number 84 at the Observatory. They are (in chronological order): Elena Amato, Niccolò Bucciantini, Valentine Wakelam (Hi Val!), Yari Something (sorry Yari, I never knew your sur-
name...), Jaron Kurk, Joao Emanuel Guerreiro Dos Santos Pinto Dias, Alessandro Bemporad.

Many thanks go to all the other people in (and outside) Arcetri who shared with me a lot of time, making my stay here so enjoyable. Sorry, but it is impossible to write all your names! So I will write only a few of them... Carmelo&Sara, Virg&Aurelie, Massimo (Isp. Canotti), Marco&Matteo, Antonio, my flatmates Jack&Miranda&Ciccio, Guido, Tommy, Dafne, Willy, the President, Claudia, Amelie and obviously Caatje.

I thank my family...and I think I don’t have to explain why!

Probably, it would have been impossible to write this thesis without the continuous and re-energizing support of the music by Vinicio Capossela.
Bibliography


BIBLIOGRAPHY


Fermi, E., Phys. Rev. 75 (1949) 1169.


Gabici, S., 2004, to be published in MmSAIt Suppl. as a proceeding of the workshop: Modelling the Intergalactic and Intracluster Media, edited by S. Borgani, A. Ferrara and V. Antonuccio-Delogu.


BIBLIOGRAPHY


Stecker, F. W., & Salamon, M. H., PRL 76 (1996b) 3878.


Strong, A. W., Moskalenko, I. W., & Reimer, O., 2003, to be published in Proc. 28th International Cosmic Ray Conference


