STRIKING NONLINEAR MODE COUPLING IN THE CMBR

Nir J. SHAVIV

Theoretical Astrophysics 130-33,
California Institute of Technology
Pasadena, CA 91125, USA
Email: nir@tapir.caltech.edu

Abstract. Fluctuations in the early universe are routinely treated within the linear approximation. We examine this linear hypothesis and its validity by adding the force that large wavelength perturbations exert on short wavelength waves during the time of recombination. We find that the large scale shears between the matter and radiation render the very short isothermal waves unstable. Moreover, the amplitude of the large scale fluctuations expected from typical cosmological models is on the limit between having a negligible effect and having nonlinear small scale structure form at recombination. Under certain circumstances, it can even affect the large, horizon size scales as well. It could have also been the source of a significant pregalactic magnetic field.

1. Introduction

Observations of the Cosmic Microwave Background Radiation (CMBR) reveal that the Universe was highly linear at the time of decoupling (e.g. Bennett et al. 1996, [1]). Since nonlinear structure is predicted to form only after gravity had a long enough time to amplify the linear amplitudes irrespective of the cosmological scenario (unless it originally existed on small scales), the hydrodynamic equations describing the behavior of the perturbations are linearized such that the behavior of each wavevector is separated from all other modes. As however will soon be shown, the dimensionless amplitude isn't the only dimensionless number and nonlinear coupling could in fact be important to the evolution of the "linear universe".

The first to develop a linear theory for the perturbed Friedmann-Robertson-Walker metric was Lifshitz (1946, [2]), and it can be found in various text books such as Weinberg (1972, [3]), Peebles (1980, [4]) and Kolb & Turner (1990, [5]). The theory has since then been applied to describe cosmologies with various components and various initial parameters. Baryonic matter, radiation and other massless particles, and cold (massive) or hot (light) dark matter are some of the ingredients that enter the primordial soup. While other parameters such as the Hubble constant H_0, the vacuum energy through the cosmological constant Λ, and the initial spectrum affect the evolution. A review of various current cosmological models, their evolution and implications can for example be found in White et al. (1994, [6]) with emphasis on the microwave background radiation and in Primack (1997, [7]). Less current reviews that cover the basic principles of structure formation are found in the aforementioned textbooks. In principle, if only linear evolution plays a role, then future high accuracy

measurements such as those that will be made by Planck or MAP can be, in a straight forward manner, be compared with linear model predictions to extract cosmological parameters such as the Hubble or cosmological constants, the density parameters, or a combination of them. This is one of the reasons why any deviation from the linear theory could be important.

In a recent paper [8], it was shown that an important coupling term arises between the large or horizon-scale modes and small, globular-cluster sized modes during the early universe recombination, a period when the opacity was a function of density. This paper summarizes this effect. Further details can be found in Shaviv (1998, [8]) where the linear analysis is extensively described and in Shaviv and Levin (1999, [9]) where the behavior of the nonlinear structure formed is analyzed, including the generation of a pre-galactic magnetic field and gravitationally unstable objects.

We start by describing the effect with simplest terms. We then continue with a more detailed linear analysis. The nonlinear study is then followed using both analytical and numerical methods. Once the behavior of the small scales is understood, its ramifications to the CMBR and structure formation can be studied more thoroughly – the generation of a magnetic field, the distortion of the CMBR energy spectrum, the generation of gravitationally collapsed objects, etc. We conclude by showing that the coupling effect is not only important to the small scales but that it could have also changed the observable CMBR power spectrum.

2. The Effect - In simple terms

At the lowest order of approximation, waves of different wavelengths are decoupled from each other. This suggests that a small-scale isothermal wave will, at this order, witness an isotropic and homogeneous medium around it. However, at the next order of approximation, one has to include the small perturbations (of order $\delta\rho/\rho \sim 10^{-4}$) of all the other scales in the vicinity of the isothermal wave.

One can simplify the problem by assuming that the isothermal wave is localized to a region of size d that is at least several times the wave's wavelength (such that the wave parameters are well defined).

Next, we need to find the waves that could contribute a coupling term to the small-scale isothermal waves that does not average out. The chosen scale d separates between two types of perturbations that could potentially affect the small-scale waves: Those with a wavelength $\lambda \lesssim d$ and those with $\lambda \gtrsim d$. The contribution to the average radiative force of the former group vanishes in the lowest order, since the spatial correlation between the wave and perturbation yields a net vanishing result. The latter group can contribute a net effect only if the temporal correlation between the perturbation and the isothermal wave does not vanish as well; namely, it will do so only if the perturbation does not have enough time to oscillate rapidly during the period considered. This implies that for a given period Δt, the contribution of waves with $\omega\Delta t \gtrsim 1$ averages out leaving no net contribution. The only waves that have a net contribution are those for which $\omega\Delta t \lesssim 1$; that is, if d is chosen smaller than the wavelengths of waves with $\omega \sim 1/\Delta t$, its exact value is unimportant. Hence, we solve for short wavelength isothermal waves in a region where the long wavelength perturbations can be considered constant and homogeneous. We cannot however, assume that the local environment of the isothermal wave is isotropic.

We assume for simplicity that the large scale perturbations are optically thick

adiabatic waves of the form:

$$\frac{\delta\rho_a}{\rho} = \delta_a \sin\left(\mathbf{k}_a \cdot \mathbf{x} - \omega_a t\right), \tag{1}$$

where $\delta\rho_a$ is the perturbation of the matter density. It is proportional to the temperature perturbations:

$$\delta\rho_a/\rho = \delta_a = 3\delta T_a/T, \tag{2}$$

namely, the fluctuations in the radiation field and in the matter are synchronized. The sole force responsible for accelerating the matter is the radiative force. If observed from the baryon rest frame, the force is due to a net radiative flux originating from an anisotropy in the radiation field. This anisotropy or shear between the radiation and matter fluids can be found through the baryon acceleration.

Using the adiabatic speed of sound c_A to relate $\mathbf{k}_a$ with ω_a and the continuity equation to relate $\mathbf{v_a}$ with $\delta\rho_a$, the velocity of the baryonic fluid in the adiabatic wave is found to be:

$$\mathbf{v}_a = \delta_a c_A \sin\left(\mathbf{k_a} \cdot \mathbf{x} - \omega_a t\right) \hat{\mathbf{n}}_\mathbf{a}, \tag{3}$$

with $\hat{\mathbf{n}}_\mathbf{a}$ being a unit bector in $\mathbf{k_a}$'s direction. The acceleration or the force per unit mass is simply $d\mathbf{v}_a/dt$; consequently, if the opacity per unit mass is κ, the net "internal" radiative flux H needed to accelerate the baryons in the wave, is given by:

$$\mathbf{H} = \frac{\mathbf{a_a}}{\kappa} = \frac{1}{\kappa}\frac{d\mathbf{v}_a}{dt} = -\frac{\delta_a c_A \omega_a}{\kappa} \cos\left(\mathbf{k_a} \cdot \mathbf{x} - \omega_a t\right) \hat{\mathbf{n}}_\mathbf{a}. \tag{4}$$

The form for an isothermal wave traveling in a small region perturbed by the wave above is:

$$\frac{\delta\rho_i}{\rho} = \delta_i \sin\left(\mathbf{k_i} \cdot \mathbf{x} - \omega_i t\right). \tag{5}$$

with $\delta T_i = 0$ as it is isothermal. It will witness a constant radiative flux $\mathbf{H}$ given by eq. 4 as it is optically thin. However, if κ is a function of density, the opacity and therefore the force per unit mass, become a function of the wave's phase. The predominant source of opacity before and during recombination is Thomson scattering off the free electrons. Consequently, the only period during which isothermal waves vary the opacity is during recombination, when the number of free electrons is proportional to $\rho^{-1/2}$. Hence, the total acceleration is:

$$\mathbf{a} = \kappa\mathbf{H} = \kappa_0\mathbf{H} + \delta\kappa_i\mathbf{H} \equiv \mathbf{a}_a + \mathbf{a}_i = \mathbf{a}_a\left(1 - \frac{1}{2}\frac{\delta\rho_i}{\rho}\right). \tag{6}$$

The first term results with the large-scale acceleration already accounted for in the large-scale oscillation. The second term however incites a force that varies synchronously with the small-scale isothermal waves. It leads to an instability.

The total power input per unit mass by the latter term into the wave is:

$$p_{\rm rad} = \mathbf{a}_i \cdot \mathbf{v}_i = \frac{1}{2} c_A \omega_a \delta_a \cos\left(\mathbf{k_a} \cdot \mathbf{x} - \omega_a t\right) \sin\left(\mathbf{k_i} \cdot \mathbf{x} - \omega_i t\right)^2 c_T \delta_i^2 \cos\theta, \tag{7}$$

where $\hat{\mathbf{n}}_\mathbf{i}$ denotes a unit vector in $\mathbf{k}_i$'s direction and θ is the angle between $\hat{\mathbf{n}}_\mathbf{i}$ and $\hat{\mathbf{n}}_\mathbf{a}$. Again, we related $\mathbf{v}_i$ to ρ_i through the continuity equation and the isothermal speed of sound c_T. When averaging the force over time one has to recall that the large

scale perturbations are assumed not to vary over the relevant period – the duration of recombination. Now we see why. Oscillating large-scale perturbations will average to a zero net contribution. Those that oscillate on a longer time scale can be assumed constant:

$$\delta_a \cos\left(\mathbf{k}_a \cdot \mathbf{x} - \omega_a t\right) \rightarrow \delta_\phi . \tag{8}$$

Using $\left\langle \cos^2 \right\rangle = 1/2$ to average the isothermal oscillations, we find:

$$\langle p_{\rm rad} \rangle = \frac{1}{4} c_A \omega_a \delta_\phi c_T \delta_i^2 \cos\theta. \tag{9}$$

The total energy per unit mass in the acoustic wave is

$$e = \delta_i^2 c_T^2 / 2. \tag{10}$$

The e-fold growth rate of the wave's amplitude is therefore:

$$r = \frac{\langle p_{\rm rad} \rangle}{2e} = \frac{1}{4} \omega_a \delta_\phi \frac{c_A}{c_T} \cos\theta. \tag{11}$$

Let the duration of recombination be Δt. The growth factor $\mathcal{G}$, or the number of e-folds the wave will grow if traveling in the direction of the large scale shear (i.e., if $\theta = 0$), is:

$$\mathcal{G} = r\Delta t = \frac{\delta_\phi}{4} \frac{c_A}{c_T} \left(\omega_a \Delta t\right) \sim \mathcal{O}\left(\frac{\delta_\phi}{10^{-4}}\right), \tag{12}$$

where we have inserted the ratio of the speeds of sound at recombination assuming radiation dominance over baryonic matter. We have also taken $\omega_a \Delta t \sim 1$, as waves with a larger ω_a average out. δ_ϕ at recombination will have a distribution of values. If however, the typical value is of order 10^{-4}, then the typical growth rate is a few e-folds. Namely, we expect that the nonlinear effect is at least as important as the linear evolution of the isothermal wave. It is now also evident why isothermal waves are unstable while adiabatic are mostly stable. One only has to replace c_T by c_A to find that the adiabatic waves are unstable only for δ_ϕ of order unity and not much less.

Note that we have implicitly assumed in the calculation of $\langle p_{\rm rad} \rangle$ that the small-scale wave has time to oscillate. This limits the validity of this simple picture to waves smaller than the isothermal Jeans scale. We have also assumed that radiation damping is negligible. When calculating the small scale evolution properly (cf §4), one finds the same result for small enough scales (that are somewhat smaller than the Jeans scale) provided that the flux H is larger than a critical flux (which is always the case in the cosmologically interesting parameter range).

Eq. 12 is useful as an estimate, but a more general expression for the growth parameter $\mathcal{G}$ should not include an assumption on the type, speed, or optical depth of the large-scale waves, but instead use directly the *radiation flux observed in the baryon rest frame*. By using eq. 4 and letting κH vary during recombination, we have:

$$\mathcal{G} \sim \frac{1}{4 c_T} \left| \int_0^{\Delta t} \kappa \mathbf{H} dt \right| . \tag{13}$$

3. The full equations of motion

We review in this section the basic equations governing the interaction of the small-scale acoustic waves with the horizon-scale diffusive radiation flow. We focus our attention on the photon-decoupling era, since it is in this period that the opacity is density dependent – a result which renders unstable the small-scale acoustic waves propagating against the photon wind [8].

In its general form, this instability isn't a recent discovery. In fact, it has been known since the early work of Hearn [10] that acoustic waves can be unstable when a radiative flux is present. Since then, the instability was primarily used for the explanation of atmospheric phenomena in stars (e.g., [11], [12] and [13]) and quasars [14].

In both this and the next section, we model the instability described above by an idealized problem. From the outset, we imagine that a net radiative flux is present, a flux that originates from the perturbations of large scales. The detailed *linear* analysis of the problem with a full description of the assumptions appears in [8]. We simplify the equations considerably by making the following assumptions:

- We first assume that the scales of interest are optically thin and are much smaller than the scales on which the driving flux changes. This is an excellent approximation since $\lambda/\ell_{\rm mfp} \sim 10^{-4}$ where λ is the small-scale size of interest and $\lambda_{\rm mfp}$ is the photon "mean free path" in the beginning of decoupling. Because of this optical thinness, the small-scale baryonic density perturbation will not cause small-scale radiation density perturbations (and therefore its temperature will be unperturbed as well). We can consequently solve the equations locally where the flux can be considered spatially constant.
- Second, the scales are much smaller than the horizon scale and can therefore be solved under the Newtonian approximation.
- Third, we assume that the large scales are unaffected by the smaller scales and are therefore given. This can be a bad approximation if the small scales can damp an appreciable amount of energy from the large scales. This damping is interesting and is briefly dealt in section §10.
- Fourth, we assume that the cooling time scale is smaller than all other scales in the system. As a consequence, the baryonic fluid can be assumed to be isothermal. This approximation holds even after recombination ends because of the residual ionization left.

Under the above approximations, the equations governing the flow are the continuity and momentum equations in which the baryonic pressure is related to the baryonic density through the isothermal equation of state: $p = c_T^2 \rho$. The Eulerian equations are:

$$\left(\frac{\partial \rho}{\partial t}\right)_r + \nabla_r \cdot (\rho \mathbf{u}) = 0 \tag{14}$$

$$\rho \left[\left(\frac{\partial \mathbf{u}}{\partial t}\right)_r + (\mathbf{u} \cdot \nabla_r)\,\mathbf{u}\right] = -c_T^2 \nabla_r \rho - \rho \nabla_r \phi + \mathbf{f}_{\rm rad}\rho, \tag{15}$$

where $\mathbf{r}$ and $\mathbf{u}$ denote the coordianate and velocity with respect to real space. ϕ is the gravitational potential and $\mathbf{f}_{\rm rad}$ is the radiative force per unit mass. In order to simplify the equations for the expanding universe, it is customary to define a co-moving coordinate $\mathbf{x} \equiv \mathbf{r}/a$. A comoving velocity $\mathbf{v}$ is then defined as: $\mathbf{u} = \dot{a}\mathbf{x} + \mathbf{v}$. We also

define a dimensionless density $\varrho = \rho/\bar{\rho}$, where $\bar{\rho}$ is the average density†. To simplify the equations more, the covariant time is defined through $dt = ad\tau$. Equations 14 & 15 are then simplified to (cf §9 in Peebles 1980, [4]):

$$\frac{\partial \varrho}{\partial \tau} + \nabla \cdot (\varrho \mathbf{v}) = 0 \tag{16}$$

$$\frac{\partial \mathbf{v}}{\partial \tau} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\dot{a}\mathbf{v} - c_T^2 \frac{\nabla \varrho}{\varrho} - \nabla\phi + \mathbf{f}_{\rm rad} a, \tag{17}$$

where for breivety we omitted the x index from ∇_x. The gravitational potential is given by:

$$\nabla^2 \phi = 4\pi G(\rho - \bar{\rho})a^2 = 4\pi G(\varrho - 1)a^{-1}. \tag{18}$$

Using the notation in [8], the force per unit mass (which is the flux times the opacity per unit mass) can be written as:

$$\mathbf{f}_{\rm rad} = \mathbf{H}\kappa = \frac{4}{3}\sigma_T \frac{\sigma}{c} T^4 \frac{\mathbf{v}_r - \mathbf{v}}{c} \frac{n_e}{\rho_b} \equiv \frac{\mathbf{v}_r - \mathbf{v}}{\tau_{e\gamma}}. \tag{19}$$

$\mathbf{v}_r$ is the "velocity" of the radiation fluid (the velocity of the frame of reference in which the radiation flux vanishes), and $\mathbf{v}$ is the co-moving velocity of the baryonic matter. σ is the Stephan-Boltzmann constant, σ_T is the Thomson cross section, and ρ_b is the mass density of baryons. The last equality is used to define $\tau_{e\gamma}$: It is a typical time scale for the transfer of momentum from the radiation to the baryons, through scattering by the electrons. We can define the velocity of the large-scale diffusive shear as:

$$\Delta\mathbf{v}_0 \equiv \mathbf{v}_r - \mathbf{v}_0. \tag{20}$$

Without loss of generality, we choose a noninertial frame of reference in which the average velocity of the baryonic fluid is zero, thus $\mathbf{v}_0 = 0$. The radiation force exerted on a unit mass of baryons is then given by:

$$\mathbf{f}_{\rm rad} = \frac{\Delta\mathbf{v}_0 - \mathbf{v}}{\bar{\tau}_{e\gamma} + \delta\tau_{e\gamma}} = \frac{\Delta\mathbf{v}_0}{\bar{\tau}_{e\gamma}} - \frac{\mathbf{v}}{\bar{\tau}_{e\gamma}} - \frac{\Delta\mathbf{v}_0}{\bar{\tau}_{e\gamma}} \frac{\delta\tau_{e\gamma}}{\bar{\tau}_{e\gamma}}. \tag{21}$$

The first term on the r.h.s. is responsible for the large-scale acceleration of the fluid and is responsible for the horizon-scale oscillations of the baryons. Since we are interested only in the small-scale motions within the this accelerated frame of reference, we should subtract this force out.

As long as ionization equilibrium is maintained, one finds that $(\tau_{e\gamma}/\bar{\tau}_{e\gamma})^{-1} = \mathcal{I}(\rho/\rho_0) = \mathcal{I}(\varrho)$. $\mathcal{I}$ is the ratio between the perturbed ionization and the unperturbed one. If the ionization fraction is in equilibrium, it will be given by the Saha equation. It yields $\mathcal{I} = \sqrt{\rho_0/\rho} = \varrho^{-1/2}$ for a small ionization fraction†.

We assume that from recombination onwards the universe was matter dominated such that $a \propto t^{2/3}$. If we define a_0 and t_0 as the expansion factor and time at the

† Note that ϱ is most often defined as $1 + \delta$ instead, however, this definition is useful only for a perturbative analysis.

† The alleviation of the equilibrium assumption is beyond the scope of this work, nevertheless, since recombination is proportional to the density, one should expect to have $\mathcal{I} \sim \rho^{-1}$, if equilibrium cannot be maintained towards the end of recombination (e.g., by integrating eq. 6.112 in Peebles [15] when neglecting the ionization term) and the small-scale oscillation is not too fast. In such a case, $\mathcal{G}$ could in fact be larger.

beginning of recombination, we have that $a/a_0 = (t/t_0)^{2/3}$. Using the relation with the conformal time we find: $\tau/\tau_0 = (t/t_0)^{1/3}$ and $a/a_0 = (\tau/\tau_0)^2$ with $\tau_0 \equiv 3\tau_0^{-1/3}a_0^{-1}$.

$\bar{\tau}_{e\gamma}^{-1}$ and the radiative force are proportional to T^4 or a^{-4}. The isothermal speed of sound is proportional to the gas temperature. If we define $\Theta(t)$ as the ratio between the gas temperature and the radiation temperature, then before and during recombination it is unity and it falls as a^{-1} afterwards, after the thermalization processes quench. It is used in the numerical simulations when integrating the evolution beyond the end of recombination. Using the definitions, the momentum equation is simplified to:

$$\frac{\partial \mathbf{v}}{\partial \tau} + (\mathbf{v}\cdot\nabla)\mathbf{v} = -\frac{2}{3}a(t)\mathbf{v} - c_{T0}^2\Theta(t)a^{-1}\frac{\nabla\varrho}{\varrho} - \nabla\phi + \frac{1}{a^3\tau_{e\gamma 0}}\left(-\mathbf{v} + \varrho\Delta\mathbf{v}_0\mathcal{I}(\varrho)\right). \quad (22)$$

4. Summary of the linear limit analysis

The first analysis of the instability of small-scale acoustic waves was simplified by being limited to small perturbations [8]. This is suitable for understanding the source of the instability. We review in this section the results of the linear analysis. We begin by decomposing $\delta \equiv \varrho - 1$, $\mathbf{v}$ and ϕ into the sum of Fourier components. Each component is characterized by a wave vector $\mathbf{k}$ such that $\delta_k, \mathbf{v}_k, \phi_k \propto \exp(-i\mathbf{k}\cdot r/a(t))$. Eqs. 16 & 17 then to first order become:

$$\begin{aligned} 0 &= \dot{\delta}_k - \frac{i\mathbf{k}}{a}\cdot\mathbf{v}_k \\ 0 &= a\dot{\mathbf{v}}_k + \dot{a}\mathbf{v}_k - i\mathbf{k}c_T^2\delta_k - i\mathbf{k}\phi_k + a\frac{\mathbf{v}_k}{\bar{\tau}_{e\gamma}} + \alpha a\frac{\Delta\mathbf{v}_0}{\bar{\tau}_{e\gamma}}\delta_k \\ 0 &= \phi_k + \frac{4\pi G\bar{\rho}}{k^2}a^2\delta_k \end{aligned} \quad (23)$$

with α defined as:

$$\alpha \equiv \left(\frac{1}{\mathcal{I}(\varrho)}\frac{d\mathcal{I}(\varrho)}{d\varrho}\right)_{\varrho=1}. \quad (24)$$

After some algebraic manipulation, one obtains that the equation for the perturbed comoving density is (eq. 25 of [8]):

$$\ddot{\delta}_k + \left(\frac{2\dot{a}}{a} + \frac{1}{\bar{\tau}_{e\gamma}}\right)\dot{\delta}_k + \left(k^2c_T^2 - 4\pi G\bar{\rho} - i\alpha\frac{\Delta\mathbf{v}_0\cdot\mathbf{k}}{\bar{\tau}_{e\gamma}}\right)\delta_k = 0. \quad (25)$$

The main effect discussed in this work arises from the contribution of the last term.

The solution of the above equation in the various limits of k can be found in [8] (table 1). The main point of interest is the solution for small wavelength waves since they have the largest amplification. If $\Delta v_0 > c_T/\alpha$ then waves of which their wavevector component in the direction opposite to the large-scale flux is larger than:

$$k_\infty = \frac{\Delta v_0}{c_T}\frac{\alpha}{c_T\bar{\tau}_{e\gamma}}, \quad (26)$$

are unstable and their growth rate is given by:

$$\Re(r) \approx \frac{\alpha}{2}\frac{\Delta\mathbf{v}_0\cdot\hat{\mathbf{n}}}{c_T}\frac{1}{\bar{\tau}_{e\gamma}}. \quad (27)$$

Under more general conditions, the growth rate is given by (eq. 30 of [8]):

$$r = -\frac{1}{2\tau_{e\gamma}} + \sqrt{\frac{1}{4\tau_{e\gamma}^2} - \left(k^2 c_T^2 - 4\pi G\bar{\rho} - i\alpha \frac{\Delta \mathbf{v_0} \cdot \mathbf{k}}{\tau_{e\gamma}}\right)}. \tag{28}$$

Since we are interested in the integrated growth during the period of recombination $[t_0, t_1]$, one can define the following growth parameter as:

$$\mathcal{G}(k, \hat{\mathbf{n}}) = \int_{t_0}^{t_1} \Re(r(k, \hat{\mathbf{n}}))dt. \tag{29}$$

If the growth rate is larger than all other time scales in the system, then $\mathcal{G} \gg 1$ and the growth is exponential:

$$\frac{\delta_{t_1}(k, \hat{\mathbf{n}})}{\delta_{t_0}} \approx \exp(\mathcal{G}(k, \hat{\mathbf{n}})). \tag{30}$$

One can then proceed to calculate the average growth of the perturbations. To do so, one must first integrate over the 4π different possible propagation directions for the isothermal waves and then average over the possible distribution of $\mathbf{G} = \mathcal{G}\hat{\mathbf{n}}$. To do so, one needs to know the distribution of relative velocities between the radiation and baryon fluids. If one assumes that the large-scale fluctuations are Gaussian then each component of $\mathbf{G}$ has a Gaussian distribution and the distribution of $\{\mathcal{G}\}$ is a Maxwellian. The average growth can then be calculated to be (eq. 37 of [8]):

$$\left\langle \frac{\delta_{t_1}}{\delta_{t_0}} \right\rangle = \int_0^\infty \sqrt{\frac{54}{\pi}} \frac{\mathcal{G}^2}{\mathcal{G}_{\rm rms}^3} \exp\left[-\frac{3}{2}\left(\frac{\mathcal{G}}{\mathcal{G}_{\rm rms}}\right)^2\right] \frac{\sinh(\mathcal{G})}{\mathcal{G}} d\mathcal{G} = \exp\left(\frac{\mathcal{G}_{\rm rms}^2}{6}\right), \tag{31}$$

where $\mathcal{G}_{\rm rms}$ is the r.m.s. value of $\{\mathcal{G}\}$. As an example, for $\mathcal{G}_{\rm rms}$ of only 12, the average growth is a staggering 10 orders of magnitude yet it is only 10% for $\mathcal{G}_{\rm rms} = 1$. Clearly, the exact value of $\mathcal{G}_{\rm rms}$ is critical as it will determine whether the amplification is negligible or remarkable. Several cosmological scenarios were considered with a Harrison-Zel'dovich (scale independent) spectrum of initial perturbations, normalized to fit the COBE measurements of the CMB quadrupole fluctuations [8]. Values ranging from $\mathcal{G}_{\rm rms} \approx 1$ for a mixed HDM scenario to $\mathcal{G}_{\rm rms} \approx 5$ for a CDM model where obtained after proper integration of the large-scale acoustic oscillations. Larger or smaller values of $\mathcal{G}_{\rm rms}$ are not unexpectable since the CMBR fluctuations are not known accurately enough. Moreover, some of the assumptions made to simplify the problem may modify $\mathcal{G}_{\rm rms}$ as well.

One can also estimate the volume fraction $\mathcal{F}$ of the universe that will actually be able to develop nonlinearities. To find it, we need to know what is the typical amplitude δ_0 of the small-scale perturbations before amplification. If given, then the regions where $\mathcal{G} \gtrsim -\ln \delta_0$ will be amplified enough to have a final amplitude of order unity or more. The fraction of space that satisfies this condition depends on the distribution of $\mathcal{G}$. For a Gaussian distribution of fluctuations, one finds that this fraction is:

$$\mathcal{F}(\delta_0, \mathcal{G}_{\rm rms}) = \mathrm{erfc}\left(-\frac{3}{2}\frac{\ln \delta_0}{\mathcal{G}_{\rm rms}}\right) - \sqrt{\frac{6}{\pi}} \exp\left(-\frac{3}{2}\left(\frac{\ln \delta_0}{\mathcal{G}_{\rm rms}}\right)^2\right) \frac{\ln \delta_0}{\mathcal{G}_{\rm rms}}. \tag{32}$$

As an example, for $\delta_0 \approx 10^{-4}$ and $\mathcal{G}_{\rm rms} \sim 5$ one finds that 1% of the volume will reach nonlinearity by the end of recombination. One should expect an even larger number for a non Gaussian fluctuation distribution that has a power law tail instead.

The behavior of the baryonic fluid once the small-scale acoustic waves reach nonlinear amplitudes is the topic of the following.

5. 1D solution

In this section, we follow the small-scale instability into the nonlinear regime, where radiatively driven shocks are formed. For simplicity, we initially restrict ourselves to 1D; §6 will consider the 2D case. We investigate the behavior of the shocks analytically and numerically.

We neglect the universe's expansion during the time of recombination. This is a good approximation since $\dot{a}/a \ll 1/\tau_{e\gamma}$. We can write the one dimensional, dimensionless form of eqs. 16,17 & 18 as:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) = 0, \tag{33}$$

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \frac{1}{\rho}\frac{\partial \rho}{\partial x} + \left(\mathcal{I}(\rho/\bar{\rho})\frac{\Delta v_0}{\bar{\tau}_{e\gamma}} - f_{\rm avr}\right) + \mathcal{I}(\rho/\bar{\rho})\frac{v}{\bar{\tau}_{e\gamma}} + \frac{\partial \phi}{\partial x} = 0, \tag{34}$$

$$\frac{\partial^2 \phi}{\partial x^2} = \rho. \tag{35}$$

A general analytical solution does not exist, we therefore solve them under a few simplifications and simulate them numerically.

5.1. Analytical Solution

An analytical 1D solution can be found under several simplifying assumptions. In view of the linear problem, it is evident that on small enough scales such that the acoustic oscillation time scale is shorter than both the dissipation time scale and the Jeans (or Hubble) time scale, the two latter terms in eq. 34 can be neglected. By doing so, we simplify our problem considerably though limit ourselves to short wavelengths. Next, it is evident that the amplification of the waves will stop when any nonlinear dissipation process will be able to damp all the energy input by the radiation. By comparing for example to the work of Hearn [16] who studied the saturation of waves undergoing a similar instability in stellar atmospheres, it is clear that the damping of energy from the acoustic waves will be through the formation of shocks. This is also confirmed in our numerical investigation.

In order to understand the behavior of the shock waves that form, it is best to analyze the system once the waves have saturated and have reached a steady state in which the energy dissipated in shocks is equal to the energy pumped in by the radiation field. To do so, we solve the equations of motion for periodic shock waves with a given wavelength λ, traveling with a constant velocity v_w. In other words, we assume that the solution is a function of only $(x - v_w t)$.

Under the above approximations, one finds the following results [9]:

- The functional form of the periodic solution depends on only one parameter:

$$\xi \equiv \frac{\lambda \Delta v_0}{\bar{\tau}_{e\gamma} c_T^2} \sim \frac{2\lambda \mathcal{G}}{\alpha \Delta t c_T}. \tag{36}$$

  The typical time scale and the typical length scale do however scale the solution. Roughly, the parameter ξ is the square of the typical velocity that a mass element

will obtain (in units of the speed of sound) when accelerated by the radiation over a length λ. Thus, $\xi \sim 1$ will separate the weak shock solutions (for $\xi \ll 1$) from the strong shock solutions ($\xi \gg 1$).

- The periodic solution is of a shock wave and a subsequent rarification, as is apparent in figure 1 for a moderately strong shock.
- For a given wavelength λ, one finds the various properties of the shock wave and the average properties of the medium. These can be found in figure 2.

Although the 1D results are limited, they are nonetheless very important because they allow us to estimate the typical parameters of the fluid at a given instant after saturation has been reached by assuming that the typical wavelengths found in it are the given λ.

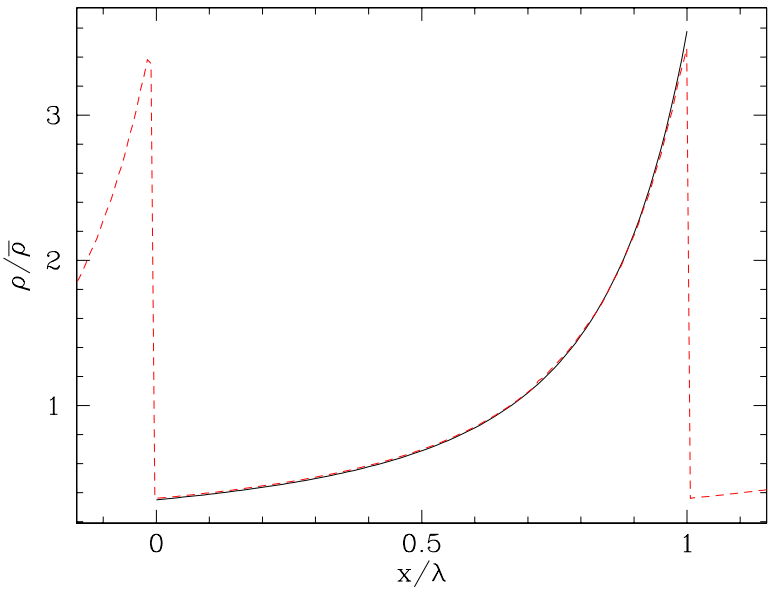

Figure 1. The saturated wave form for $\lambda = 0.01\lambda_J$ and $\xi = 12$ giving a moderately strong shock. A short wavelength was chosen so that the effects of the radiation drag and the gravitational fields will be negligible. The saturated numeric solution is given by the dashed line and the analytic result appropriate for short wavelengths is given by the solid line. The good agreement is evident. The slight discrepancies appear at the shock were the finite resolution of the numeric code is apparent.

5.2. Numerical Solution

The numerical codes written (both the 1D and the 2D) use the Flux Corrected Transport (FCT) scheme to integrate the isothermal hydrodynamic equations. The FCT method is a conservative, positivity ensuring, monotone technique developed by Boris & Book [17]. It combines integration schemes of low and high spatial accuracy together. The low order scheme provides a monotone solution by introducing a numerical diffusive flux. The high order scheme provides the high accuracy in regions where the solution is smooth and has small gradients. The high-order solution is obtained by correcting ("anti-diffusing") the low order monotone solution, but only to

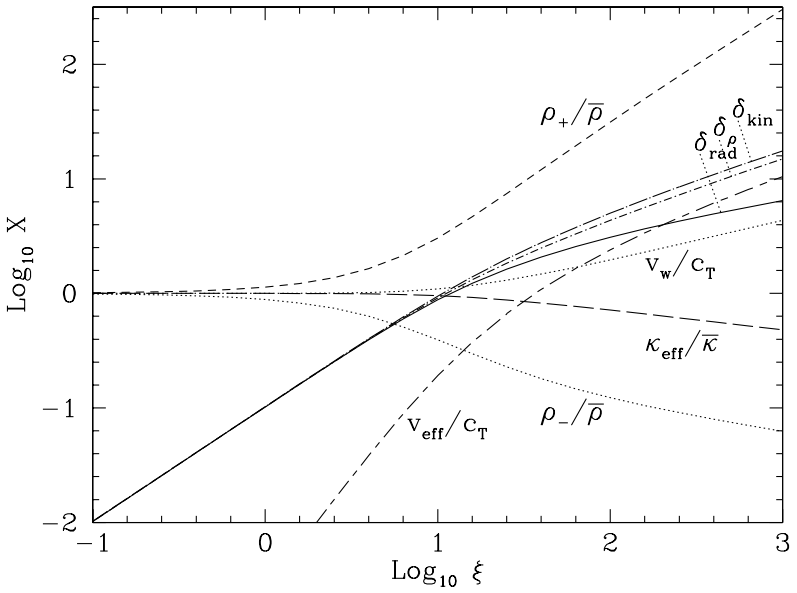

Figure 2. Summary of the analytical 1D results for saturated waves. Depicted as a function of $Log_{10}\xi$ with $\xi = \lambda\Delta v_0/c_T^2\bar{\tau}_{e\gamma}$ are the Log_{10} values of the various parameters that describe the saturated waves. $\rho^{-,+}$ are respectively the density before and after the shock. $\kappa_{\rm eff}$ is the averaged opacity. v_w is the speed of the shock. $v_{\rm eff}$ is the opacity weighted velocity of the medium. Even though the medium is at rest, the radiation field will "perceive" a moving medium. $\delta_{\rho,{\rm kin,rad}}$ are respectively the amplitude that a linear sinusoidal wave should have in order to have the same density r.m.s, kinetic energy or amount of radiation damping as the solution has.

the extent that no new extrema are created and no existing extrema are accentuated. This is achieved by limiting, or "correcting," the antidiffusive fluxes of the high-order scheme (hence its name). A detailed discussion and comparison with other methods can be found in Boris et al. [18].

The integrator subroutine itself is part of the publicly available FCT package of Boris et al. [18]. Besides its numerical advantages such as its good localization of shocks, its simplicity allows the easy modification of the hydrodynamic equations without the need to reprogram any Riemann solvers or adjust other problem specific code. This enables the easy implementation of the isothermal equations of motion (eqs. 16,22) and the description of the magnetic field dynamics (eq. 54). The numeric code uses periodic boundary conditions as they are most appropriate for the simulation.

The numeric code was checked by comparing the results of the simulations to the known analytical results as well as verifying the conservation laws. The conserved quantities were found to be conserved to always better than 0.3%, with typical accuracies that are better than 0.1%. The growth rate of the linear perturbations was compared to the analytical linear theory and a good agreement was found.

The saturated waveform of a short wave mode ($0.01\lambda_J$) was compared with the analytical solution found in §5.1. The results can be seen in figure 1. The agreement is, like the previous result, more than satisfying, and it demonstrates that the code is reliable even after nonlinear shocks are produced.

Properties at Saturation: The analytical solution of §5.1 was developed under the assumptions that $\lambda \ll \lambda_J$. However for very large wavelengths, the growth rate is inhibited by radiation drag and the saturation amplitude will be smaller. The simplest way to compare the actual numerical results to the analytical results which neglected the radiation drag and the gravitational potential is to define ξ using the actual growth rate $r(\lambda)$ given by eq. 28 and defining $\mathcal{G}(\lambda)$ using it. This allows us to define ξ more generally as $\xi(\lambda) = 2\lambda\mathcal{G}(\lambda)/\alpha\Delta t c_T$. Doing so, we find that to a good approximation (better that the numerical accuracy of roughly 1%), the properties at saturation of the finite sized waves all have the same value as the properties of the short-wave limit of modes with the same ξ.

The nonlinear growth rate: The second property of interest is the time it takes the shocks to form after the instability begins. Unlike the saturation properties which depend on $\lambda\mathcal{G}(\lambda)$, the growth rate is expected to be predominantly a function of $\mathcal{G}(\lambda)$ alone and not a function of λ explicitly (it isn't an explicit function of the wavelength at all for linear amplitudes). By plotting the time $t_{1/2}$ that it takes to reach half the saturation amplitude as a function of $\mathcal{G}(\lambda)$, for different wavelengths, it was found that there is some nontrivial scatter between cases with different wavelengths. Nevertheless, one can write for all practical purposes that the saturation time is roughly:

$$\frac{t_{1/2}}{\Delta t} \sim \frac{\ln(1/2) - \ln(\delta_0)}{\mathcal{G}(\lambda)}. \tag{37}$$

In other words, the time to reach half the saturation amplitude is roughly the time it takes to amplify the linear waves to an amplitude of 1/2. That is to say, the final growth before saturation is relatively very fast and no appreciable "slowing down" from the exponential growth is apparent. This serves as a justification of the linear analysis argument used to estimate of the fraction of regions that reach nonlinear amplitudes (eq. 32) since it implies that effectively all the way up to the saturation in the nonlinear regime, the growth rate is given by the linear one.

Merging of shock fronts: The analytical analysis assumed that once the waves saturate, they are forced to have a given wavelength λ. In reality however, no limit is imposed on the system. Such a case is depicted in figure 3 which describes the evolution of a system which initially has small amplitude white noise. The figure plots the location of the discontinuities (i.e., shocks) as a function of time. As the system begins to evolve, its density variations are very small and it has no discontinuities. Once however the noise is amplified sufficiently, small shocks begin to emerge. These shocks then start to interact with each other through an effective "attraction" force. The shocks attract each other and merge to form larger and larger structures. Since the typical wavelength is increased, the typical strength of the shocks increases as well and the density variations obtained become larger. This is reflected in the graph for δ_ρ.

The key point to note is that although the shock wave speed is close to c_T for weak shock "trains" with a wavelength λ and two weak shocks would be expected to have a small relative speed to each other, the "merging" speed of shocks is actually of order c_T as well. The reason is that if the shock fronts are not in their equilibrium position (as is the case for the traveling wave), then their mutual interaction changes their speed by $\mathcal{O}(c_T)$, even if they are weak.

We consequently find that as time will progress, the saturation length scales will grow at roughly c_T. If $\tilde{t}$ signifies the time since saturation, we should expect to have typical saturated wavelengths of order $\lambda_{\rm typ} = f_\lambda \tilde{t} c_T$ where f_λ is a factor of order unity

that could perhaps depend weakly on ξ. Its numerical value is found from the 2D numerical simulations (cf §6).

The strength of the saturated shocks will grow as well as a result of the wavelength growth. It will be given by:

$$\xi \sim \frac{2\mathcal{G} f_\lambda \tilde{t}}{\alpha \Delta t}. \tag{38}$$

For example, if $f_\lambda \sim 1$, $\alpha = 1/2$ and $\tilde{t}/\Delta t \sim 1/2$ at the end of recombination, we find that $\xi \sim 2\mathcal{G}$. Since saturation takes place only when $\mathcal{G}$ is large, once saturated the shocks are most likely to be in the strong shock regime. The physical scale of the shock separation for the above typical parameters is $\lambda_{\rm typ} \sim c_T \Delta t/2 \sim l_J/10$ (or roughly a tenth of the isothermal Jeans scale at recombination).

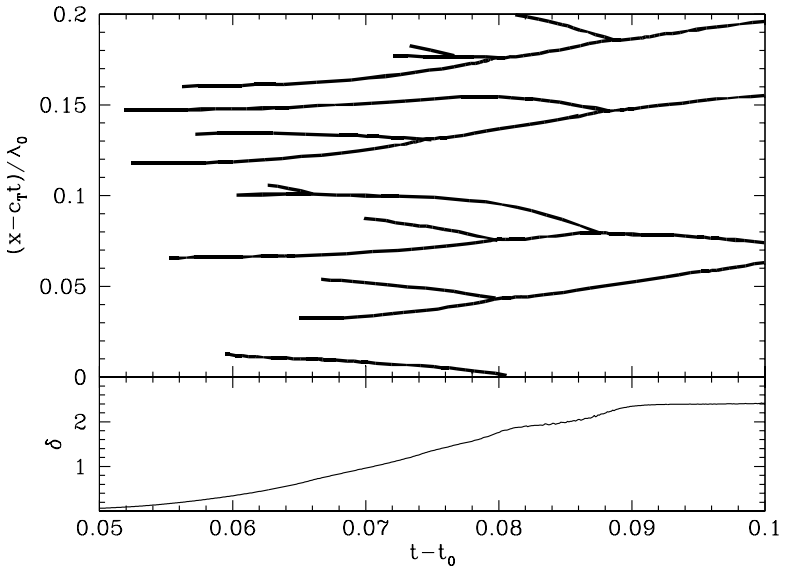

Figure 3. The shock fronts as a function of time for a 1D simulation (with $\mathcal{G} = 20$, $\delta_0 = 10^{-5}$, $c_T = 1$, $\Delta t = 0.1$, $\tau_{e\gamma} = 0.1$). The lower panel describes the dimensionless amplitude. According to the linear growth theory, shocks should start forming at $t - t_0 \approx (\ln(1/\delta_0)/\mathcal{G})\Delta t = 0.057$, as is roughly the case in the evolution. Once shocks have formed they on average propagate at roughly c_T if they are weak and faster if they are strong. However, because they interact with each other they are attracted and merge such that the typical distance between the shocks increases as c_T as well.

6. 2D solution

We extend here our numerical simulation to two dimensions. The most important finding is that the radiative shocks generate vorticity in the baryonic fluid.

To achieve the simulations of two dimensions, the simplest extension of a 1D simulation was used, namely, dimensional splitting. For each time step Δt, the system (eq. 16 and 22) is first advanced in the $\hat{x}$-direction according to the equations:

$$
\begin{aligned}
1: \quad \frac{\partial \varrho}{\partial \tau} &= -\frac{\partial}{\partial x}(\varrho v_x) \\
\frac{\partial \varrho v_x}{\partial \tau} &= -\frac{\partial}{\partial x}(\varrho v_x v_x) - \frac{2}{3}a(t)\varrho v_x - c_{T_0}^2 \Theta(t) a^{-1}\frac{\partial \varrho}{\partial x} \\
&\quad - \varrho\frac{\partial \phi}{\partial x} - \frac{\varrho}{a^3 \tau_{e\gamma 0}}\left(v_x + \varrho v_0 f(\varrho)\right), \\
\frac{\partial \varrho v_y}{\partial \tau} &= -\frac{\partial}{\partial x}(\varrho v_y v_x),
\end{aligned}
$$

and then again in the $\hat{y}$-direction:

$$
\begin{aligned}
2: \quad \frac{\partial \varrho}{\partial \tau} &= -\frac{\partial}{\partial y}(\varrho v_y) \\
\frac{\partial \varrho v_x}{\partial \tau} &= -\frac{\partial}{\partial y}(\varrho v_x v_y). \\
\frac{\partial \varrho v_y}{\partial \tau} &= -\frac{\partial}{\partial y}(\varrho v_y v_y) - \frac{2}{3}a(t)\varrho v_y - c_{T_0}^2 \Theta(t) a^{-1}\frac{\partial \varrho}{\partial y} \\
&\quad - \varrho\frac{\partial \phi}{\partial y} - \frac{\varrho}{a^3 \tau_{e\gamma 0}} v_y.
\end{aligned}
$$

It is assumed that the large scale flux is in the $\hat{x}$-direction. We have seen in the 1D analysis that the system exhibits an inverse cascade – the small scales grow first and they nonlinearly amplify the larger scales, that is, the typical length scales found in the system grow. We expect to see the same behavior in the 2D case. Moreover, the effective "attractive force" between shocks and the lack of any forces in the $\hat{y}$-direction, save the pressure, tends to homogenize the system. We should expect the shocks to become parallel to each other with the shock fronts aligning perpendicular to the $\hat{x}$-direction. Indeed, this is what is observed.

The summary of the sixteen 2D simulations that were performed, can be found in [9]. All simulations had a 256×256 grid. The initial perturbation was a Gaussian uncorrelated white noise field with initial r.m.s.w amplitudes that vary from run to run. The grid size was changed as well such that the typical box sizes corresponded to $\sim 0.5 - 2l_J$. Larger scales could not be simulated because the coarsing which is then forced on the small scales could not resolve the structure of individual shocks. A typical simulation took about 2 days of CPU time on an HP Workstation. Most of the simulation time is spent on the "saturation phase" where small time steps are required to adequately simulate the shocks. This forced a limit on the maximal $\mathcal{G}$ that could be simulated. Any larger $\mathcal{G}$ would have had a "saturation phase" for which the required simulation time is too long.

The simulations assume that the driving large-scale photon flux is turned on for a finite duration of time Δt (that is varied from simulation to simulation) and then switched off immediately. The simulations are then carried on further assuming the gas is still isothermal. This implies that the post-recombination results should be taken cautiously.

Figure 4 depicts snapshots of the density perturbations at different instants for a particular simulation.

The three stages of the evolution: linear growth, saturated behavior and dissipation are apparent. At first (fig. 4a), all the small scales are amplified according to the

exponential behavior. However, once the small-scale perturbation amplitude becomes of order unity, shocks are formed (fig. 4b). They start to merge together and amplify as a result of these mergers. Once the shocks becomes strong, they prevent the formation of additional shocks. The shock mergers increase the average distance between the shock fronts roughly as $\lambda \sim c_T t$ where t is the time since saturation, as is expected from a naive extrapolation of the 1D simulation. Typically, one finds that the correlation lengths in the direction perpendicular to the photon wind are as much as 5 to 10 times larger than the typical horizontal separations, implying that the fluid along a line parallel to the radiative force effectively behaves as a 1D system. After the driving flux is switched off, the energy input needed to sustain the shocks disappears and the shocks start to dissipate (fig. 4c). Interestingly, not long after the switching off, the appearance is that of an isotropic system (fig. 4d).

The final "strength" of the shocks depends on a single parameter $\tilde{\mathcal{G}}$ which is a function of the growth factor $\mathcal{G}$ and the initial small-scale perturbation amplitude δ_0:

$$\tilde{\mathcal{G}} \equiv \mathcal{G} + \ln(\delta_0). \tag{39}$$

The parameter $\tilde{\mathcal{G}}$ represents the integrated radiative drive of the photon wind *after* the small-scale waves have saturated. It is related to the expected strength of the shocks ξ_1 at the end of recombination:

$$\xi_1 = \frac{2\mathcal{G} f_\lambda \widetilde{\Delta t}}{\alpha \Delta t} = \frac{2\tilde{\mathcal{G}} f_\lambda}{\alpha} \tag{40}$$

where $\widetilde{\Delta t}$ is the $\tilde{t}$ at the end of recombination. It is the recombination time from which the linear growth time was excluded: $\widetilde{\Delta t} = \Delta t \tilde{\mathcal{G}}/\mathcal{G}$. The typical separation between shocks $\lambda_{\rm typ}$ scales with the actual duration of recombination and with the speed at which the fronts merge:

$$\lambda_{\rm typ} = f_\lambda c_T \widetilde{\Delta t}. \tag{41}$$

To find the proportionality constant f_λ which gives the average speed at which shocks merge and increase their typical separation distance, we measure $\lambda_{\rm typ}$ by finding the wavenumber k_0 at which $|\delta_k|$ is maximum, were the Fourier transform of δ is:

$$\delta_k = \frac{1}{2\pi} \int \delta(\mathbf{x}) \exp(-i\mathbf{k}\cdot\mathbf{x}) d\mathbf{x}. \tag{42}$$

We then define $\lambda_{\rm typ}$ as $2\pi/k_0$. When $\lambda_{\rm typ}/c_T\widetilde{\Delta t}$ is plotted as a function of $\tilde{\mathcal{G}}$, one obtains a horizontal line [9] which is consistent with $f_\lambda = 0.87 \pm 0.3$. Hence, shocks merge at a constant speed that is of order the isothermal speed of sound.

Using the merging speed and eq. 40, we can calculate the expected shock strength ξ_1 at the end of recombination, and from it estimate the expected amplitude $\delta_\rho^{(\rm est)}$ (or other variables). We can then compare it to the measured value δ_ρ. The two sets of amplitudes should agree if our heuristic picture in which the results of the 1D saturated solution can be applied to the 2D simulations. When compared (cf [9]), it was found that $\delta_\rho^{(\rm est)}$ and δ_ρ roughly agree even though there is some scatter. To quantify it, we calculate $\langle\, \delta_{\rm typ}/\delta_\rho^{(\rm est)} \,\rangle_{\rm runs}$ from 16 different simulations, and find that it is 0.98 ± 0.07, in very good agreement.

To conclude, we can use the heuristic picture of shock mergers to estimate the fluid variables (e.g., δ_ρ) once the small scale waves have reached saturation. All we need to

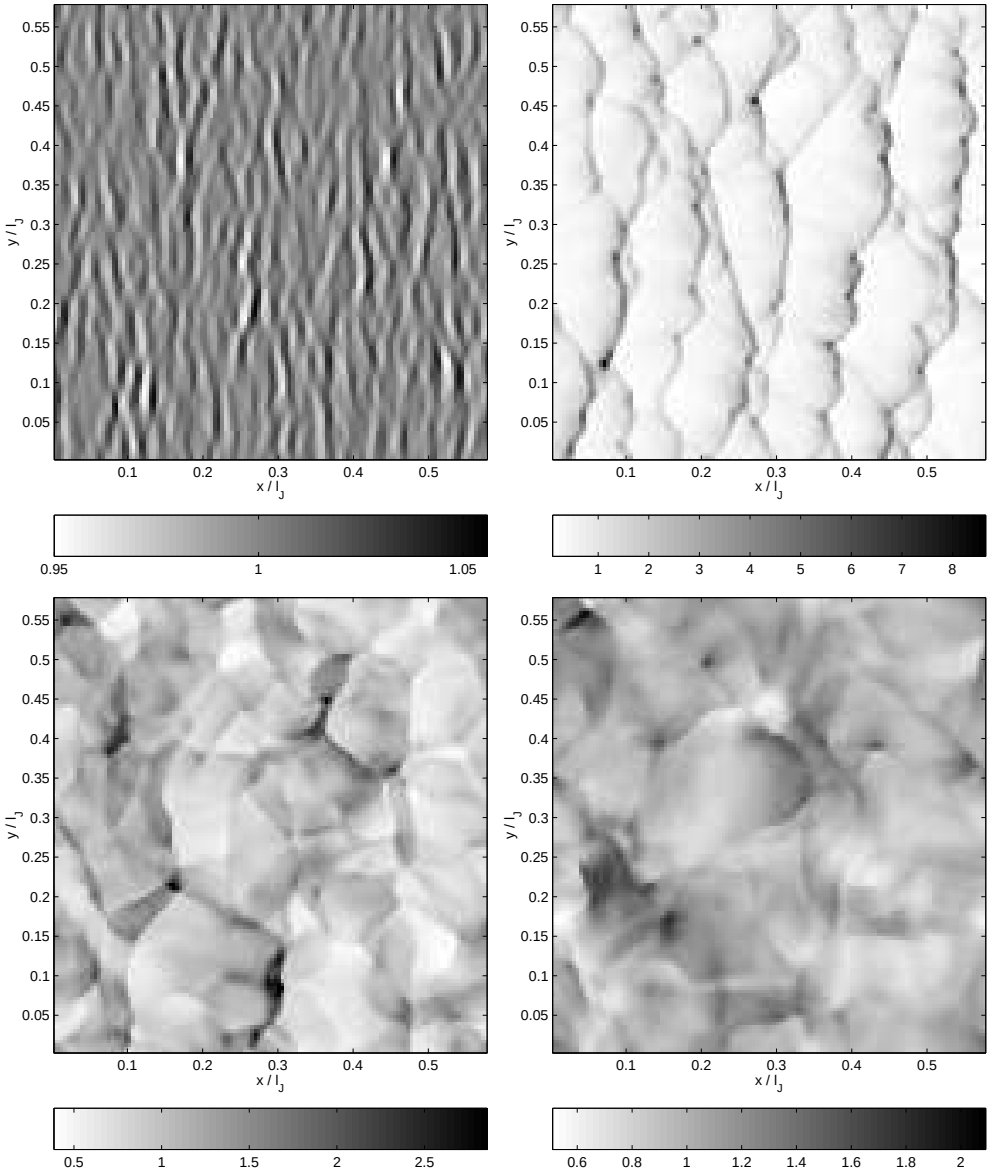

Figure 4. Snapshots of ρ in a 2D simulation for which the run parameters where: $\delta_0 = 10^{-6}$, $\Delta t/t_0 = 0.4$, $d/l_J = 0.576$, $\Delta v_0/c_T = 60$ corresponding to $\mathcal{G} = 20$. The size of the grid is 256×256. The time at which the snapshots were taken is $t = t_0 + \{\{0.5, 1\}, \{1.5, 4\}\}\Delta t$ (from top left to bottom right respectively).

do is to know the time elapsed since saturation had started. From it, we calculate the typical length scale (eq. 41) and the shock strengths ξ using the measured merging speed (eq. 40). Finally, we can estimate the variables using the semi-analytic results of the 1D saturated solution for the given ξ (fig. 2).

The evolution of several averaged variables are seen in fig. 5. The various stages of the evolution are clear – exponential growth, saturation when the amplitudes become

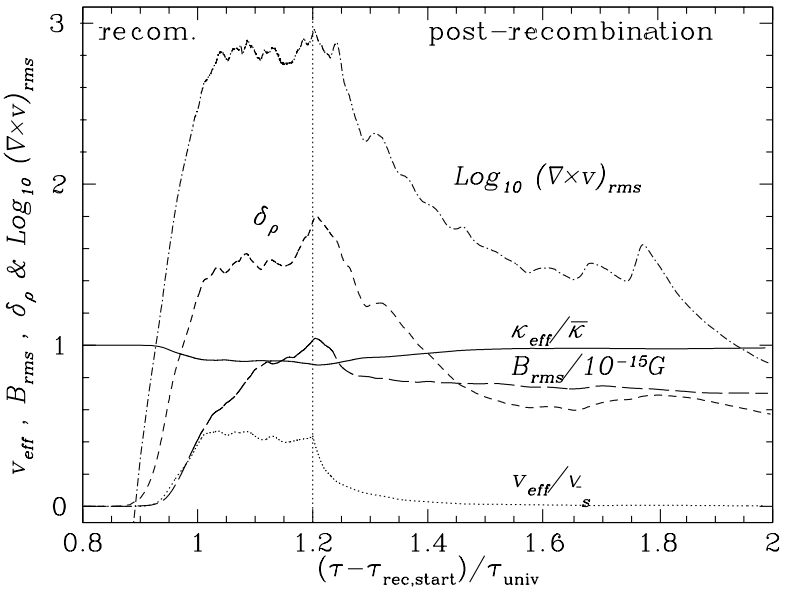

Figure 5. The (conformal) time evolution of several averaged variables in the simulation depicted in the previous figure. δ_ρ is the equivalent amplitude of a sinusoidal wave that has the same density r.m.s. Its initial exponential growth is replaced with a much slower evolution when it reaches saturation. After recombination ends, the shocks can start dissipating the energy and the amplitude falls. Note that this part is not accurate since after recombination, since the isothermal assumption breaks down. The other depicted variables exhibit similar evolution. These are the effective velocity with which the radiation perceives the baryon fluid, the effective opacity of the fluid, the vorticity and the magnetic field.

nonlinear (including a slight growth due to shock mergers) and subsequent dissipation. Another interesting property is the vorticity formed and the consequent magnetic field produced. This is extensively described in §7.

7. Formation of a magnetic field

In this section, we revive the old idea of Harrison of how to produce magnetic fields in the early Universe [19]. He argued that if a baryonic field has some vorticity (e.g. a rotating sphere of gas) then the radiation drag acting on the electrons alone would create a non-zero EMF (electro-motive force) which in turn would produce a magnetic field through Faraday's law:

$$\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}. \tag{43}$$

However, Harrison considered relatively large cosmological scales on which it is not possible to achieve high vorticities.

By contrast, we obtain in our simulations vorticity on scales smaller than the isothermal Jeans scale; therefore, the baryonic fluid can perform many revolutions during the decoupling era. As will be shown later in this section, the typical magnetic

field generated by the mechanism is given by:

$$B \sim \frac{4}{3}\frac{\sigma_T \sigma T_{\rm rec}^4}{c} n_{\rm rev} \sim 3 \times 10^{-16} \text{Gauss } n_{\rm rev} \tag{44}$$

where σ_T is the Thompson cross-section, σ is the Stephan-Boltzmann constant, $T_{\rm rec}$ is the temperature at recombination and $n_{\rm rev}$ is the number of revolutions that the fluid element performs during the decoupling era and shortly thereafter.

In the rest of this section, we derive eq. 44 and perform a 2D simulation of the magnetic field generation in shocked baryonic fluid. We will then try to estimate the effects that 3D MHD amplifications might have.

The source for a rotational electric field is the radiation drag experienced by electrons that is 2000 times larger than that experienced by the protons. This distinction between electrons and protons will produce a polarization and an electric field. To see this, we start with Ohm's law that is used to describe the flow of the electrons. It is permissible because the electron-proton relaxation time τ_{ep} is very short when compared to other time scales in the problem. The equation is:

$$\mathbf{j} = \frac{\tau_{ep} n_e e}{m}\left(\mathbf{f}_e + e\mathbf{E} + \frac{\mathbf{u}}{c} \times \mathbf{B}\right), \tag{45}$$

with $\mathbf{j}$ being the current density. n_e, e and m_e are the electron density, charge and mass and $\mathbf{f}_e$ is the radiation force exerted on each electron:

$$\mathbf{f}_e = \frac{4}{3}\sigma_T \sigma T^4 \frac{\mathbf{u}_r - \mathbf{u}}{c} \tag{46}$$

with $\mathbf{u}_r$ the radiation velocity, and $\mathbf{u}$ represents the material flow. By taking the curl of Ampère's law:

$$\nabla \times \mathbf{B} = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \tag{47}$$

and using Ohms equation, we find:

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{B}) = & \frac{1}{c}\frac{\partial(\nabla \times \mathbf{E})}{\partial t} + \frac{4\pi}{c}\frac{\tau_{ep} n_e e}{m_e}(\nabla \times \mathbf{f}_e) \\ & + \frac{4\pi}{c}\frac{\tau_{ep} n_e e^2}{m}\left(\nabla \times \mathbf{E} + \nabla \times \left(\frac{\mathbf{u}}{c} \times \mathbf{B}\right)\right). \end{aligned} \tag{48}$$

Using Faraday's equation we have:

$$\begin{aligned} & \nabla(\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B} + \frac{1}{c^2}\frac{\partial^2 \mathbf{B}}{\partial t^2} + \\ & \frac{4\pi}{c^2}\frac{\tau_{ep} n_e e^2}{m}\left(\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B})\right) = \frac{4\pi}{c}\frac{\tau_{ep} n_e e}{m}(\nabla \times \mathbf{f}_e). \end{aligned} \tag{49}$$

We see that to form a magnetic field we need a non conservative force $\mathbf{f}_e$ to exist. The high conductivity of the plasma implies that the first three terms can be neglected (in fact, they are 10 order of magnitudes smaller). We thus have:

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B}) = \frac{e}{c}\nabla \times \mathbf{f}_e = \frac{4}{3e}\sigma_T \sigma T^4 \nabla \times (\mathbf{u}_{\rm rad} - \mathbf{u}). \tag{50}$$

Since the photon wind originates from Horizon-scale perturbations and it isn't perturbed, we have that $\nabla \times \mathbf{u}_{\rm rad} = 0$.

If we define a "magnetic source function" $\mathcal{B}$ as:

$$\mathcal{B}(t) \equiv \frac{4}{3e}\sigma_T \sigma T^4 = 3 \times 10^{-16} \times (T(t)/T_{rec})^4 \text{ Gauss}, \tag{51}$$

we can write down the simplified equation of motion for the magnetic field:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla_r \times (\mathbf{u} \times \mathbf{B}) - \mathcal{B}(\nabla_r \times \mathbf{u}), \tag{52}$$

where we have specifically written here that the differentiation is with respect to a non-expanding frame. In a frame co-moving with the expansion of the universe, we write $\mathbf{u} = \mathbf{v} + \dot{a}\mathbf{x}$ and find:

$$\frac{\partial \mathbf{B}}{\partial t} = \frac{1}{a}\nabla_x \times ((\mathbf{v} + \dot{a}\mathbf{x}) \times \mathbf{B}) - \frac{\mathcal{B}}{a}\nabla_x \times \mathbf{v}. \tag{53}$$

By switching to conformal time and defining a "comoving" magnetic field in the form $\tilde{\mathbf{B}} \equiv a^2\mathbf{B}$ (that actually has the dimension of a magnetic flux), the evolution of $\mathbf{B}$ is simplified to a form similar to that in a non expanding universe:

$$\frac{\partial \tilde{\mathbf{B}}}{\partial \tau} = \nabla \times (\mathbf{u} \times \tilde{\mathbf{B}}) - a_{rec}^2 \mathcal{B}_0 \left(\frac{a_{rec}}{a}\right)^2 (\nabla \times \mathbf{v}), \tag{54}$$

where $\mathcal{B}_0 \equiv \mathcal{B}(t = t_{rec})$ and the x index is omitted.

A magnetic field will be produced by the vorticity of the baryonic fluid (the last term) and it will be advected by the plasma (the second term). By assuming that the typical length scales involved are of order the distance that isothermal waves can traverse during recombination, we have that $k^{-1} \sim \tau c_T$ and therefore the typical magnetic fields will be of order $B_{\text{typical}} \sim \int \mathcal{B}\nabla \times \mathbf{v} d\tau \sim \mathcal{B}_0 n_{\text{rev}}$, where $n_{\text{rev}} \sim \mathcal{O}(1)$ is the typical number of revolutions of a fluid element during decoupling (cf eq. 44). To conclude, magnetic fields of order a few times 10^{-16} Gauss are expected. Vorticity is however a complex quantity that is predominantly formed by oblique shocks or by shocks for which the shock parameters (such as their strength or angle) vary in a direction perpendicular to the flow†. Consequently, one has to resort to numerical simulations in order to get any quantitative predictions. Moreover, since vorticity is formed only in 2 or more dimensional shocks (it cannot be described by a 1D flow), a 2D simulation is required as a minimum. One should however still be cautious with its prediction because 2D and 3D MHD dynamics have several qualitative differences. In particular, turbulent amplification does not exist in 2D. To see how the vorticity is formed, we take the curl of the momentum equation to get:

$$\frac{d[\nabla \times \mathbf{u}]}{dt} = -\nabla \times \mathbf{f}_{\text{rad}} = -\frac{\nabla \tau_{e\gamma} \times \mathbf{u}_{\text{rad}}}{\tau_{e\gamma}^2} - \frac{\nabla \times \mathbf{u}}{\tau_{e\gamma}} \tag{55}$$

where $\mathbf{f}_{\text{rad}}$ is the radiative force per unit mass. We see that even if $\nabla \times \mathbf{u} = 0$, there is a source term if the gradient in the density (and therefore also $\tau_{e\gamma}$ through the opacity) is not in the direction of the radiation "velocity" in the material rest frame. This will be the case in oblique shock waves.

The equations for the magnetic field can be further simplified for the 2D simulations. Using that $\nabla \cdot \mathbf{B} = 0$ and that $\mathbf{B} = B\hat{\mathbf{z}}$, the advective term in eq. 54 becomes:

$$\nabla \times (\mathbf{v} \times \mathbf{B}) = -\nabla \cdot (\mathbf{v}B)\hat{\mathbf{z}}. \tag{56}$$

† By taking the curl of eq. 17 with eq. 21 (e.g. [20], §114), one can see that a small negligible vorticity will be formed even before the formation of shocks.

Namely, one has the following equation for the $\hat{z}$ component of the magnetic field:

$$\frac{\partial \tilde{B}}{\partial \tau} = -\nabla \cdot (\mathbf{v}\tilde{B}) - a_{rec}^2 \mathcal{B} \left(\frac{a_{rec}}{a}\right)^2 (\nabla \times \mathbf{v}) \,. \tag{57}$$

Like the other hydrodynamic equations, the numerical solution is achieved through dimensional splitting. The two steps are then given by:

$$\begin{aligned} 1: \quad & \frac{\partial \tilde{B}}{\partial \tau} = -\frac{\partial \cdot (v_x \tilde{B})}{\partial x} + a_{rec}^2 \mathcal{B} \left(\frac{a_{rec}}{a}\right)^2 \frac{\partial v_y}{\partial x} \\ 2: \quad & \frac{\partial \tilde{B}}{\partial \tau} = -\frac{\partial \cdot (v_y \tilde{B})}{\partial y} - a_{rec}^2 \mathcal{B} \left(\frac{a_{rec}}{a}\right)^2 \frac{\partial v_x}{\partial y} \end{aligned} \tag{58}$$

Typical results of the numerical simulations can be found in figure 6. The typical magnetic fields obtained are of order 10^{-15}Gauss. However, one should bear in mind that 3D effects of twisting and folding of the magnetic field can amplify it and that the 2D simulations that we carried out are uncapable of simulating these effects entirely. The typical magnetic fields produced by the end of recombination can be fitted by the rough formula (cf [9]):

$$B_1 \equiv \langle B_{\mathrm{rms},1} \rangle (\tilde{\mathcal{G}}) \approx (1 \pm 0.4)\tilde{\mathcal{G}} \times 10^{-16}\mathrm{Gauss} \quad \text{for} \quad \tilde{\mathcal{G}} \lesssim 12. \tag{59}$$

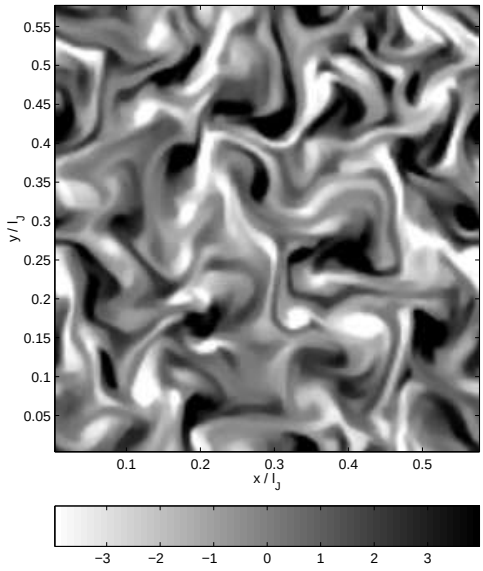

Figure 6. A snapshot of the magnetic field for the 2D simulation depicted in figure 4. The snapshot was taken when $t = t_0 + 1.5\Delta t$. The units of the magnetic field are in $\mathcal{B}_0 = 3 \times 10^{-16}$Gauss

7.1. 3D effects

3D effects are known to be important to the evolution of turbulent MHD. Unlike 2D systems, 3D systems include "twisting" and "folding" which can amplify the fields if they are below their equipartition value. We *conjecture* here that he shocked fluid will

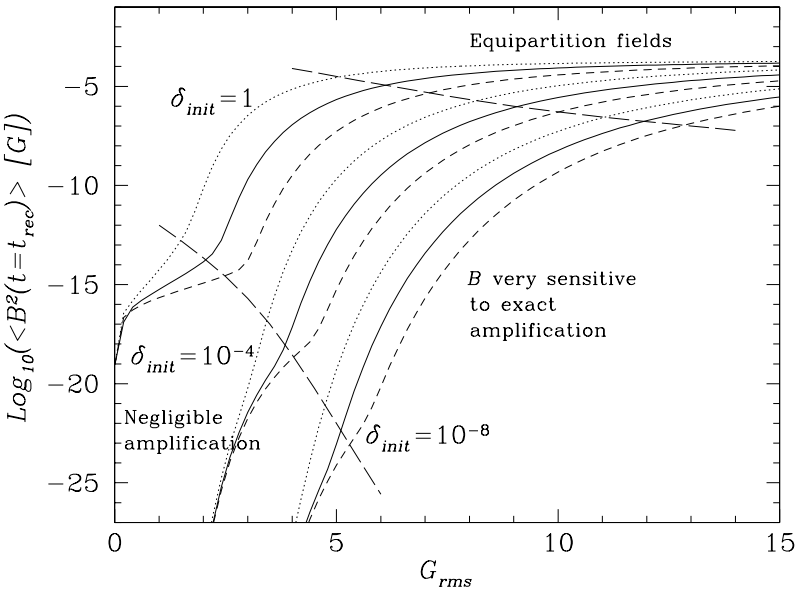

Figure 7. The average magnetic field obtained as a function of $\mathcal{G}_{\rm rms}$ for cosmologies with a Gaussian distribution for the initial large-scale perturbation spectrum. The three different curves are for different initial small-scale amplitudes. The solid lines assume that the 3D amplification is $\propto \exp(\mathcal{T})$. The dashed lines assume $\exp(\mathcal{T}/2)$ while the dotted lines assume $\exp(2\mathcal{T})$. They depict the theoretical uncertainties in the exponential growth. Since the universe expands, a factor of $(z/z_{rec})^2$ gives the dilated field at z. The results above the lower long-dashed line all depend on whether 3D amplification does take place. Since without 3D simulation it is only a conjecture, the results above this line should be taken cautiously.

develop full 3D turbulence and that it will result with 3D turbulent MHD amplification that is known to take place in other systems (e.g. [21]). It should therefore be taken cautiously until it will be proved by a full 3D numerical investigation.

To quantify the importance of the 3D effects, we define a "twisting" number as:

$$\mathcal{T} \equiv \frac{1}{2\pi}\int_{\tau_0}^{\tau_{\rm end}} \langle \nabla_x \times \mathbf{v}\rangle_{\rm rms}\, d\tau. \tag{60}$$

Although it is impossible to know exactly by how much the field will actually grow without a detailed 3D simulation, a rough guesstimate would be that the magnetic field grows by a factor of order $\epsilon\mathcal{T}$ e-folds, where $\epsilon \sim \mathcal{O}(1)$ is the amplification efficiency. This arises because the expected growth in MHD turbulence is exponential with a typical time scale of order the eddy turn over time (e.g. §17.6 in [21]), and the twist parameter as defined is roughly the number of revolutions or eddy turn overs.

By plotting the twist number [9], one finds values that vary from a few for small $\tilde{\mathcal{G}}$ to more than 30 for $\tilde{\mathcal{G}} > 10$. Its behavior can be approximated by the fitting formula:

$$\mathcal{T}(\tilde{\mathcal{G}}) \approx 1.5\tilde{\mathcal{G}} + 1.2\times 10^{-3}\tilde{\mathcal{G}}^4 \quad \text{for} \quad \tilde{\mathcal{G}} \lesssim 12. \tag{61}$$

Clearly, when $\tilde{\mathcal{G}} \gtrsim 10$, the amplification will be so large that the *equipartition* value of the magnetic field can be reached. The equipartition will be with the local acoustic

density, that is†:

$$B_{\rm eq} \sim \sqrt{4\pi\bar{\rho}c_T^2} \sim 2\times 10^{-4}{\rm Gauss}\left(\frac{z}{z_{\rm rec}}\right)^2. \tag{62}$$

To reach this values, a growth of typically 25 e-folds should take place (From typical values of $\mathcal{B}$ to $B_{\rm eq}$).

We proceed now to estimate the average magnetic field produced as a result of Harrison's mechanism and 3D turbulent amplification. We use $B_1(\tilde{\mathcal{G}})$ from eq. 59 for the seed field generated by the shocks and the 3D turbulent e-fold twist parameter $\mathcal{T}(\tilde{\mathcal{G}})$ from eq. 61. The average magnetic field is then given by:

$$\langle B_{\rm rms}\rangle \approx \int_0^{\tilde{\mathcal{G}}_{\rm sat}} P(\tilde{\mathcal{G}}-\ln\delta_0)B_1(\tilde{\mathcal{G}})\exp(\epsilon\mathcal{T}(\tilde{\mathcal{G}}))d\tilde{\mathcal{G}} \tag{63}$$

$$+\int_{\tilde{\mathcal{G}}_{\rm sat}}^{\infty} P(\tilde{\mathcal{G}}-\ln\delta_0)B_{\rm eq}d\tilde{\mathcal{G}}. \tag{64}$$

$P(\mathcal{G}) = P(\tilde{\mathcal{G}}-\ln\delta_0)$ is the probability density for the growth factor $\mathcal{G}$ (arising from the variability of the large scales). If the fluctuations are Gaussian, then $P(\mathcal{G})$ is given by $\sqrt{54/\pi}(\mathcal{G}^2/\mathcal{G}_{\rm rms}^3)\exp[-(3/2)(\mathcal{G}/\mathcal{G}_{\rm rms})^2]d\mathcal{G}$ (cf [8]). $\tilde{\mathcal{G}}_{\rm sat}$ is the growth factor above which the magnetic field is amplified enough as to reach the equipartition value. It is given by the solution of the equation:

$$B_1(\tilde{\mathcal{G}}_{\rm sat})\exp(\epsilon\mathcal{T}(\tilde{\mathcal{G}})) = B_{\rm eq}. \tag{65}$$

The values found for $B_{\rm rms}$ vs. the growth factor $\tilde{\mathcal{G}}$ are depicted in figure 7. The magnetic field shown is the field produced shortly after decoupling, at $z\sim 1000$. Since the expansion of the universe dilates the magnetic field, an additional factor of $(z/z_{rec})^2$ should be used for finding a later field strength.

Cosmologically interesting magnetic fields (e.g., as seeds for galaxy formation) are of order 10^{-15}Gauss at recombination (corresponding to 10^{-19}Gauss at $z\sim 10$, when galaxies form); whence, above a critical perturbation amplitude (corresponding to $G_{\rm rms}\sim 4$ for $\delta_0\sim 10^{-4}$ or $G_{\rm rms}\sim 6$ for $\delta_0\sim 10^{-8}$) a cosmologically interesting field will be produced in the "shocked" part of the Universe. However, its exact amplitude is generally hard to predict accurately since the exact amplification factor is unknown. Only if $\mathcal{G}_{\rm rms}$ is large enough to reach values close to the equipartition (if $\mathcal{G}_{\rm rms}\gtrsim 10$) is it easier to predict the field.

8. Globular cluster size structure from nonlinear evolution at decoupling

The numerical simulations have shown that nonlinear structure in the form of shocks can form by the time recombination ceases if the local $\mathcal{G}$ is large enough (greater than $\sim -\ln(\delta_0)$. Nevertheless, the typical length scale of this structure (i.e., the typical distance between the neighboring shock fronts) is roughly 10 times smaller than the isothermal Jeans scale. It arises from the finite duration of recombination ($\Delta a/a\sim 1/6$) and the finite speed with which the shocks can propagate (at most a few times the speed of sound). Consequently, gravitationally unstable baryonic structure does not form until a later epoch.

† The field obtained by equipartition at a redshift z or the field obtained by equipartition at an earlier epoch and a subsequent dilation through the Universe's expansion are the same.

In this section, we attempt to estimate the mass and the time of formation of the first gravitationally unstable objects. There are two different scales on which the first gravitationally unstable objects can form. The first is the sub-Jeans scales which corresponds to the average distance between shocks and that can collapse only after the evolving co-moving Jeans length had decreased to this scale.

The second preferred scale is the Jeans scale. The typical perturbations at this scale will be small but their collapse initiates immediately after recombination. If the density power spectrum of the shocked fluid does not fall too steeply for small k's, then the Jeans scale at recombination could be the formation scale of the first objects.

The two preferred scales that can therefore form are the Jeans scale and the typical "structure" scale. We assume that the structure formed is predominantly that of baryons. The collapse is therefore of the baryon fluid which can subsequently drag in any cold dark matter component†. As we shall see, the preferred scale will be determined by the spectrum of $|\delta_k|$. If collapse proceeds on a given scale λ, it will proceed on all the larger scales as well. Consequently, we are interested in the total fluctuations on all scales larger than a given one. We therefore define the dimensionless variance σ^2 for the variability above a scale λ (or below $k = 2\pi/\lambda$) as:

$$\sigma^2(\lambda) \equiv \int_{|\mathbf{k}|<2\pi/\lambda} \delta_k(\mathbf{k}) d\mathbf{k}. \tag{66}$$

8.1. Collapse of the baryon fluid on the typical structure scale

If by the end of recombination, there are perturbations with $\sigma \sim 1$ over a sub-Jeans scale of $\lambda_{\rm typ}$, then these perturbations collapse only after the Jeans mass had reduced sufficiently. Since the Jeans co-moving scale evolves as $z^{1/2}$, it is evident that the structure will start to collapse at:

$$1 + z_b \approx 1000 \left(\frac{\lambda_{\rm typ}}{\lambda_J}\right)^2. \tag{67}$$

However, if the baryons are not the major component of the cosmological soup (i.e., if $\Omega_b/\Omega_{\rm matter} < 1$, then gravitational nonlinearities will set it only when z has decreased by a further factor of $\Omega_b/\Omega_{\rm matter}$, namely, the z for which gravitational nonlinearities commence will be:

$$1 + z_c \approx 1000 \left(\frac{\lambda_{\rm typ}}{\lambda_J}\right)^2 \frac{\Omega_b}{\Omega_{\rm matter}}. \tag{68}$$

The additional factor of $\Omega_b/\Omega_{\rm matter} < 1$ arises because gravitational nonlinearities need nonlinear perturbations in the total density field and not just that of the baryons.

From our understanding of the numerical results, it is clear that the typical length scale will be at most $\lambda_{\rm typ} \lesssim c_T \Delta t$; that is, the largest typical scales expected are at most $c_T t/6 \sim \lambda_J/12\pi$. Therefore, $1 + z_b$ is at most of order unity but could be less. This structure formation path is apparently not very efficient because of the long time needed to wait until the Jeans scale is reduced sufficiently.

One should nevertheless be cautious because we have implicitly assumed that in the long duration it takes the Jeans scale to be reduced to the shocks' length scale, nothing

† The full analysis of the modes which includes cold dark matter for mass scales smaller than the Jeans scale shows that besides the oscillating baryonic mode, there is a mode which is largely the collapse of CDM. Although it starts its collapse during recombination, its amplitude can be shown to be too small to be of any importance.

interesting has taken place in the shocked fluid. However, since it is already with nonlinear density variations, processes like cooling and a consequent earlier collapse could take place.

8.2. Collapse of the baryon fluid on the Jeans scale

Structure which is on the scale of the Jeans mass at recombination will start to collapse immediately. However, it will start with an amplitude σ that is smaller than unity. Since linear structure grows linearly with z (for a matter dominated, flat universe), one has that the baryons will reach a nonlinear amplitude at a redshift of:

$$1 + z_b \approx 1000\sigma_{\lambda_J}. \tag{69}$$

And the gravitational nonlinearities will set in at:

$$1 + z_c \approx 1000\sigma_{\lambda_J}\frac{\Omega_b}{\Omega_{\rm matter}}. \tag{70}$$

Evidently, if σ behaves as λ^{-p} with $p < 2$, then forming Jeans scale objects will be more favorable than sub-Jeans objects.

To estimate the z at which gravitationally bound objects will form in this picture, we need to know the amount of power in baryonic density fluctuations on the recombination Jeans scale. The amount of power depends on the length (or time) scale, and on $\tilde{\mathcal{G}}$. By plotting the slope of $\sigma(k)$ for several cases, we conclude that $\sigma(k) \propto k^{1/2}$ for k of order the (inverse) Jeans scale. Thus, if we wish to scale the amount of structure for runs with a different time scale, we should fit $\sigma(k)/\Delta t^{1/2}$, because $k \propto \lambda^{-1} \propto \Delta t^{-1}$. We find: $\sigma/(\Delta\tau/0.1\tau_0)^{1/2} \approx (3 \pm 1) \times 10^{-3}\tilde{\mathcal{G}}$. The typical redshift at which structure on the recombination Jeans scale will form is:

$$z_c \approx (3 \pm 1)\tilde{\mathcal{G}}\left(\frac{\Delta\tau}{0.1\tau_0}\right)^{-1/2}\frac{\Omega_b}{\Omega_{\rm matter}}. \tag{71}$$

When we compare it to eq. 68, it is evident that structure will typically form at a higher z on the recombination Jeans scale than on the shocks' scale.

One of the interesting implication of structure the gravitationally unstable structure formed is reionization of the galactic medium. To achieve this, it is sufficient to have 10^3 of the baryonic material collapse [22] which corresponds to a 3σ fluctuation. Roughly 10^{-3} of the baryons will form Jeans scale structure at a red shift of:

$$1 + z_{c,3\sigma} \approx (9 \pm 3)\tilde{\mathcal{G}}\left(\frac{\Delta\tau}{0.1\tau_0}\right)^{-1/2}\frac{\Omega_b}{\Omega_{\rm matter}}. \tag{72}$$

This is the most important equation of this section. It implies that for a local $\tilde{\mathcal{G}} \gtrsim \Omega_b/\Omega_{\rm matter}$, reionization through the gravitational collapse of the shocked fluid will take place at $z \gtrsim 10$, and it could be important for cosmological evolution.

We can also estimate the volume fraction of the universe in which the most of the mass collapsed into globular-cluster-size ($M \sim 10^5 M_\odot$) objects by a given z. If we assume that $\Delta\tau/0.1\tau_0 \approx 1$, we have that for a given z, the minimal $\tilde{\mathcal{G}}$ required to have most of the mass collapse is:

$$\tilde{\mathcal{G}}_{\rm min} \approx (z/3)(\Omega_{\rm matter}/\Omega_b). \tag{73}$$

If we are only interested in 3σ fluctuations, then the required $\tilde{\mathcal{G}}$ is smaller:

$$\tilde{\mathcal{G}}_{{\rm min},3\sigma} \approx (z/9)(\Omega_{\rm matter}/\Omega_b). \tag{74}$$

The fraction of regions having this required $\tilde{\mathcal{G}}$ is $\int_{\tilde{\mathcal{G}}_{\rm min}}^{\infty} P(\mathcal{G}+\ln\delta_0)d\tilde{\mathcal{G}}$. It is the function $\mathcal{F}(\mathcal{G}_{\rm rms},\delta)$ of the fraction of regions that reach nonlinear amplitudes with an effectively smaller initial amplitude δ of $\delta_0\exp(-\tilde{\mathcal{G}}_{\rm min})$. This allows to calculate, using equation 32, the fraction of the universe's volume that will contain collapsed objects by a given redshift z (either by most of the fluid using $\tilde{\mathcal{G}}_{\rm min}$, or of 3σ objects using $\tilde{\mathcal{G}}_{{\rm min},3\sigma}$).

Unless matter in the universe is baryonically rich, i.e., unless $\Omega_b/\Omega_{\rm matter}$ is not a small fraction or unless $\mathcal{G}_{\rm rms}$ is somewhat higher than what is inferred from the CMBR (if $\mathcal{G}_{\rm rms}\gtrsim 10$), then $10^5 M_\odot$ gravitationally unstable object will form relatively late, later than predicted by other CDM models (e.g., Haiman & Loeb 1997, [22]).

On a more speculative note, there still could be several unaccounted effects that could change the expected z at which gravitationally bound objects form. For example, if the original perturbation spectrum is one which corresponds to a $\mathcal{G}_{\rm rms} > 5\Omega_{\rm matter}/\Omega_b$ then the reionization that will necessarily take place at high redshifts ($z\gtrsim 50$) will rescatter the CMB photons, homogenize it, and have it appear to have a smaller power spectrum and a smaller inferred $\mathcal{G}_{\rm rms}$.

9. Direct Effect on the CMBR – Compton scattering

The unstable small-scale isothermal modes have typical length scales which are at least 5 orders of magnitudes smaller than the horizon scale. This implies that direct observations of CMBR spatial variability arising from the small scales will not be observable in the near future. The resolution needed is simply too small to be resolved. Nonetheless, since small-scale structure is expected to form in regions correlated with large scales (specifically where the temperature or density gradients are large – where the radiative fluxes will be important), it can leave a detectable imprint. Since nonlinear structure will form shocks, their dissipation after recombination will partially reionize the medium. This can be detectable as a non vanishing Compton scattering y-parameter that distorts the CMBR energy spectrum.

Once the small-scale waves saturate, the energy content in them is $\delta_{\rm kin}^2 c_T^2/2$. $\delta_{\rm kin}$ is related to ξ by $\delta_{\rm kin}\approx 0.5\xi^{1/2}$ for $\xi\gtrsim 10$ (cf fig. 2). At the end of recombination one has $\xi\approx 2\tilde{\mathcal{G}}f_\lambda/\alpha\approx 3.5\tilde{\mathcal{G}}$. Therefore, an energy of $0.4\tilde{\mathcal{G}}c_T^2$ (per unit mass) is there waiting to be dissipated in shocks. This energy is of order the internal energy of the baryons.

At the time of recombination, the heat capacity is very large because a large amount of energy is needed to reionize the Hydrogen. Thus, after recombination, when the temperature is of order but less than the recombination temperature $T\lesssim T_{rec}\approx 0.26eV$, injecting an amount of order the thermal energy of the plasma will ionize at most only $\lesssim 0.4\tilde{\mathcal{G}}kT_{rec}/E_{\rm ion}\approx 0.4\tilde{\mathcal{G}}0.26eV/13.6eV\approx\tilde{\mathcal{G}}/125$ of the Hydrogen atoms. The free electrons released by ionization rescatter photons and leave a mark in the CMBR.

Heating the plasma back to the recombination temperature implies that the plasma and the microwave background are not in equilibrium anymore. The plasma consequently upsets the Planck distribution of the radiation through the number conserving Compton scattering process. The process is evident from Kompaneets' equation (1957, [23]) and its cosmological application by Sunyaev and Zel'dovich (1969, [24]). The dimensionless parameter describing the importance of the effect

is the parameter y defined as:

$$y = \frac{k \langle T_e \rangle}{m_e c^2} \int^t \sigma_T n_e c dt, \tag{75}$$

with standard notations. n_e denotes the number of *free* electrons. The analysis of the COBE data revealed no deviation from a Planckian spectrum and set a limit of (Fixsen et al. 1996, [25]):†

$$y \lesssim 1.5 \times 10^{-5} \quad (95\% CL). \tag{76}$$

If the fraction of ionized electrons increases to $\iota_e \sim \tilde{\mathcal{G}}/125$, and the temperature T increases to τT with $\tau \sim 1$, one can show that the resulting y parameter should be of order (e.g. Peebles, 1993 [15]):

$$y \sim 2 \times 10^{-11} (\tau - 1)(1+z)^{5/2} h \Omega_b \Omega^{-1/2} \iota_e. \tag{77}$$

with $z \sim 1000$ at recombination. Thus,

$$y_{max} \sim 1.5 \times 10^{-6} h_{0.75} \Omega_{b0.05} \tilde{\mathcal{G}} \Omega^{-1/2}. \tag{78}$$

The aforementioned value for the y parameter is however only in regions (and directions in the sky) where the gradient at recombination was large enough to have isothermal waves reach saturation. The average y will be smaller and given by:

$$y_{avr} = y_{max} \mathcal{F} \left(\frac{\log \delta_0}{\mathcal{G}} \right). \tag{79}$$

Unless the fraction $\mathcal{F}$ of the sky that reaches nonlinearity is of order unity, the y_{avr} is too small to be detectable by present means. Nevertheless, investigations that correlate y with the gradients in future sub-degree maps will be able to detect smaller values for y.

If the observations rule out the existence of such structure then the theory imposes strict limits on the cosmological parameters, including the fact that the fluctuation distribution function does not have a power law tail. However, it can predict black-body deformations correlated with the CMBR gradients (i.e., a specific correlation). If indeed this is the case, it will be proven as the origin for small scale structure, and it could be used for setting strict limits on cosmological parameters.

10. Indirect Effect on the CMBR – The damping of large scales

Besides affecting the CMBR directly, the formation of nonlinear structure will have an indirect effect as well. The observed CMBR power spectrum is the radiation left by the acoustic oscillations at recombination. Instead of changing this radiation after it was last scattered at recombination, one could also envision a possibility that the large-scale acoustic oscillations themselves are affected. Specifically, if large amounts of energy can be damped from the large scales into the small scales through the amplification process, the large scales could be damped.

† Note that a smaller upper limit of $y \lesssim 3 \times 10^{-6}$ (Fixsen et al. 1997) was found when the signal was correlated with the DMR fluctuation map, however, this type of limit assumes that y is correlated with the temperature fluctuations, in our case however, the correlation will be with the temperature gradients.

Although a full treatment of the feedback on the large scales is still missing, one can show using a very simplified picture how the large scales can be affected.

Since energy is conserved, large scales will be affected if the energy transferred to the smaller scales is comparable to the original energy in the large-scale perturbations. The energy per unit mass in the large-scale waves is roughly given by $(\delta\rho_a/\rho)^2 c_A^2/2$ while the energy per unit mass in the small-scale isothermal waves is roughly $(\delta\rho_i/\rho)^2 c_T^2/2$. Inspection of eq. 12 reveals that when the small isothermal waves have a dimensionless amplitude of order unity over a large fraction of the universe and $\mathcal{G}_{\rm rms} \sim 1$, the energy content in both types of waves is comparable. A significant fraction of the energy in the large scales can therefore be transferred to the small scales.

Using the terminology and approximations of §2, it is evident that the energy in the large scales (per unit mass) can be expressed in the form $e = \delta_a^2 c_A^2/2 \equiv A_l^2 \mathcal{G}_{\rm rms}^2 c_T^2/2$. $\mathcal{G}_{\rm rms}$ serves here as the amplitude of the large scales, while A_l is a constant of order unity which normalizes the energy content in the large scales to $\mathcal{G}_{\rm rms}$. c_A and c_T are the adiabatic and isothermal speeds of sound.

During the exponential growth, the energy in the small scales is negligible when compared to the large scales, thus, energy from the large scales will only be damped after the small scales have saturated.

At a time t after recombination, one has:

$$\xi \approx \frac{2\mathcal{G}_{\rm rms}}{\alpha} f_\lambda \frac{t - t_{\rm sat}}{\Delta t} \equiv \frac{2\mathcal{G}_{\rm rms}}{\alpha} f_\lambda \bar{t}, \tag{80}$$

where $t_{\rm sat}$ is the time saturation is reached, Δt is the duration of recombination, and $\bar{t}$ is a dimensionless time since saturation.

The damping rate is therefore given by:

$$\frac{de}{dt} \approx -\frac{d}{dt}\left(\frac{1}{2}\delta_{\rm rad}^2 c_T^2\right) \approx -\frac{\mathcal{G}_{\rm rms}(\xi)}{\Delta t}\delta_{\rm rad}^2 c_T^2 \approx -\frac{\mathcal{G}_{\rm rms}^{3/2} c_T}{\Delta t}\bar{t}^{-1/2}, \tag{81}$$

where $\delta_{\rm rad} \approx 1.1\xi^{1/4}$ for $\xi \gg 1$ is the radiation damping amplitude found from the 1D saturation solution (cf fig. 2). Or in other words, it is the equivalent amplitude of a sinusoidal wave that results with the same transfer of energy from the large scales. Using the energy in the large scales $e \approx A_l^2 \mathcal{G}_{\rm rms}^2 c_T^2/2$, we find:

$$\frac{d\mathcal{G}_{\rm rms}}{d\bar{t}} = -\frac{2\mathcal{G}_{\rm rms}^{1/2}\bar{t}^{1/2}}{A_l^2}. \tag{82}$$

If integrated, we find:

$$\mathcal{G}_{\rm rms}(\bar{t})^{1/2} = G(\bar{t}=0)^{1/2} - \frac{8}{3A_l^2}\bar{t}^{3/2}. \tag{83}$$

The result is valid only after saturation has reached ($\bar{t} > 1$) and that we are in the strong shock regime for which $\delta_{\rm rad} \approx 1.1\xi^{1/4}$ is a good approximation. If $\mathcal{G}_{\rm rms}$ decreases enough for the system to return back to the weak regime, then the damping rate will be much smaller and the system will "quench" itself off. This happens when $\xi \sim 10$ or when $\mathcal{G}_{\rm rms} \lesssim 1 - 3$.

Apparently, the exact value of A_l is crucial to whether $\mathcal{G}_{\rm rms}$ can actually be reduced significantly or not. If A_l is roughly unity or less, then $\mathcal{G}_{\rm rms}$ can be reduced during the

time of recombination to a significantly smaller value. At the end of recombination, $\bar{t} = 1 - \ln(1/\delta_0)/\mathcal{G}_{\rm rms,0}$ giving that:

$$\mathcal{G}_{\rm rms,1} = \left[\mathcal{G}_{\rm rms,0}^{1/2} - \frac{8}{3A_l^2}\left(1 - \frac{\ln(1/\delta_{init})}{\mathcal{G}_{\rm rms,0}}\right)\right]^2. \quad (84)$$

For example. if $\mathcal{G}_{\rm rms,0} = 30$ before recombination (corresponding to a power spectrum that is perhaps 5 to 10 times larger than what is inferred directly from the CMBR, and $A_l^2 = 1/2$ then the final $\mathcal{G}_{\rm rms}$ inferred from the CMBR would only be $G_{\rm rms,1} \sim 3$ if $\delta_0 \sim 10^{-4}$ and $G_{\rm rms,1} \sim 10$ if $\delta_0 \sim 10^{-8}$.

If A_l is somewhat larger than unity then the effect will not be important, as the reduction of $\mathcal{G}_{\rm rms}$ will not be significant since the larger scales have too much energy to be dampled. Clearly, a proper evolution of the large scales should be done in order to find whether energy could actually be damped. If it could, then a somewhat small $G_{\rm rms}$ today does not imply the the coupling instability did not take place. The instability could have been important enough as to erase the evidence that could have proved it to be important!

11. Discussion

Small-scale isothermal waves are potentially unstable in the presense of large-scale perturbations. They exhibit exponential growth if the amplitude of the large-scale perturbations that drive the instability is large enough. The extent to which the instability takes place was characterized by a local $\mathcal{G}$ - the number of e-folds that small-scale waves grow during the time of recombinations at a given location or by its spatial r.m.s average of $\mathcal{G}_{\rm rms}$. For large enough $\mathcal{G}_{\rm rms}$'s as is the case for some cosmological parameters, some regions in the universe have a local $\mathcal{G}$ that is large enough to develop this instability to the extent that the isothermal waves are amplified out of the linear regime. In some cosmologies they will occupy only a small fraction of the volume while in others they can occupy effectively all of it or none at all. For example, in an optimistic scenario in which $\mathcal{G}_{\rm rms} = 10$, with a Gaussian distribution, 40% of the universe will have $\mathcal{G} > 10$, almost 1% will have $\mathcal{G} > 20$ and 6×10^{-6} of its volume will have $\mathcal{G} > 30$. (It will be larger for a power-law distribution, as can be the case in a model in which topological defects are the source of the perturbations).

Under large enough local fluxes, small-scale isothermal waves can be amplified enough to reach amplitudes of order unity. The amplitude however cannot grow much further because the waves saturate through the formation shocks. Once they appear, they dissipate any additional energy driven into the waves in the amplification process. To understand this saturation process, the problem of a traveling shock-wave "train" was analyzed and its was found to have an analytical solution. It yields the properties of the saturated waves (e.g., profile, amplitude, or effective opacity). It was found that the larger the typical wavelength is, the stronger is the shock.

Using an isothermal hydrodynamic simulation, the 1D results where extended to understand the growth of the waves while in the nonlinear regime. It was found that shocks merge and the correlation distance between shocks grows with a velocity of $\sim c_T$.

Using these results, one can heuristically describe the evolution of the small scales and their effect on the average plasma properties (e.g., effective amplitude and velocity). This will prove to be important in the next work in which the proper feedback on the large scales will be taken into account. In the full problem, the large

scales should be evolved concurrently with evolution of the small scales. Namely, at each different location, a different hydrodynamic simulation for the small scales should be carried out. Since this task is impossible, the evolution is amendable only if the small scales are approximated with a simplified behavior.

Next, a 2D simulation was developed. It has shown that some of the basic properties of the 1D simulation adequately describe the more complex system as well. For example, a 1D cut of the 2D evolution in the direction of the driving flux resembles the 1D evolution. The properties missing from the 1D simulation are the behavior in the perpendicular direction as well as the formation of vorticity. It was found that the shocks have a correlation length in the perpendicular direction that is roughly 5 to 10 times larger than the typical scales in the horizontal direction. The expected 3D behavior is the generation of "walls" that are separated by typical distances of order $\lambda_J/10$ in the horizontal direction and which merge (or have typical perpendicular correlation lengths that is at least 5 times as large in the perpendicular directions). After recombination, the dissipation of the shocks smears this picture into a more smooth one.

The generation of considerable vorticity is an effect that appears only when the system is at least 2D and only once it has formed shocks. Although the vorticity dies away after the shocks dissipate because of the universe's expansion, it can generate an important pre-galactic magnetic field in the duration that it exists. The 2D simulation has shown that the typical magnetic field produced is larger than 10^{-15}Gauss, however, it is clear that the lack of a 3D simulation in this case is crucial because a 3D simulation will include the effects of twisting and folding that clearly lack in the 2D case and which can amplify the field. Since the number of twisting can be anything up to several dozen folds until the vorticity is quenched, amplifications which are large enough to reach equipartition fluids are not unreasonable. A meaningful pre-galactic magnetic field can be obtained.

The last question addressed is the formation of gravitationally collapsed objects from the nonlinear structure that has formed. Since the structure formed is smaller than the isothermal Jeans scale at recombination, it cannot collapse gravitationally immediately after recombination. If nonlinear effects do not play a role, then in baryonically rich cosmologies one could expect to have structure collapse already at $z \sim 10$. However, since the structure is already nonlinear on smaller scales, it isn't unreasonable that the nonlinear effects, such as the merging and cooling of dense blobs, could in fact have objects collapse earlier. This is an interesting topic which will be addressed elsewhere.

Another interesting possibility is that by growing nonlinearities one could damp energy from the large scales and have a smaller CMBR power spectrum and a smaller inferred $\mathcal{G}_{\rm rms}$. A different effect that could affect the CMBR is that by growing the instability, the universe could be reionized early enough as to rescatter the radiation, thereby homogenizing it. This too will have the tendency to mimic a smaller apparent $\mathcal{G}_{\rm rms}$. That is to say, the mode coupling effect has the tendency to erase its "traces" when the original value of $\mathcal{G}_{\rm rms}$ is slightly (by only a factor of a few) larger than what is directly inferred from the CMBR.

Besides the proper analysis of the feedback on the large scales, another important issue that should be addressed in future work is the alleviation of several assumptions made. For example, it was assumed that the cooling time scale is much smaller than all other scales in the system. This is not necessarily the case, especially towards the end of recombination. Interestingly, should the small scales waves have

had temperature perturbations similar to the density perturbations, one could have expected to increase $\mathcal{G}$ by a factor of 20! Another assumption is that the equilibrium given by the Saha equation does describe the number of free electrons and thus the opacity as well. Clearly, the variation in opacity will be reduced in fast oscillations that cannot follow the equilibrium values. It can however be increased by up to a factor of 2 for slow variation in cases where the number of free electrons is determined by the recombination rate instead of Saha's equation. This work is in progress.

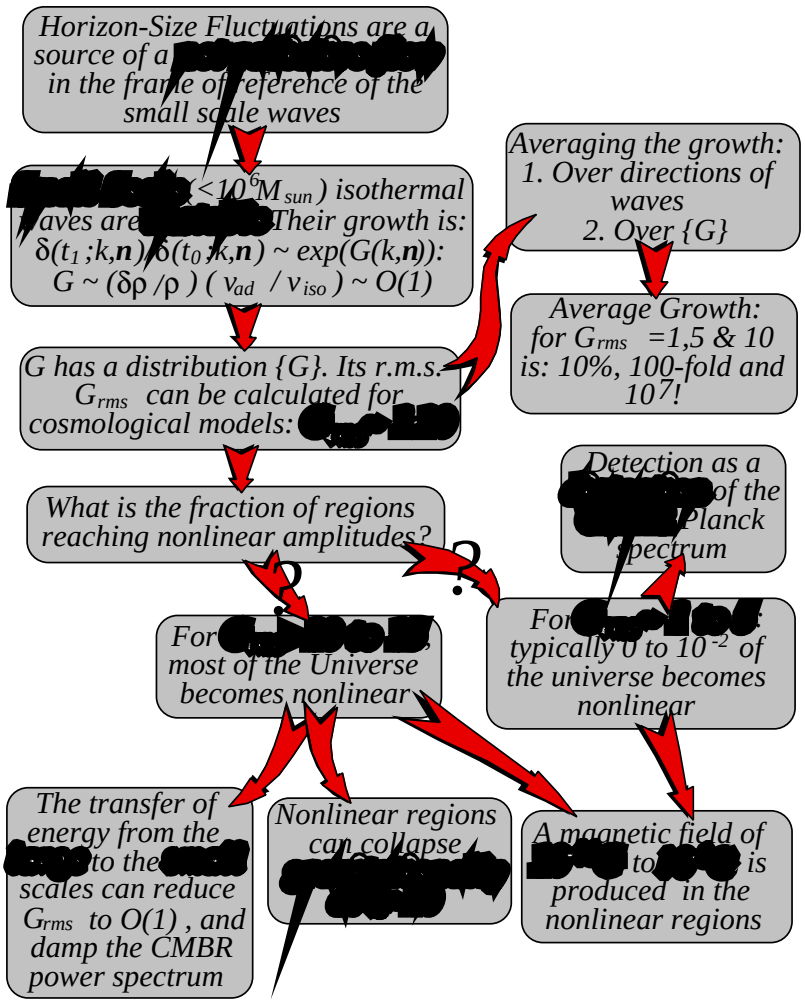

Figure 8. Summary of the effect of coupling at recombination and its consequences.

Acknowledgments

The author wishes to thank Caltech for the DuBridge Prize Fellowship support and the group of Boris et al. [18] for having their FCT code library available to the public.

References

[1] C.L. Bennett, A.J. Banday, K.M. Gorski, G. Hinshaw, P. Jackson, P. Keegstra, A. Kogut, G.F. Smoot, D.T. Wilkinson, & E.L. Wright, *Astrophys. J.* **464**, L1 (1996).

[2] E.M. Lifshitz, *J. Phys. USSR*, **10**, 116 (1946).

[3] S. Weinberg, *Gravitation and Cosmology*, New York: Wiley (1972).

[4] P.J.E. Peebles, *The Large-Scale Structure of the Universe*, Princeton: Princeton Univ. Press (1980).

[5] E.W. Kolb, & M.S. Turner, *The Early Universe*, Addison Wesley (1990).

[6] M. White, D. Scott, & J. Silk, *Annu. Rev. Astron. Astrophys.*, **32**, 319 (1994).

[7] J.R. Primack, in Proceedings of the Jerusalem Winter School, edited by A. Dekel & J.P. Ostriker, Cambridge: Cambridge Univ. Press, chap. 1. Also as astro-ph/9707285 (1997).

[8] N.J. Shaviv, *Mon. Not. Roy. Ast. Sc.* **297**, 1245 (1998).

[9] N.J. Shaviv, & Y. Levin, *Mon. Not. Roy. Ast. Sc.* , submitted (1999).

[10] A.G. Hearn, *Astron. & Astrophys.* **19**, 417 (1972).

[11] R.G. Carlberg, *Astrophys. J.* **241**, 1131 (1980).

[12] A. Feldmeier, J. Puls, & A.W.A. Pauldrach, *Astron. & Astrophys.* **332**, 878 (1997).

[13] M. Asplund, *Astron. & Astrophys.* **330**, 641 (1998).

[14] L. Mestel, & D.W. Moore, *Astron. & Astrophys.* **156**, 121 (1986).

[15] P.J.E. Peebles, "Principles of Physical Cosmology", Princeton university press, Princeton, p. 179, 635 (1983).

[16] A.G. Hearn, *Astron. & Astrophys.* **23**, 97 (1973).

[17] J.P. Boris, & D.L. Book, *J. Comput. Phys.* **11**, 38 (1976).

[18] J.P. Boris, A.M. Landsberg, E.S. Oran, & J.H. Gardner, *LCPFCT - A Flux-Corrected Transport Algorithm for Solving Generalized Continuity Equations*, NRL Memorandum Report #93-7192, April, 1993.

[19] E.R. Harrison, *Mon. Not. Roy. Ast. Sc.* , **147**, 279 (1970).

[20] L.D. Landau, & E.M. Lifshitz, *Fluid Mechanics* 2nd ed., Pergamon Press, Oxford (1987).

[21] E.N. Parker, *Cosmological Magnetic Fields*, Clarendon Press, Oxford (1979).

[22] Z. Haiman, & A. Loeb, *Astrophys. J.* **483**, 21 (1997).

[23] A.S. Kompaneets, *Soviet Physics - JETP* **4**, 730 (1959).

[24] Ya.B. Zel'dovich, & R.A. Sunyaev, *Ap. Space Sci.* **4**, 301 (1969).

[25] D.J. Fixsen, E.S. Cheng, J.M. Gales, J.C. Mather, R.A. Shafer, & E.L. Wright, *Astrophys. J.* , **473**, 576 (1996).