

Klein onderzoek: Large scale simulations for LOFAR

Gijs de Goeijen

August 21, 2009

Supervisors: S. Zaroubi & R. M. Thomas

Abstract

An efficient method for simulating the reionization history of the Universe in very large boxes (≥ 500 Mpc/h comoving) while resolving very small mass halos has been developed. The method has two main steps: Firstly, a large number of points, that satisfy a given correlation function, are distributed in space using the Mandelbrot prescription for such a point process. Secondly, a mass is assigned to each point using the Press-Schechter formula as a probability distribution function (PDF). As an outcome the method gives the halo list positions and masses as well as the large scale density and velocity within the box. This method makes it possible to quickly explore a multitude of reionization scenarios for the LOFAR Epoch-of-Reionization key science project data. Here we demonstrate the method by creating a comoving 500 Mpc/h box at redshift 10 with halo mass resolution of $10^8 M_\odot$ resulting in $\sim 6 \times 10^9$ halos needed to get the box to the average density of the Universe and the point process to converge. We also run, as an example, a radiative transfer code to create an ionization box at $z=10$.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 1.1 | Epoch of Reionization | 3 |
| 1.2 | LOFAR | 5 |
| 1.3 | Goal of this project | 6 |
| 2 | The method | 8 |
| 2.1 | Outline | 8 |
| 2.2 | Halo positioning: the Mandelbrot prescription | 11 |
| 2.3 | Press-Schechter formalism | 11 |
| 2.4 | Density field | 13 |
| 2.5 | Velocity field | 13 |
| 3 | Results | 14 |
| 4 | Ionization simulations | 20 |
| 4.1 | Ionization code | 20 |
| 4.2 | Ionization map | 21 |
| 4.3 | Observable quantity: differential brightness temperature | 22 |
| 5 | Conclusion | 24 |
| | Future work | 24 |
| | Acknowledgments | 24 |
| | Bibliography | 25 |

1 Introduction

1.1 Epoch of Reionization

At about 400,000 years after the Big Bang ions and electrons were able to recombine to neutral hydrogen and helium (abundances of heavier elements are negligible at this early time) as the Universe's density and temperature had dropped enough by that time. After matter and radiation decoupled the Universe entered the so called Dark Ages which ended about 400 Myrs later when the first radiation emitting objects have formed. The first objects, e.g., Pop III stars and mini-quasars, started to emit UV photons thus creating ionized bubbles of hydrogen around themselves. After some time enough ionizing sources have formed to almost completely ionize the Universe. The time between which the gas went from entirely neutral to almost entirely ionized is called the Epoch of Reionization (EoR). Figure 1 illustrates the various stages that gas in the Universe went through since recombination.

Studying this epoch can give us insight on many important topics in cosmology, such as galaxy formation and the formation of quasars and metal poor stars (such as Pop III stars). Though a lot of theoretical effort is put into understanding the physical processes that drive the EoR there is little observational evidence most of which is either indirect or dependent upon the model being used. However in the coming years a number of projects plan to probe this very important cosmological epoch, among which are LOFAR (LOW Frequency ARray), MWA (Murchison Wide Array) and SKA (the Square Kilometer Array).

Currently, there are two main observational constraints on the EoR, the CMB and the Lyman-alpha forest, the other observational evidence is either too weak and/or model dependent. The first of the strong constraints comes from the Cosmic Microwave Background (CMB) temperature and polarization which have been measured by the WMAP satellite. This can be used to obtain the Thomson scattering of CMB photons off free electrons which are being produced in the ionization processes during the EoR along the line of sight, known as the Thomson optical depth. From this optical depth it can be inferred that the Universe was mostly neutral until about 400 million years after the Big Bang, after this the ionization started. The second constraint comes from the Lyman-alpha forest measurements found along the spectra of distant quasars. The Lyman-alpha forest appears due to neutral hydrogen, so these measurements can give us some information about reionization. It tells us that the Universe is almost completely ionized at low redshift, only about 1 hydrogen atom in 10,000 is neutral. It also tells us the reionization process finished at about a redshift of 6.5.

These constraints tell us between which times the Universe should have been ionized, but it does not tell us anything about the process itself; how long it took to ionize the Universe, what source(s) ionized the Universe or how it spread through the Universe (will dense clusters or voids be ionized the fastest?). These are questions which we hope to answer with the next generation of radio telescopes mentioned earlier (Jelić et al. 2008, Labropoulos et al. 2009, Zaroubi, 2009 to be published).

One of these next generation radio telescopes is LOFAR which is being build in The Netherlands, which will soon be starting to observe the EoR. The next subsection will deal with LOFAR and its specialized EoR project.

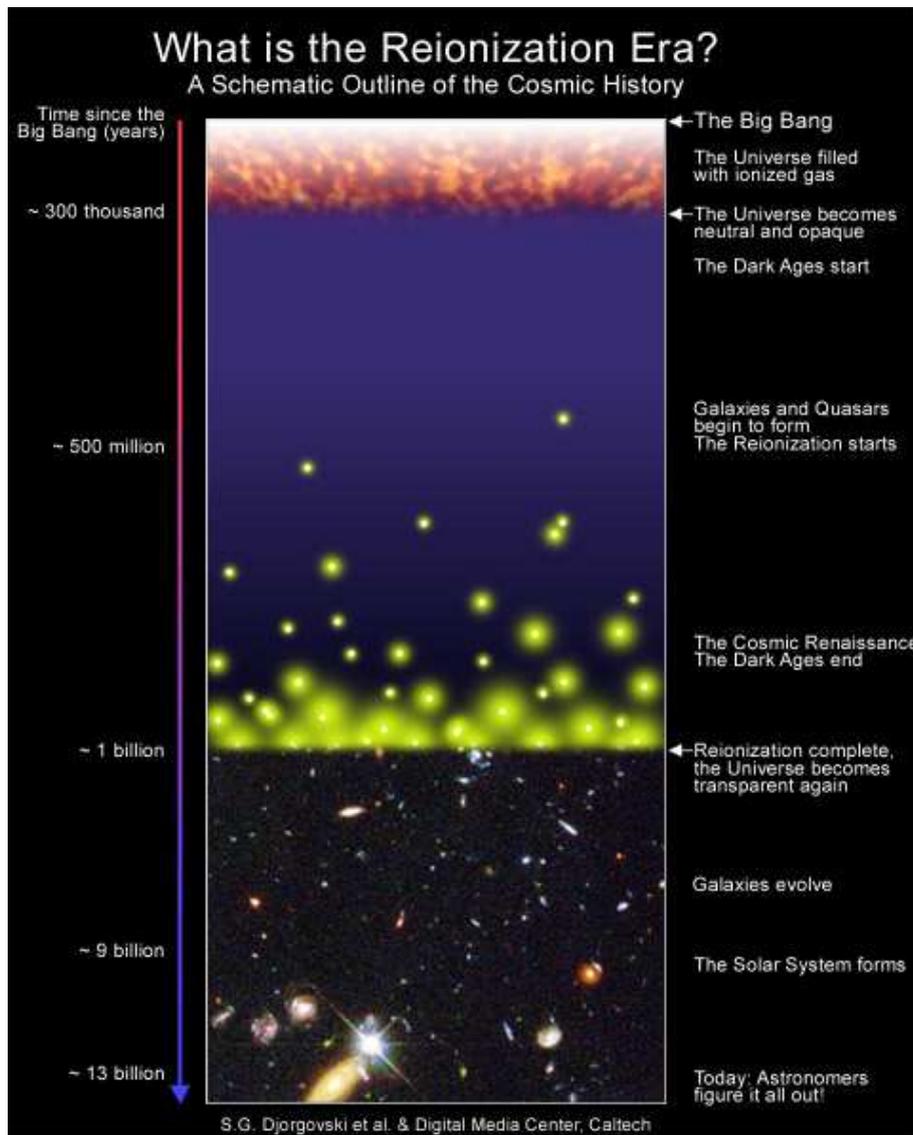


Figure 1: We see the Universe is completely neutral until about 400 million years after the Big Bang when the first objects are being formed and start to create ionized bubbles in the otherwise neutral Universe. The amount and sizes of the sources increases with redshift until about a billion years after the Big Bang when they fill the entire Universe meaning that it is almost completely ionized at this point.

1.2 LOFAR

Early radio telescopes were large dishes that collected the signal that was then processed and analyzed. However to get better and better resolution one would need to build bigger and bigger dishes, which would simply be unaffordable, so another approach is required. LOFAR gets around this problem by using a new technology; it consist of an array of antennas spread over a large area instead of using a single giant dish. The signals received by these antennas are fed into a central processor which then emulates a 'standard' telescope. This central processor is the Blue Gene/P supercomputer located at the 'Rijksuniversiteit Groningen'. The end goal is to have an array of about 19000 antennas spread over an area with a diameter of 350 km with the core area located near Exloo, The Netherlands¹.

LOFAR can be used for many applications in astronomy, such as looking at the Epoch of Reionization (EoR) and the dark ages and looking at massive high redshift galaxies. But because LOFAR can also be turned into a more generic Wide Area Sensor Network it can be used for non-astronomical applications as well, such as geographical and agricultural applications. The project of interest here, is the EoR-project. The primary scientific goals of which are: ²

- To arrive at a statistical estimate of the variation in the 21-cm neutral hydrogen signal as a function of redshift.
- To determine the spatial-frequency power spectrum of the brightness temperature fluctuations on angular scales of 3 arc-minutes to 5 degrees and frequency scales of 0.01 MHz to 10 MHz.
- To image Strömgren bubbles around bright sources and to make targeted observations to detect the 21-cm forest, which is so far undetected.
- Cross-correlation of the 21-cm data set from the EoR with astronomical surveys in other wavebands like the near-infrared and also with the new CMB observations.

This can be done by comparing the signal we receive to ionization simulations. Though this might sound easy and straightforward there are a lot of difficulties involved, for example the removal of foregrounds from the relatively weak EoR signal (Jelić et al. 2008) as well as the large number of free parameters involved in creating a single reionization scenario. This second point is addressed by BEARS, the reionization code written by R. M. Thomas (Thomas et al. 2009) which can explore a large part of the parameter space in an efficient way. But there is a related problem with this, the large field of view of LOFAR ($\sim 5^\circ \times 5^\circ$) requires big simulation boxes of the order of $1 Gpc^3$. Which brings us to the goal of this project.

¹<http://www.lofar.org/p/geninfo.htm>

²Adopted from R. M. Thomas' thesis; Cosmological reionization simulations for LOFAR

1.3 Goal of this project

As stated above it is necessary to simulate big reionization boxes to use for LO-FAR, because of the large field of view, but there are also other reasons why this is required. The first reason has to do with the analysis of the power spectrum. We know the power spectrum peaks at about a scale of 120 Mpc, to analyze the power spectrum on this scale one needs to look at very large volumes; at least 400-500 Mpc to make sure that one is not constrained to just one mode in the simulation box. The second is to be able to compare the simulations with the large-scale CMB data, for example when looking at the polarization of the CMB. And the third has to do with reducing the amount of repetition when one creates a simulated datacube. The way one usually creates such a datacube is by stacking the same box at different times behind each other to create an ionization history, but this causes the structures in that box to repeat themselves. If one increases the box sizes the amount of boxes one has to stack reduces, thus reducing the amount of repetition.

While in theory it is doable to create such a large simulation box with standard N-body simulations it is not very practical, as in order to simulate the reionization process one needs to resolve the bulk of the halo masses ($M_{halo} \approx 10^9 M_{\odot}$) in which ionization sources reside which makes such simulations practically impossible to obtain as it requires a huge amount of active memory as well as storage capacity. The second reason is that N-body simulations take a long time to run, part of this is that they include a lot of different effects as well as that they keep information on all the cells in the grid even if most are empty. This causes a lot of memory to be needed if one would like to simulate a big box (for reference we want to get a box with sides of 500 comoving Mpc/h) with a reasonable degree of accuracy. Especially, if one takes in to account that we do not typically have a single box, but rather a series of boxes at different redshifts to be able to simulate the ionization history of the Universe.

Therefore, it becomes clear that we will need a different method of creating large simulation boxes while resolving the halos responsible for reionization. Here we propose an approximate method which circumvent the memory, storage and running time issues by keeping only the positions and masses of the halos responsible for reionization while roughly estimating the IGM density and velocity fields.

The halo list and IGM density are created by randomly distributing points (halos) in the simulation box that satisfy a given correlation function, the masses of these points are then assigned using the Press-Schechter mass function as a PDF. This process is very simple and greatly reduces the amount of memory and run time needed for these simulations. This method also helps us with the second issue as we just need to have a list with the halo positions and masses instead of keeping information on the entire grid, thus also greatly reducing the storage space required.

Of course this method sweeps a lot 'under the carpet' when compared to N-body simulations, but it is good enough for our purpose as well as a lot faster than using a N-body simulation. The goal of this project is to create the aforementioned method to use with the BEARS ionization code allowing us to obtain different reionization histories on very large scales in a time, storage and memory efficient manner.

The rest of the report is structured in the following way: Section 2 will

outline the steps taken to realize the method followed by an explanation of each step in more detail. In section 3 we will present results obtained by using the method. After that there will be a summary of the BEARS ionization code along with some ionization maps it produced after plugging in the data obtained with this method in section 4 followed by the conclusion and a look at possible future work in section 5.

2 The method

2.1 Outline

As mentioned in the previous section we want to create a list of halos which contains their masses and their positions. First we would like to obtain positions for these halos. Of course we can not just assign random positions to these halos as we know that in the real Universe objects are clustered due to clustering of the primordial matter distribution, hence their positions have to reflect that. Our starting point is choosing a certain correlation function which we assume the halos adhere to; obviously this correlation function will change as a function of redshift. It will also change as a function of halo mass, however, this effect we ignore as it complicates the simulation method; although we note that even then our method could be modified to include this effect. To obtain such positions we use the Mandelbrot prescription (Mandelbrot 1975, 1977). In this prescription the clustering properties of the objects are taken into account.

Secondly, we need a way to assign masses to the halos we generated with the Mandelbrot prescription. To do this we need a PDF that gives the probability for a halo to have a certain mass M at redshift z . This function can then be used to assign masses to the halos in our box. For this we choose to use the Press-Schechter formalism (Press & Schechter, 1974), this formalism predicts the mass distribution of bound objects at a certain redshift z .

In the following subsections a more detailed description of how this works exactly is given. Here we just outline the steps we follow to obtain a very large box simulation: First, we calculate the PDF of the mass distribution at a given redshift as predicted by the Press-Schechter formalism. Next, we limit the number of halos we use for populating our box such that the total mass in our simulation box is equal to the average mass such a volume would have in the real Universe. Furthermore we make sure that the number of halos is enough for the Mandelbrot point process to converge. Throughout the report we adopt the value for the energy content of the Universe found by WMAP, $9.9 \cdot 10^{-30} g cm^{-3}$ and an Ω_M of 0.27.

Furthermore we need to introduce a minimum mass (M_{min}) the halos we are accepting must have. We can not go all the way down to very small masses, because that would require an enormous amount of halos. This would increase the amount of memory and storage space needed dramatically, which goes against the purpose of creating a time efficient method to obtain these boxes. Of course we need to balance this against the fact that one would like a reasonable degree of sensitivity. The minimum mass we decided upon is ($M_{min} = 10^8 M_{\odot}$), for reference this results in about 6×10^9 halos.

We are also interested in creating the cosmological density field, which one must have in order to apply the radiative transfer code, BEARS. The density field is used to determine the ionization profile around a given source. Since we already obtained the masses and positions of the halos getting the density field from this is quite straightforward. The only thing we need to do is grid our box, calculate the density inside each cell and smooth the field afterwards to prevent huge jumps in the density between gridcells. We can take a reasonably large gridscale here as we mostly care for the density on large scales.

The last thing we would like to have is the velocity field, the reason for

this is the redshift distortions that will be in the observations done by LOFAR. The biggest effect of this for our purpose is the Kaiser effect (Kaiser, 1987). This effect is only observable on large scales, it is caused by infalling motions along the line of sight which will cause the structure in redshift space to be ‘squashed’. The effect that this will have on real space is that the contrast between overdense and underdense regions appears larger than it in reality is, i.e. overdense regions appear to be even more overdense while underdense regions appear to be even more underdense. To be able to simulate the LOFAR observations with our method redshift distortions must be included. This can be accomplished by using the velocity field, which can be created, in the linear case, from the density field by using the mass conservation formula.

Once we have obtained these four quantities; masses, positions, density field and velocity field, we can use the BEARS radiative transfer code to create ionization maps out of this for different reionization scenarios, for example using different ionization sources (Pop III stars, QSO’s). BEARS will create ionized bubbles around the most massive halos ($M \geq 10^{10} M_{\odot}$), which we assume to have the strongest ionizing sources. Especially for the case in which QSO’s play a role in ionization you expect them to be in the most massive halos, because smaller halos are unlikely to be able form such an object.

The entire process is depicted in the flow chart given in figure 2. The last block in the figure is the observable EoR, which can be simulated by combining the ionization maps with the redshift distortions which we can obtain using the velocity field. Due to the fact that at the time of writing we have only created one box we can not say much about this yet, but we plan to do this in the future (see also section 5).

To summarize: the steps described above are as follows:

- Obtain positions for the halos using the Mandelbrot prescription
- Assign masses to these halos using the Press-Schechter formalism
- Use the above two to get a large scale density field
- Obtain the velocity field using the density field

Next we will describe the steps outlined in more detail, specifically how the various quantities we need for the simulations are obtained.

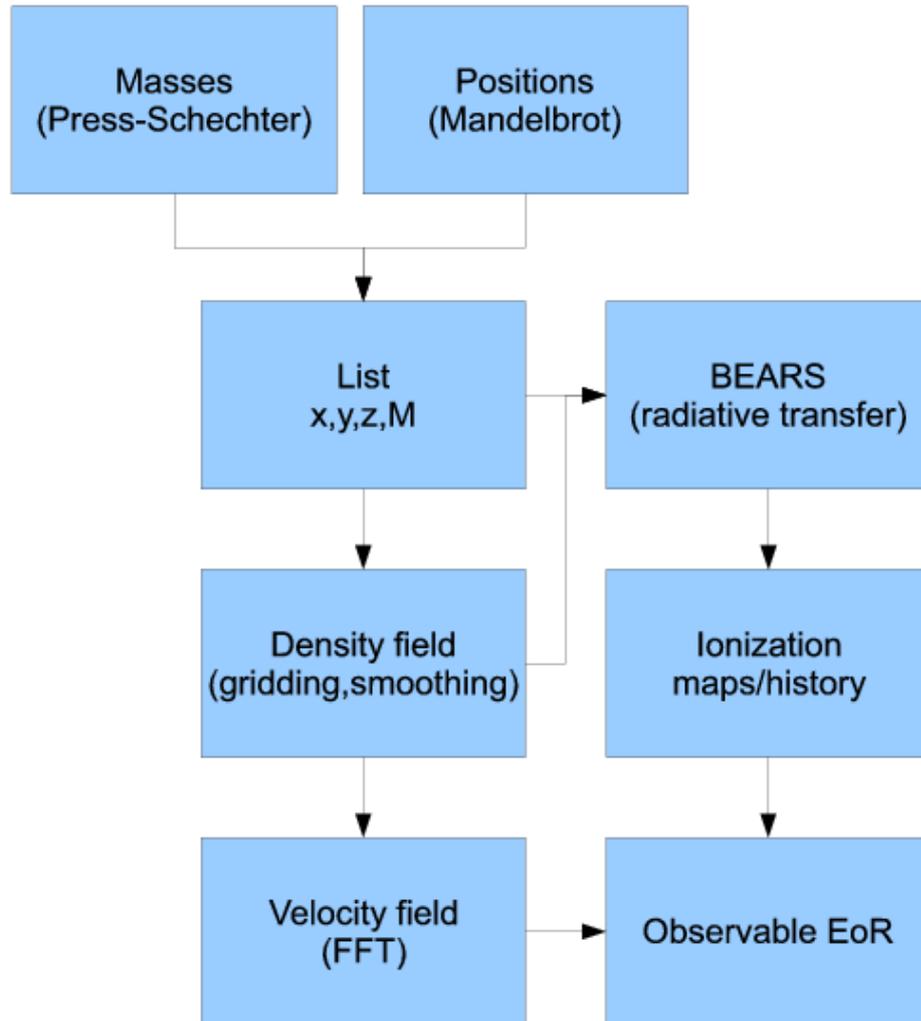


Figure 2: Flowchart showing the various steps of the method and how they relate

2.2 Halo positioning: the Mandelbrot prescription

More than three decades ago Mandelbrot (1975, 1977) proposed an elegant way to determine positions of galaxies that follow a certain correlation function. His prescription uses a Rayleigh-Lévy random walk, at every step of the walk a galaxy is placed. One starts at one of the galaxies and then chooses a random direction and a distance l which is drawn from a probability distribution function which has a higher chance to assign a short distance than a long distance. As previously said, one needs to do this because we know objects are clustered and not just randomly distributed in space. The correlation function ξ is then given by:

$$\xi(r) = \begin{cases} (r/r_0)^{-\alpha} & \text{if } r \geq r_0; \\ 1 & \text{if } r < r_0. \end{cases}$$

Which leads to a cumulative pdf as follows (Peebles 1980):

$$P(> l) = \begin{cases} (l_0/l)^\alpha & \text{if } l \geq l_0; \\ 1 & \text{if } l < l_0. \end{cases}$$

where $\alpha > 0$ always. This process is repeated a large number of times each time choosing a random and independent direction. One needs to do this often to make sure that the distribution of drawn points converges to the desired pdf. Since we have on the order of 6×10^9 halos this is not a big worry, still we verify that the distribution has the correct correlation function (see section 3). For our box we took $l_0 = 2$ Mpc and $\alpha = 1.5$. But we need to include the dependence of this on redshift by multiplying with a factor $D^2(z)$, where $D(z)$ is the linear growth factor of density, which we need to include to take the (linear) growth of density fluctuations into account.

2.3 Press-Schechter formalism

In 1974 the Press-Schechter formalism was developed (Press & Schechter 1974), it predicts the mass distribution of objects formed through hierarchical clustering. Hierarchical clustering is the formation mode of structures in the Universe according to cold dark matter scenarios. Hierarchical clustering tells us that small objects merge to form bigger objects, which form bigger objects again themselves and so on. What this means is that it predicts that there are a lot of small structures (or halos in the context of this project) compared to larger, more massive, structures. Although this method has some limitations it is in remarkable agreement with N-body simulations.

The Press-Schechter formalism is based on assuming a Gaussian random field for the density perturbations, linear gravitational growth and spherical collapse. To determine the amount of halos at redshift z the density field smoothed on a mass scale M is used, δ_M . Using that we know that δ_M is distributed as a Gaussian variable with a mean of zero and a standard deviation $\sigma(M)$ we can write the probability that δ_M is larger than some δ as (White 1995):

$$\int_{\delta}^{\infty} d\delta_M \frac{1}{\sqrt{2\pi}\sigma(M)} \exp\left[-\frac{\delta_M^2}{2\sigma^2(M)}\right] = \frac{1}{2} \operatorname{erfc}\left(\frac{\delta}{\sqrt{2\pi}\sigma(M)}\right) \quad (1)$$

The ansatz is to identify the above probability with the fraction of particles enclosed in collapsed halos of mass M or greater at a redshift z . Then there are two more things needed. The first is that the value used for δ is $\delta_{crit}(z)$, which is the critical density found for a spherical top-hat extrapolated to the present. $\delta_{crit}(z)$ is given by:

$$\delta_{crit}(z) = \frac{1.686}{D(z)} \quad (2)$$

Where $D(z)$ is the linear density growth factor, with $D(z = 0) = 1$. The second is that the fraction of matter in halos above a mass M should be multiplied by a factor 2 to make sure that every particle ends up in a halo with a mass $M > 0$. Using this we can write the formula for the mass fraction in halos above mass M at redshift z as (White 1995):

$$F(> M|z) = \text{erfc}\left(\frac{\delta_{crit}(z)}{\sqrt{2\pi}\sigma(M)}\right) \quad (3)$$

The factor 2 is necessary to make sure we are not only including positive fluctuations of δ_M . However this factor of 2 was also found from the “cloud-in-cloud” problem as described in Bond et al. (1991). If we differentiate the fraction of matter in halos above mass M we obtain the mass distribution. Let dn be the comoving density of halos between masses M and $M + dM$ we find (White 1995):

$$\frac{dn}{dM} = \sqrt{\frac{2}{\pi}} \frac{\rho_m}{M} \frac{-d(\ln\sigma)}{dM} \nu_c e^{-\nu_c^2/2} \quad (4)$$

With $\nu_c = \delta_{crit}(z)/\sigma(M)$. Later on numerical simulations showed that the standard Press-Schechter function over-estimates the number of high mass halos while underestimating the low mass ones as predicted by Peacock and Heavens (1990). To remedy this Sheth & Tormen (1999) developed a new mass function, which we use here.

First let $\nu \equiv [\delta_{crit}(z)/\sigma(M)]^2$, for initially scale-free spectra in models with $\Omega_0 = 1$ and $\Lambda_0 = 0$ we get (Sheth & Tormen 1999):

$$\nu f(\nu) = M^2 \frac{n(M, z)}{\rho_m} \frac{d \log m}{d \log \nu} \quad (5)$$

When fitting to numerical simulations Sheth and Tormen found the following fit, which is a modification to the Press-Schechter function:

$$\nu f(\nu) = A \left(1 + \frac{1}{\nu^p}\right) \left(\frac{\nu'}{2}\right)^{1/2} \frac{e^{-\nu'/2}}{\sqrt{\pi}} \quad (6)$$

With $\nu' = a\nu$, with $a = 0.707$, $p = 0.3$ and $A \approx 0.322$, which follows from the requirement that the integral of $f(\nu)$ over all ν equals 1. For reference the original Press-Schechter formula has values of $a = 1$, $p = 0$ and $A = 0.5$ (Sheth & Tormen 1999).

We used the Sheth and Tormen mass function to create the probability distribution functions from which we have drawn the masses to fill the box. The halo mass which contributes the most mass in total is called M_* . Note that M_* is not the mass which most of the halos have, rather halos with mass M_* contribute the most to the total mass in a given volume.

2.4 Density field

Once a list of halo positions and masses is obtained we can use it to create an underlying density field. The density field is needed for three reasons. The first is that we need it for the radiative transfer code to determine the required ionization profiles for the ionizing sources. The second reason is that we need the density field to be able to obtain the velocity field, which we need in turn to simulate redshift distortions. The last reason is that we need it to calculate the differential brightness temperature from our simulations, which is the quantity that is observable with LOFAR.

To obtain the density field we grid the box and calculate the average density in every gridcell, a 512^3 grid was used for this. After that the field was smoothed using a Gaussian smoothing filter with a smoothing radius of 2 Mpc. This means every point gets smoothed by taking a weighted average of the surrounding points. This is necessary to make sure that there aren't big differences in density on the 'borders' between neighboring gridcells, i.e. to prevent 'jumps' in the density, but instead to make it a smooth transition from one cell to the other.

2.5 Velocity field

The velocity field can be obtained by using the density field, since the two are coupled. Particles move away from underdense regions and towards overdense region because of gravitational attraction. In Fourier-space there is an easy relation between density and velocity field (Coles & Lucchin 2002):

$$\delta(x) = -\frac{\nabla \cdot V}{aHf} \quad (7)$$

To get the velocity field we need to transform our density field to Fourier-space. Once we have that we can use the above relation to obtain the velocity field in Fourier-space. After that we can transform back to real space to get the real velocity field. Note that we will in fact get three velocity fields from this since $V = \{V_x, V_y, V_z\}$. We normally pick one of these as the line of sight direction.

3 Results

This section will be devoted to showing some results obtained by using the method described above. The first result is shown in figure 3. It shows the Sheth & Tormen mass function for a number of redshifts, it was obtained from the code which creates the pdf for the mass function. The pdf shown here for redshift $z = 10$ is the one the masses that fill the box have been drawn from (at the moment of writing only one box has been created for a redshift of 10, but this will surely be extended in the near future). As we can see the chances of picking a halo of mass M decreases fast with halo mass. Also we can see that for increasing redshift the chances of picking a high mass halo drops as well as the maximum mass that can be assigned to a halo. This is what one would expect based on hierarchical clustering, the lower redshift one takes the more time the halos have had to merge to bigger structures. At a higher redshift there has not been enough time yet for the most massive objects we see today to have formed.

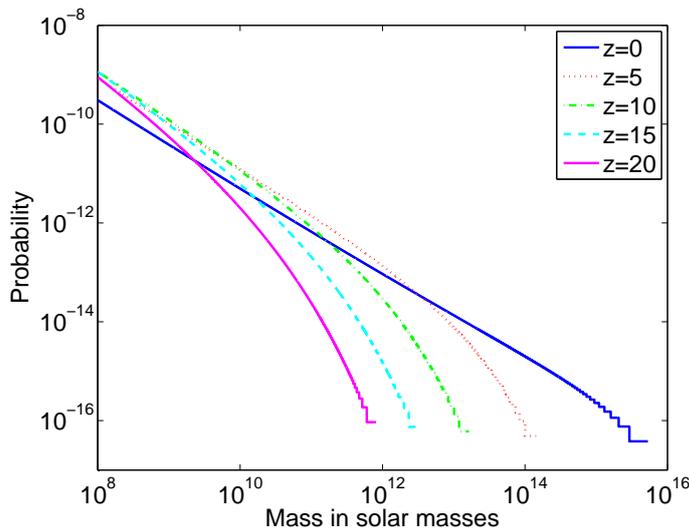


Figure 3: Showing the PDF according to the Sheth & Tormen mass function for a number of redshifts, obtained by using the code that calculates the mass function at different redshifts.

Coupled to this is figure 4, which shows the total mass in halos of mass M versus halo mass M . First of all we can see it has the expected knee-shape, the maximum in this plot is the aforementioned M_* . As expected the cutoff at the high mass end as well as the value for M_* shifts to lower masses as the redshift increases, as we already saw in the previous figure as well. The last thing to note about this figure is that at redshift $z = 10$ the value for M_* is about $10^9 M_\odot$, this is above the minimum mass we used for our halos ($10^8 M_\odot$). Which is comparable to the halo resolution obtained in small scale N-body simulations.

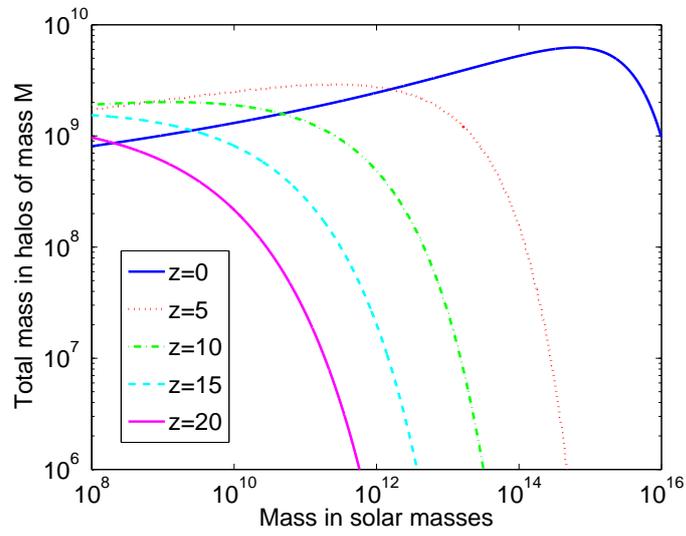


Figure 4: Showing the distribution of masses according to the Sheth & Tormen mass function, obtained by using the code that calculates the mass function at different redshifts.

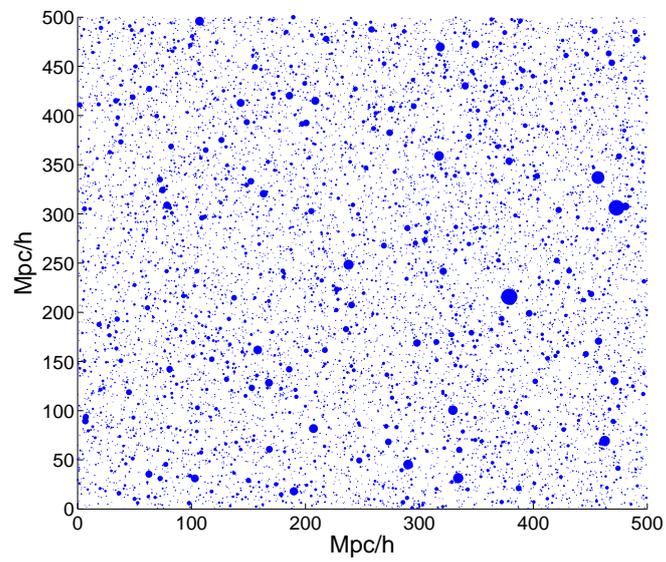


Figure 5: Showing a slice from the box, the dot size is (linearly) scaled with halo mass

Figure 5 shows the positions of halos coupled to masses drawn from the Sheth & Tormen mass function. The figure only shows the heaviest halos ($M > 10^{10} M_{\odot}$), which are the ones used for the ionization code. The points in the figure represent the halos, the point size is scaled with the mass of that halo in a linear way. Again we see that the number of halos with mass M drops dramatically with increasing mass.

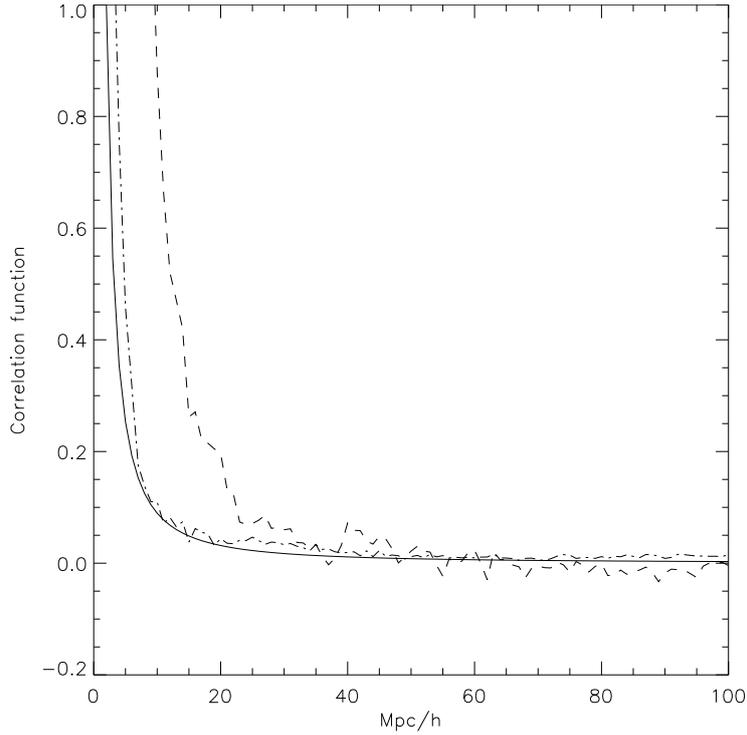


Figure 6: The dashed line is the correlation function estimated using the Landy & Szalay estimator for a sample of 100,000 data points and 5,000,000 random points, the dot-dashed line is the correlation function estimated using the Landy & Szalay estimator but for samples 10 times as big. The solid line is the input correlation function.

We also checked whether the points are actually correlated to see if the code produces the correct distribution for the halo positions. For this we used the estimator for the correlation function suggested by Landy & Szalay (1993), which can be written as:

$$\xi_{LS} = 1 + \left(\frac{N_{rand}}{N_{data}}\right)^2 \frac{DD(r)}{RR(r)} - 2 \frac{N_{rand}}{N_{data}} \frac{DR(r)}{RR(r)} \quad (8)$$

Where $DD(r)$ are points from the data compared to other points from the data, $DR(r)$ are points from the data compared with points from a random sample, $RR(r)$ are points from a random sample compared with other points from this sample, N_{data} is the number of data points and N_{rand} is the number of points in the random sample. However, this formula assumes one compares

every pair of points in the two samples, but this is not possible due to run time issues for the large samples we need to make the distribution converge to the desired PDF. In our case we only took a small number of points from the entire sample and compared all other points to those, because of this a factor of $(\frac{N_{rand}}{N_{data}})$ drops out and we need a slightly different formula to estimate the correlation function:

$$\xi_{LS} = 1 + (\frac{N_{rand}}{N_{data}}) \frac{DD(r)}{RR(r)} - 2 \frac{DR(r)}{RR(r)} \quad (9)$$

The result of this is plotted in figure 6. In this figure the correlation function estimator has been plotted for two samples of different sizes as well as the input correlation function as given in section 2.2. Due to time constraints we were unable to make a plot of this for the latest box we ran while having a reasonable samplesize, but this plot still serves to show that we obtain the desired correlation function. We can see that for the smaller of the two samples the estimator of the correlation function has not converged to the correlation function we put in. The larger sample is already much closer to the correlation function that was put in, it has not entirely converged to it yet but it is much better compared to the smaller sample. It should be noted that the larger sample still only contained 10^6 points while the simulation box has about 6×10^9 points. From this it is safe to assume that for such a large sample we obtain the correlation function we initially put in.

Finally we can look at the density and velocity fields, in figure 7 a slice of the density field has been plotted along with the corresponding slice from the velocity field. The density field shows over- and underdensities, i.e. $\rho/\rho_{mean} - 1$. The black contours are at an overdensity of 0, the red contours show overdensities and the blue ones underdensities, the spacing between the contour levels is 0.05. As we can see they match very well, at the positions of overdensities we see velocity vectors pointing towards it and at underdense sites we see velocity vectors pointing away as expected.

Figure 7 shows the density field without redshift distortions, for comparison we included a plot of a density field with and a density field without redshift distortions (figure 8). The colors and level spacing in this plot are the same as for figure 7. We see the expected differences, overdense regions appear to be even more overdense than they are in reality while underdense regions appear to be even more underdense. This results in a much larger observed contrast between overdense regions, such as clusters and underdense regions, such as voids, which in reality is a lot shallower. We can see the effect is quite large confirming again the need to include this effect in our simulations.

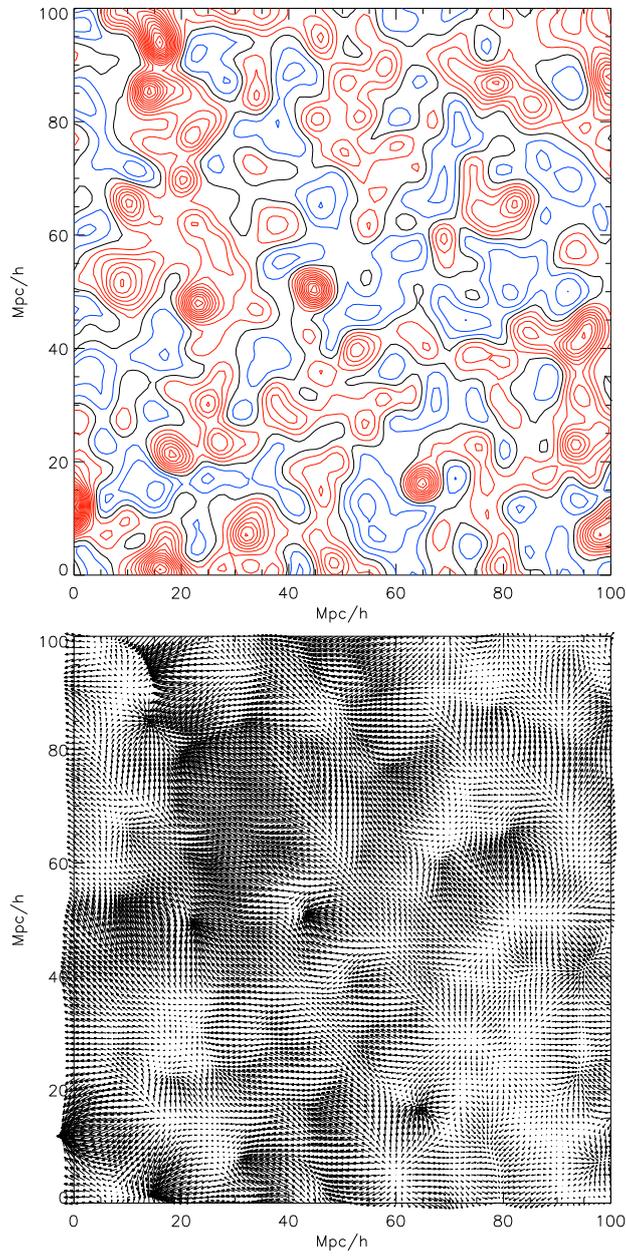


Figure 7: The top panel shows the density field, the bottom panel shows the corresponding velocity field

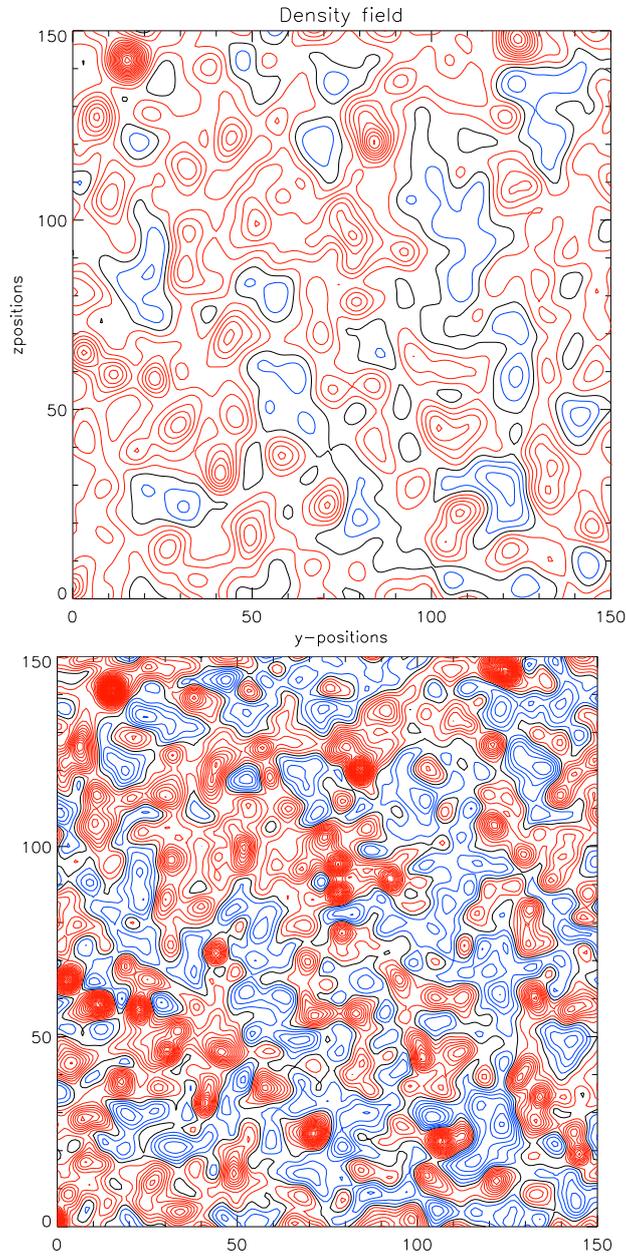


Figure 8: The top panel shows a slice of the density field before redshift distortions are taken into account, the bottom panel shows the same slice including the redshift distortions. Due to time constraints this plot is from an older box, but it still serves to show the large effect redshift distortions have on the density field.

4 Ionization simulations

When the LOFAR EoR-project is going to start collecting data it will be necessary to have algorithms that can extract the signal from this. To test whether these algorithms are stable and reliable one has to simulate a large range of reionization scenarios (Thomas et al. 2009). Due to the large number of simulations that need to be run it is inefficient to use a full 3-D radiative transfer code as these take a long time to run. To solve this a new 1-D radiative transfer code, BEARS (Bubble Expansion Around Radiative Sources) (Thomas et al. 2009), was created. Here a short summary will be presented of how this works, for the full treatment see Thomas et al. (2009).

4.1 Ionization code

Instead of using a full 3-D radiative transfer code BEARS assumes the ionizing bubbles around radiation sources to be spherically symmetric, the ionized fraction and temperature profiles for these bubbles are derived from a pre-made catalogue using the 1-D radiative transfer code developed by Thomas & Zaroubi (2008). This is done by solving a set of rate equations at every gridcell as a function of time. The time-evolution of the following species is followed: H_I , H_{II} , He_I , He_{II} and He_{III} as well as the temperature. The code used for this is modular, making it a straightforward process to put in different spectra corresponding to different ionization sources (stars, qso's).

Using the 1-D radiative transfer code a number of ionization profiles are computed which are stored in a table to ensure faster execution. Once one knows the positions and masses of the halos, their velocities, the underlying density field and the ionizing luminosity if the simulation also contains gas the following steps are taken to create ionization bubbles around the sources:

- Given the redshift, the ionizing luminosity (if there is gas) and the type of source an ionization profile is chosen from the previously generated catalogue of profiles.
- The source is usually in an overdense region with a certain profile, which varies from source to source. Therefore the following approximation is used: the overdensity within a certain radius R_{od} of the source is computed. It is then assumed this source is in a uniform density field with a density equal to the average overdensity within R_{od} . Which translates to taking the same ionization profile but at a higher redshift.
- Then a correction is made for overlapping ionized regions (becomes especially important at lower redshifts). The photons which ionize the overlapping region are redistributed in other regions which are still neutral, thus conserving the amount of ionizing photons.
- When computing the ionization history, ionized regions are mapped one-to-one from the current simulation snapshot to the next.

The averaging radius R_{od} is calibrated by taking one point and calculating the ionization profile using radiative transfer. Then a very small R_{od} around one calibration point is taken which is increased until the extent of the ionization

profile for this R_{od} is equal to the radiative transfer calculation done before. After this all the other points can be scaled to this calibrated point.

Finally the correction for overlapping ionization regions, which is especially important for lower redshifts where the ionizing sources become more numerous as well as heavier. A region of overlap is a region which is being ionized by more than one source, which results in unused photons in this region. What is done to compensate for this is that all the bubbles which are overlapping in this particular region are increased in size until the increased volume equals that of the overlap. This makes sure that the amount of ionizing photons is conserved as well as ensuring that all the previously unused photons are homogeneously distributed among the entire region.

These simulations have proven to be a lot more time efficient than a full 3-D radiative transfer code, while still maintaining a good degree of accuracy as shown in Thomas et al. (2009).

4.2 Ionization map

Using the ionization code on the created box produces a box with ionized bubbles around the heaviest halos ($M \geq 10^{10} M_{\odot}$). This box gives a snapshot of the reionization history of the Universe at a redshift $z = 10$ Figure 9 gives an example of a slice taken from this box. For this box the reionization scenario assumes that the Universe was ionized by QSO's, but due to the way the code is written it is easy to probe other scenarios as well. Because of the fact that we only have one snapshot of the reionization history we can not say much about it yet, we know that at high redshift all the hydrogen is neutral and that at low redshift nearly everything is ionized with just one image in between. So to get a better understanding on how this process has developed through

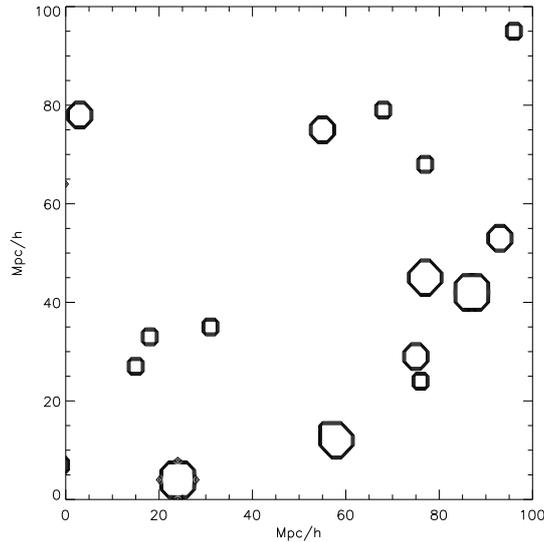


Figure 9: Shows a slice from the box with ionized bubbles around the heaviest halos

time we need to have more snapshots, which will be addressed in future work (see next section). What we can see in this picture is that at redshift $z = 10$ for this particular scenario we expect there to be a number of ionization sources already which are creating ionized bubbles around them. What we can also see is that at this redshift the bubbles are far from volume filling as one would expect them to be.

4.3 Observable quantity: differential brightness temperature

In radio astronomy, where the Rayleigh-Jeans law is usually valid, the radiation intensity, $I(\nu)$ is expressed in terms of the brightness temperature, as follows (Zaroubi & Silk 2005):

$$I(\nu) = \frac{2\nu^2}{c^2} kT_b, \quad (10)$$

where ν is the radiation frequency, c is the speed of light and k is Boltzmann's constant (Rybicki & Lightman 1979). This can only be observed differentially as a deviation from the cosmic microwave background T_{CMB} . The predicted differential brightness temperature deviation from the cosmic microwave background radiation is given by (Field 1958; 1959; Ciardi & Madau 2003),

$$\delta T_b = 20 \text{ mK} (1 + \delta) \left(\frac{x_{HI}}{h} \right) \left(1 - \frac{T_{CMB}}{T_s} \right) \left(\frac{\Omega_b h^2}{0.0223} \right) \left[\left(\frac{1+z}{10} \right) \left(\frac{0.24}{\Omega_m} \right) \right]^{1/2}. \quad (11)$$

Here T_s is the spin temperature and Ω_m and Ω_b are the mass and baryon density in units of critical density and δ is the overdensity ($\rho/\rho_{mean} - 1$) (Thomas, 2009, PhD thesis). In this report we take T_s to be much larger than T_{CMB} . The differential brightness temperature is the physical quantity that LOFAR will observe.

In the figure below (figure 10) we plotted the differential brightness temperature as defined above, there x_{HI} is the neutral fraction of hydrogen, which is zero inside the bubbles and one everywhere else. In the figure, brighter colors mean higher temperatures and the black circles are the ionization bubbles around their sources. The differential brightness temperature in these bubbles is zero because there $x_{HI} = 0$.

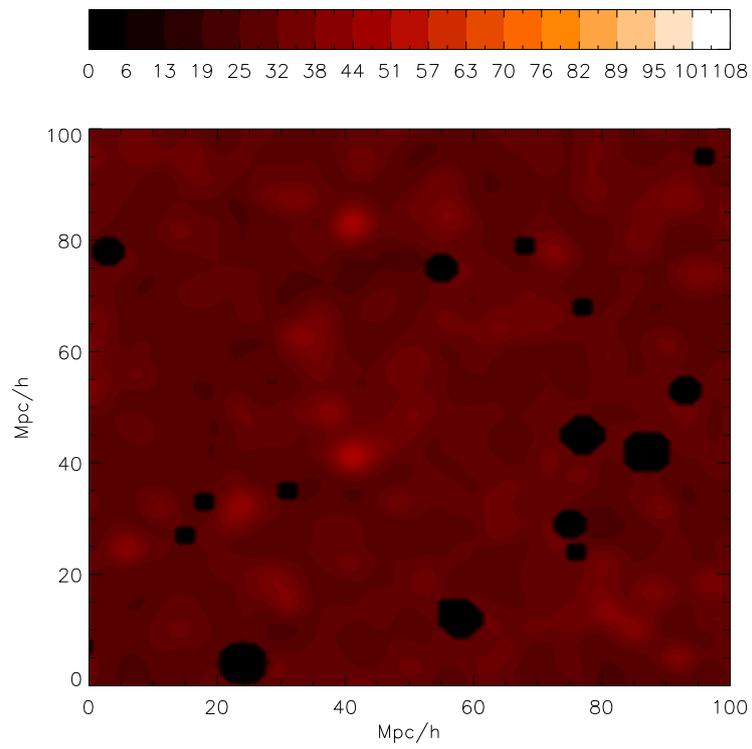


Figure 10: Shows the differential brightness temperature in the same slice as depicted in figure 9

5 Conclusion

In this project we set out to develop a method capable of creating a large box (500 comoving Mpc/h sides) in a time and memory efficient manner. We used the Press-Schechter formalism and Mandelbrot prescription to get distributions for halo masses and positions respectively, from this we obtained a list of halos which compared to N-body simulations is a very storage space efficient way of simulating. Then we obtained density and velocity fields. We knew that redshift distortions would be an important effect, so we took those in to account by using the velocity field. This method is able to fill a box of 500 comoving Mpc/h a side with halos starting from a minimum mass of $10^8 M_\odot$, resulting in $\sim 6 \cdot 10^9$ halos, in a few days time, using $\sim 300 - 350$ GB of storage space. While this may seem long it should be noted that to reach this kind of sensitivity with a traditional N-body simulation would have take a lot longer as well as take up a lot more storage space. Also of note is that this was not done on a state of the art supercomputer as opposed to N-body simulations as well as the likely possibility of further optimizing the method.

Future work

At the moment we created a box with sides of 500 comoving Mpc/h with a minimum halo mass of $10^8 M_\odot$, in the future it might be desirable to go to lower masses to get a better approximation of reality or to make even larger boxes. However since the PDF is non-linear going down to a mass of $10^7 M_\odot$ would require a lot more halos. We choose not to go down to that mass for the moment, but it is most likely possible once the code is further optimized

The second point is that so far we only have one box, so the only thing we can do at the moment is get a 'snapshot' of the ionization at redshift $z = 10$. What we would like to have is a series of boxes at different redshifts to create entire ionization histories because we could get a lot more information out of these especially when comparing different scenarios with one another. This will be done in the very near future, since we have a working method to create a box at a given redshift it is rather straightforward to create a series of boxes over a range of redshifts.

The last thing to note is that should one desire this method can also be used to look at what happens on the smaller scales with the ionization sources, since both the Press-Schechter formalism and the Mandelbrot prescription are also valid on small scales.

Acknowledgments

I would like to thank S. Zaroubi for being a great supervisor who has answered a lot of my questions and has been giving helpful insights over the course of the previous months. I would also sincerely like to thank R. M. Thomas for the many times he helped me with out when I was stuck writing the code for the method as well as for providing useful ideas and comments.

Bibliography

- Bond, J. R., Cole, S., Efstathiou, G. and Kaiser, N., Excursion set mass functions for hierarchical Gaussian fluctuations, *ApJ*, 379, 440, 1991
- Ciardi, B. and Madau P, Probing beyond the Epoch of Hydrogen Reionization with 21 Centimeter Radiation, *ApJ*, 596, 1-8, 2003
- Coles, P. and Lucchin, F., *Cosmology: The Origin and Evolution of Cosmic Structure*, John Wiley & Sons, Ltd, 2002
- Field, G. B., Excitation of the Hydrogen 21-cm Line, *Proc IRE*, 46,240,1958
- Field, G. B., The time relaxation of a resonance-line profile, 129, 551, 1959
- Jelić, V., Zaroubi, S., Labropoulos, P., Thomas, R. M., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Ciardi, B., Harker, G., Koopmans, L. V. E., Pandey, V. N., Schaye, J. and Yatawatta, S., Foreground simulations for the LOFAR-epoch of reionization experiment, *MNRAS*, 389, 1319-1335, 2008
- Kaiser, N., Clustering in real space and in redshift space, *MNRAS*, 227, 1-21, 1987
- Labropoulos, P., Koopmans, L. V. E., Jelić, V., Yatawatta, S., Thomas, R. M., Bernardi, G., Brentjens, M., de Bruyn, G., Ciardi, B., Harker, G., Offringa, A., Pandey, V. N., Schaye, J., Zaroubi, S., The LOFAR EoR Data Model: (I) Effects of Noise and Instrumental Corruptions on the 21-cm Reionization Signal-Extraction Strategy, *MNRAS*, 000, 1-20, 2009
- Landy, S. D. and Szalay, A. S., Bias and variance of angular correlation functions, *ApJ*, 412, 64-71, 1993
- Loeb, A., First light, eprint arXiv:astro-ph/0603360, 2006
- Mandelbrot, B., On a decomposable model of a hierarchical universe - Derivation of galactic correlations on the celestial sphere, *CRASM*, 280, 1551-1554, 1975
- Mandelbrot, B., *The fractal geometry of nature*, W. H. Freeman, 1977
- Peacock, J. A. and Heavens, A. F., Alternatives to the Press-Schechter cosmological mass function, *MNRAS*, 243, 133-143, 1990
- Peebles, P. J. E., *The large-scale structure of the Universe*, Princeton University Press, 1980
- Press, W. H. and Schechter, P., Formation of galaxies and clusters of galaxies by self-similar gravitational condensation, *ApJ*, 187, 425-438, 1974
- Rybicki, G. B. and Lightman, A. P., *Radiative processes in astrophysics*, Wiley Interscience, 1979
- Sheth, R. K. and Tormen, G., Large-scale bias and the peak background split, *MNRAS*, 308, 119-126, 1999
- Thomas, R. M. and Zaroubi, S., Time-evolution of ionization and heating around first stars and miniquasars, *MNRAS*, 384, 1080-1096, 2008
- Thomas, R. M., Zaroubi, S., Ciardi, B., Pawlik, A. H., Labropoulos, P., Jeli, V., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Harker, G. J. A., Koopmans, L. V. E., Mellema, G., Pandey, V. N., Schaye, J. and Yatawatta, S., Fast large-scale reionization simulations, *MNRAS*, 393, 32-48, 2009
- Thomas, R. M., *Cosmological Reionization Simulations for LOFAR*, PhD thesis, 2009 ISBN electronic version: 978-90-367-3758-6
- Zaroubi, S., Silk J., LOFAR as a probe of the sources of cosmological reionization, *MNRAS*, eprint arXiv:astro-ph/0412080, 2005
- Zaroubi, S., Probing the End of the Universe's Dark Ages with LOFAR, *Europhysics News*, 2009 (to be published)