Mapping baryonic physics to a dark matter field

Autor:
Manuel Sánchez Benavente

Tutor:
Francisco Shu Kitaura Joyanes

Trabajo de fin de Máster

2020

Universidad de La Laguna
Facultad de ciencias
Máster en astrofísica

A mis padres, María Isabel y Manuel,
por aguantarme todos estos años
y apoyarme sin reservas. A mis amigos,
que me han animado todos estos meses.
A Tino e Iván, porque es justo y necesario.

Contents

1. Introduction **1**

2. Hydrodynamical simulation **2**
2.1. Collisionless dynamics 3
2.2. Hydrodynamics 4
2.3. Additional physics 5
2.4. Outputs of the simulation 7

3. Methodology: Bias Assignment Method (BAM) **8**
3.1. Calibration 8
3.2. Generation of mock density fields 11
3.3. Extension of the `BAM` method to a continuous distribution 11

4. Results and analysis **12**
4.1. Gas density field 14
4.1.1. Low resolution case: mesh with 128^3 cells 14
4.1.2. High resolution case: mesh with 512^3 cells 19
4.2. Neutral hydrogen: HI number density field 21
4.2.1. Low resolution case: mesh with 128^3 cells 21
4.3. Temperature field 26
4.3.1. Low resolution case: mesh with 128^3 cells 26

5. Summary and conclusions **29**

6. Acknowledgements **31**

Resumen

La distribución de materia bariónica en el Universo contiene una vasta riqueza de información tanto cosmológica como astrofísica. Viene dada como resultado de la formación de estructuras, dominada por la materia oscura en un Universo en expansión, dentro del contexto de la interacción con una serie de complejos procesos físicos. Las simulaciones hidrodinámicas que pueden llevarse a cabo en la actualidad incluyen la mayoría de los procesos físicos involucrados en los fenómenos que incumben a la materia bariónica, produciendo una descripción bastante fehaciente de ésta en el Universo. Sin embargo, estas simulaciones son extremadamente costosas en términos de tiempo y memoria de cómputo, y, por lo tanto, ni los volúmenes cosmológicos equivalentes a los que se obtendrán de los diferentes sondeos astronómicos observacionales venideros, ni el número necesario de este tipo de simulaciones, están disponibles para poder desarrollar análisis estadísticos en profundidad. Este hecho hace que sea de vital importancia el desarrollo de diferentes técnicas que permitan extraer la información estadística y cosmológica de esos costosos cálculos, aprenderla y poder mapearla sobre simulaciones que únicamente incluyan partículas de materia oscura, mucho menos costosas en todos los términos. Este trabajo de fin de máster explora esta última posibilidad, enfocándose en una serie de cantidades derivadas de diferentes procesos involucrados en la física bariónica, tales como la densidad de gas, la densidad de hidrógeno neutro, o la temperatura, presentes tanto en galaxias como en medio intergaláctico. Para ello, se dispone de una simulación hidrodinámica de referencia obtenida a partir del código GADGET3-OSAKA, de la que se comentan las particularidades en el tratamiento de la formación estelar, de la creación de elementos metálicos y de polvo a través de la retroalimentación en explosiones de supernova (tanto de tipo II como Ia) y en estrellas de la rama asintótica de las gigantes, así como del tratamiento que se le da a la evolución de la abundancia de polvo con el tiempo, además de las generalidades básicas presentes en toda la simulación hidrodinámica. Se extrae aquí la relación de *bias* presente entre esas cantidades y el campo de materia oscura subyacente a todas ellas como función de la densidad local de materia oscura. Asimismo, se extiende este estudio para tener en cuenta algunas relaciones no locales a esta densidad presentes en la relación de *bias* a través del análisis del tensor de mareas. Para llevar a cabo esta tarea, este trabajo se cimentará sobre el recientemente desarrollado *Bias Assignment Method*, describiendo el método seguido para poder reproducir la estadística de dos puntos con precisión a partir de la extracción de la relación de *bias* directamente de la simulación de referencia y de la calibración de un *kernel* en un proceso iterativo, y poder generar catálogos que sigan esta estadística únicamente a partir de las condiciones iniciales de la simulación de referencia, y lo extenderá para lidiar con distribuciones continuas más allá de la distribución discreta de un número de halos para el que fue originalmente concebido este método. Además, se implementarán una serie de configuraciones para los intervalos de clase de los histogramas empleados para obtener la relación de *bias* distintas a la original en un intento de representar de forma óptima cada una de las cantidades mencionadas anteriormente con un número finito de estos intervalos. Esta serie de experimentos numéricos demuestran que es posible mapear la densidad de gas, con un error relativo por debajo del uno por ciento en el espectro de potencias en diferentes resoluciones. En particular, se obtiene una distribución tridimensional del campo de densidad de gas completamente indistinguible de aquél provisto directamente por la simulación hidrodinámica de referencia a escalas de Mpc. A escalas por debajo de los Mpc comienzan a encontrarse discrepancias entre los campos generados y los de la referencia, que deberán ser investigadas en el futuro mediante relaciones de *bias* más complejas. Asimismo, se investiga la distribución de hidrógeno neutro y se obtienen espectros de potencia que no se desvían más allá de un dos por ciento de aquel de la referencia. Sin embargo, al acudir a las distribuciones tri-

dimensionales de hidrógeno neutro producidas experimentalmente, las diferencias entre éste y el campo de la referencia empiezan a ser visibles en escalas de Mpc. Como paso final, se estudia el mapeado del campo de temperaturas. Este último tiene una dificultad añadida a los anteriores, pues la temperaturas se encuentran mucho más difundidas que el campo de densidad de materia oscura. El uso de una configuración específica para los intervalos de las clases del histograma, como se ha comentado con anterioridad, y la transformación a una escala logarítmica que se hace sobre el campo de materia oscura en todos los casos, parecen atenuar este problema. De hecho, se obtiene convergencia para el caso de la temperatura, aplicando esta configuración, para escalas mayores a $k = 2h\text{Mpc}^{-1}$ (en espacio de Fourier). Asimismo, la distribución tridimensional producida experimentalmente se asemeja a la de la referencia hasta escalas de Mpc. De todo esto, se concluye que las técnicas de mapeado son muy prometedoras como herramienta para el futuro y podrían ayudar a reproducir este tipo de simulaciones hidrodinámicas, si bien es necesario implementar una serie de mejoras, especialmente en lo que respecta a la relaciones de *bias* no locales. Estos resultados podrían ser de especial interés para los próximos sondeos de corrimiento al rojo de galaxias que mapearán la distribución del bosque de Lyman alpha, como DESI y WEAVES-JPAS.

Abstract

The distribution of baryonic matter in the Universe has a wealth of cosmological and astrophysical information. It results from structure formation dominated by dark matter in an expanding background Universe within the interplay of a series of complex physical processes. State-of-the-art hydrodynamical simulations include the majority of the involved physics to produce a fair description of the baryonic matter in the Universe. However, they are extremely expensive to run, and thereby neither the cosmic volumes, which will be provided by forthcoming astronomical observational campaigns, nor the required number of simulations to perform robust statistical analysis are available. This calls for the development of techniques which permit us to learn extract the information from such expensive calculations and map them onto a far less expensive dark matter only simulation. The present master thesis explores this possibility by focusing on a series of baryonic quantities such as the gas density, the neutral hydrogen, and the temperature. We extract the bias relation between those quantities and the underlying dark matter field as a function of the local density. We further extend the study to account for non-local bias through the tidal field tensor. To this end we rely on the recently developed Bias Assignment Method and extend it to cope with continuous quantities beyond halo number counts, for which it was conceived. Furthermore, we implement different binning schemes in the aim of optimally representing each quantity with a finite number of bins. Our numerical experiments demonstrate that we can map the gas to percentage accuracy in the power spectrum to different resolutions. In particular, we obtain a three dimensional gas density distribution which is indistinguishable with respect to the one from the hydrodynamical simulation at Mpc scales. On sub-Mpc scales we start to see deviations, which should be further investigated with more complex bias relations. We also investigate the neutral hydrogen density and obtain power spectra in agreement with the reference one within a few percent. However, in this case the differences in the distribution are visible already at Mpc scales. Finally, we investigate the mapping of the temperature. This quantity shows some additional difficulty, because the temperature of the baryonic gas is further extended than the cold dark matter density. Our specific binning of the temperature in the bias extraction, and the logarithmic transformation of the dark matter density field mitigate this problem. In fact we obtain convergence of the temperature beyond $k = 2h\mathrm{Mpc}^{-1}$. Also the three dimensional distribution resembles the reference one. We conclude that mapping techniques are very promising and could help to reproduce hydrodynamical simulations, but further improvements need to be done, especially in the non-local bias relation. These findings could be of special interest for upcoming galaxy redshift surveys which will map the Lyman alpha forest, such as DESI and WEAVES-JPAS.

1. Introduction

Structure formation in the Universe is a complex process resulting from the interplay between baryonic matter and dark matter through gravity and a series of physical phenomena, such as the forces that govern fluid dynamics, star formation, active galactic nuclei (AGN) or supernovae (SN) feedback. The great amount of data provided by large galaxy surveys such as eBOSS (Dawson et al., 2016) or DES (The Dark Energy Survey Collaboration, 2005), or the huge cosmological volumes that are expected to be explored with experiments such as Euclid (Amendola et al., 2018) and DESI (Levi et al., 2013), represent a great challenge in terms of analysis and will lead to set constraints to the model of Universe that is currently accepted, requiring a quickly evolution of Cosmology into a much more precise and accurate science. For instance, those last surveys have demanded the measure of covariance matrices for appraising the uncertanties in the measurements of cosmological observables such as redshift space distorsions (e.g. Kaiser, 1987) and baryonic acoustic oscillations (e.g. Eisenstein & Hu, 1998), and, ultimately, galaxy clustering analysis (e.g. Dodelson & Schneider, 2013; Taylor et al., 2013; Percival et al., 2014; Paz & Sánchez, 2015; Pearson & Samushia, 2016; Howlett & Percival, 2017; O'Connell & Eisenstein, 2019).

The production of large sets of mock catalogues based on detailed N-body simulations has been the path chosen as the way to obtain that covariance matrices and estimates on the errors of cosmological observables. However, the generation of such catalogues with this approach could be described as unpractical, as N-body simulations require a large amount of computation time and memory to get to the desired cosmological volumes and resolutions. Producing mock catalogues with a different approach of that resource-consuming based on N-body simulations has also been considered: some of them rely on dark matter fields obtained from approximate gravity solvers and populate it with dark matter halos or galaxies following a pre-defined bias relation (*bias mapping technique*). The precision of this type of approach in the resulting halo catalogs is only acceptable, according to scientific requirements of forthcoming surveys, until intermediate scales ($k \sim 0.2 - 0.3\,h\,\mathrm{Mpc}^{-1}$ in Fourier space), (see e.g. Kitaura et al., 2016; Blot et al., 2019; Colavincenzo et al., 2019; Lippich et al., 2019).

New observational campaigns will deliver an unprecedented map of the inter-galactic medium, which is even more challenging, since the above mentioned baryonic physics will have to be included in the cosmological anaysis. However, running hydrdynamical simulations which include all the necessary physics is computationally extremely expensive. In addition to gravity, cooling of the gas, star formation, AGN, SN, stellar winds feedback and chemical reactions have to be modelled. These efforts are necessary, and need to be compared to observations to unveil and calibrate the relevant physical processes yielding our observable Universe. However the available computational resources do not allow to make these efforts covering the volumes which will be available by the next generation of surveys such as DESI, EUCLID, WEAVES-JPAS (Benitez et al., 2014). These surveys will also provide a better picture of the Lyman alpha forest distribution on high redshift, which will be huge importance on assessing the structure of the Universe at scales that have not yet been investigated (Bautista et al., 2017).

Having this in mind, we explore in this work the extension of the recently developed Bias Assignment Method (BAM) (Balaguera-Antolínez et al., 2019, 2020; Kitaura et al., 2020), which has shown to reach higher precision ($\sim 1-2\%$) towards smaller scales beyond $k = 1\,h\,\mathrm{Mpc}^{-1}$ (including the Nyquist frequency) than any other method mentioned above, to map baryonic properties onto a dark matter mesh. The production of mock catalogues including as much baryonic physics as possible with that accuracy and without the need of consuming a huge amount of resources is, without any doubt, a very powerful tool that must be investigated. In particular, the first step will be to analyze detailed hydrodynamical simulations and their products. Subsequently the bias relations between those properties and the underlying dark matter field will be extracted. Then the required kernel will be sampled within a Markov Chain Monte Carlo approach the required kernel, which will enable to reproduce the property's three dimensional distribution. The results from a series of numerical tests at different resolutions and the conclusions are presented below.

2. Hydrodynamical simulation

In this section the reference hydrodynamical simulation is described, which is aimed to be mapped from the dark matter field.

In particular, the cosmological smoothed particle hydrodynamics (SPH) code, GADGET3-OSAKA (Aoyama et al., 2018; Shimizu et al., 2019), is the one on which this work relies, being a modified version of GADGET-3 (originally described as GADGET-2 by Springel, 2005). This code includes the density-independent formulation of SPH (Hopkins et al., 2013; Saitoh & Makino, 2013) and the time-step limiter (Saitoh & Makino, 2009). The quintic spline kernel (Morris, 1996) is adopted, and the number of neighbour particles for each SPH particle is set to 128 ± 8. The radiative cooling is calculated using the Grackle-3 chemistry and cooling library [1] (Smith et al., 2017), which solves the primordial chemistry network for atomic H, D, He, as well as molecular H_2 and HD. The library also includes photo-heating and photo-ionization under the UV background (UVB), for which Haardt & Madau (2012) model is adopted.

In order to avoid artificial numerical fragmentation when the Jeans mass at low temperatures is not resolved, a Jeans pressure floor is introduced, following Hopkins et al. (2011); Kim et al. (2016):

$$P_{\mathrm{Jeans}} = \frac{1}{\gamma\pi} N_{\mathrm{Jeans}}^2 G \rho_{\mathrm{gas}} r_{\mathrm{sys}}^2, \tag{1}$$

where $\gamma = 5/3$, $N_{\mathrm{Jeans}} = 4.0$ and r_{sys} is chosen from the larger one of either the smoothing length or the gravitational softening of an SPH particle. In this prescription, it is ensured that the Jeans length is always resolved with N_{Jeans} system lengths.

A total of 2×512^3 particles are employed for dark matter and gas in a comoving volume of $(100\,h^{-1}\ \mathrm{Mpc})^3$. The mass of a dark matter particle is $5.38\times10^8\,h^{-1}\mathrm{M}_\odot$, and that of a gas

[1] https://grackle.readthedocs.org/

particle is initially $1.00 \times 10^8 \, h^{-1} \mathrm{M}_\odot$. The softening length for the gravitational force is set to be $6 h^{-1}$ kpc in comoving units.

The CELib library (Saitoh, 2016, 2017), which allows a separate treatment of Type II supernovae (SN-II), Type Ia SNe (SN-Ia), and asymptotic giant (AGB) stars, is also employed. Stellar lifetimes, metallicity-dependent metal yields, and mass losses from SN-II, SN-Ia and AGB stars with time-dependent SN rate are considered. The CELib library can treat the evolution of 13 important elements with H, He, C, N, O, Ne, Mg, Si, S, Ca, Fe, Ni, and Eu, which are the major coolants in the ISM.

The following cosmological parameters (Planck Collaboration et al., 2016) are adopted: baryon density parameter $\Omega_\mathrm{b} = 0.049$, total matter density parameter $\Omega_\mathrm{m} = 0.32$, cosmological constant parameter $\Omega_\Lambda = 0.68$, Hubble constant $H_0 = 67$ km s^{-1} Mpc^{-1}, power spectrum index $n_\mathrm{s} = 0.9645$, and density fluctuation normalization $\sigma_8 = 0.831$. The non-dimensional Hubble constant is also used as follows $h \equiv H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.67$.

A brief summary of the main equations that are studied in cosmological simulations of structure formation, focusing on the discretised form of the equations, —describing the dynamics of a collisionless component (dark matter or stars in galaxies) and of an ideal gas (ordinary baryons, mostly hydrogen and helium), both subject to and coupled by gravity in an expanding background space— is presented here, following the model of Springel (2005), as long as a recapitulation of the main processes involving other physical phenomena that are taken into account in the code, following Aoyama et al. (2018); Shimizu et al. (2019).

2.1. Collisionless dynamics

Dynamics of non-interacting dark matter are described by the collisionless Boltzman equation coupled to the Poisson equation in an expanding Friedmann-Lemaître-Robertson-Walker model of Universe. These equations are solved with the N-body method, where phase-space density is sampled with a finite number N of tracer particles. With this scheme, the dynamics are described by the Hamiltonian

$$H = \sum_i \frac{\vec{p}_i^2}{2 m_i \, a(t)^2} + \frac{1}{2} \sum_{ij} \frac{m_i m_j \, \varphi(\vec{x}_i - \vec{x}_j)}{a(t)}, \tag{2}$$

where $H = H(\vec{p}_1, \ldots, \vec{p}_N, \vec{x}_1, \ldots, \vec{x}_N, t)$. The $\vec{x}_i$ are comoving coordinate vectors, given $\vec{p}_i = a^2 m_i \dot{\vec{x}}_i$; the dependence with time comes from the evolution $a(t)$ of the scale factor.

Assuming periodic boundary conditions for a cube of size L^3, the interaction potential $\varphi(\vec{x})$ is the solution of

$$\nabla^2 \varphi(\vec{x}) = 4\pi G \left[-\frac{1}{L^3} + \sum_{\vec{n}} \tilde{\delta}(\vec{x} - \vec{n} L) \right], \tag{3}$$

where the sum over $\vec{n} = (n_1, n_2, n_3)$ extends over all integer triplets. Actually, the solution corresponds to the *peculiar potential*, being the dynamics of the system driven by $\nabla^2 \phi(\vec{x}) =$

$4\pi G[\rho(\vec{x})-\overline{\rho}]$, as the mean density is subtracted here. In a discretised particle system, that potential is defined as

$$\phi(\vec{x}) = \sum_i m_i \varphi(\vec{x}-\vec{x}_i). \tag{4}$$

Density distribution profile for each particle, $\tilde{\delta}(\vec{x})$, is defined as the Dirac δ-function convolved with a normalised gravitational softening kernel of comoving scale ϵ: the quintic spline kernel (Morris, 1996) as previously mentioned, setting $\tilde{\delta}(\vec{x}) = W(|\vec{x}|, \alpha\epsilon)$, where $W(r)$ is given by

$$W(r,h) = \sigma \begin{cases} (3-\frac{r}{h})^5 - 6(2-\frac{r}{h})^5 + 15(1-\frac{r}{h})^5, & 0 \le \frac{r}{h} < 1, \\ (3-\frac{r}{h})^5 - 6(2-\frac{r}{h})^5, & 1 \le \frac{r}{h} < 2, \\ (3-\frac{r}{h})^5, & 2 \le \frac{r}{h} < 3, \\ 0, & \frac{r}{h} > 3. \end{cases} \tag{5}$$

with $\sigma = [1/120, 7/478\pi, 1/120\pi]$ for each case.

If the scale factor is set to 1, the dependence of time in the Hamiltonian disappears; for vacuum boundary conditions, the interaction potential simplifies to its Newtonian form, i.e. $\varphi(\vec{x}) = -G/|\vec{x}|$ for a point mass —modified for small separations by the softening. Independently of the type of boundary conditions, the computation of the gravitational force results in a N^2-scaling for computational cost, as each particle interacts with the rest of them, implying a double sum. However, the force accuracy needed for this kind of dynamics, where particles do not collide, allow the use of approximate force computations by methods such as TreePM, the one that is used on GADGET.

2.2. Hydrodynamics

Smoothed particle hydrodynamics (SPH) uses a set of discrete tracer particles to describe the state of a fluid, with continuous fluid quantities being defined by a kernel interpolation technique. Lagrangian specification, following each fluid element as a particle with coordinates $\vec{r}_i$, velocities $\vec{v}_i$, and masses m_i is the one followed here. The thermodynamic state of each element is defined in terms of the entropy per unit mass, s_i, although an entropic function defined by $A \equiv P/\rho^\gamma$ will be used to work, as it is more convenient, being $A = A(s)$.

The density is estimated in the form

$$\rho_i = \sum_{j=1}^{N} m_j W(|\vec{r}_{ij}|, h_i), \tag{6}$$

where $\vec{r}_{ij} \equiv \vec{r}_i - \vec{r}_j$, and $W(r,h)$ is the SPH smoothing kernel defined in equation (5). The smoothing lengths of each particle, h_i, is defined following

$$\frac{4\pi}{3} h_i^3 \rho_i = N_{\text{sph}} \overline{m}, \tag{7}$$

where N_{sph} is the typical number of smoothing neighbours, and $\overline{m}$ is an average particle mass.

The equations of motion for each SPH particle is given by (Springel & Hernquist (2002))

$$\frac{\mathrm{d}\vec{v}_i}{\mathrm{d}t} = -\sum_{j=1}^{N} m_j \left[f_i \frac{P_i}{\rho_i^2} \nabla_i W_{ij}(h_i) + f_j \frac{P_j}{\rho_j^2} \nabla_i W_{ij}(h_j) \right], \tag{8}$$

where $W_{ij}(h) = W(|\vec{r}_i - \vec{r}_j|$ and the coefficients f_i are defined by

$$f_i = \left[1 + \frac{h_i}{3\rho_i}\frac{\partial \rho_i}{\partial h_i}\right]^{-1}, \tag{9}$$

The particle pressures are given by $P_i = A_i \rho_i^{\gamma}$, and A_i stays constant in such a flow.

Although these equations define fluid dynamics in SPH, given no discontinuities, flows of ideal gases usually develop them in the form of different shocks. The entropy generated by this kink of phenomena need to be captured using an artificial viscosity, that generate entropy at a rate

$$\frac{\mathrm{d}A_i}{\mathrm{d}t} = \frac{1}{2}\frac{\gamma - 1}{\rho_i^{\gamma-1}} \sum_{j=1}^{N} m_j \Pi_{ij} \vec{v}_{ij} \cdot \nabla_i \overline{W}_{ij}, \tag{10}$$

transforming kinetic energy of gas motion irreversibly into heat. $\overline{W}_{ij}$ denotes the arithmetic average of the two kernels $W_{ij}(h_i)$ and $W_{ij}(h_j)$, and the viscosity can take a wide spectrum of parametrisations, such as the one derived by Monaghan (1997),

$$\Pi_{ij} = -\frac{\alpha}{2}\frac{\left[c_i + c_j - 3w_{ij}\right]w_{ij}}{\rho_{ij}}, \tag{11}$$

where $w_{ij} = \vec{v}_{ij} \cdot \vec{r}_{ij} / \left|\vec{r}_{ij}\right|$ is the relative velocity projected onto the separation vector, provided the particles approach each other, i.e. for $\vec{v}_{ij} \cdot \vec{r}_{ij} < 0$, otherwise set $w_{ij} = 0$, and c_i, c_j are the sound speeds of each particle. Viscosity adds an excess pressure to the particles in their motion equation $P_{\mathrm{visc}} \simeq \frac{1}{2}\rho_{ij}^2 \Pi_{ij}$. To avoid angular momentum transport when shear-flows appear, a viscosity-limiter is used, multiplying the viscous tensor by a factor that measures the relative amount of shear flow around a particle.

2.3. Additional physics

In this simulation, stars are formed in dense and cold gas particles whose number density is $n_{\mathrm{th}} > 0.1$ cm^{-3} and temperature is $T_{\mathrm{th}} < 10^4$ K, using the following SFR:

$$\frac{d\rho_*}{dt} = \varepsilon_* \frac{\rho_{\mathrm{gas}}}{t_{\mathrm{ff}}}, \tag{12}$$

where ρ_* and ρ_{gas} are the local mass density of formed stars and gas, ε_* is the star formation efficiency given a free-fall time ($\varepsilon_* = 0.05$ in this simulation), and $t_{\mathrm{ff}} \equiv \sqrt{3\pi/\left(32G\rho_{\mathrm{gas}}\right)}$ is the local free-fall time. The thresholds applied to temperature and density are chosen so that the conditions to meet gas collapse (where Lyman α cooling already allows it, Sutherland & Dopita (1993)) and bounded objects like molecular clouds (following Larson's law, Larson (1981)) are given, respectively. The gas particles that satisfy above conditions spawn star particles stochastically with Chabrier IMF with mass range 0.1 to 120 $\mathrm{M}_\odot$ (Chabrier (2003)).

Metal and dust production by SNe II, SNe Ia and AGB stars is also included in the code. With the mentioned implementation of the CELib package the formation of heavy elements is also taken into account. SN feedback is treated as follows. As the Sedov-Taylor solution (Sedov (1959); Taylor (1950)) gives very small scales both in space and time to resolve the size and the duration of a SN, a subgrid model needs to be implemented. Using the duration time of hot phase (t_{hot}) and the hot SN bubble radius (R_{bub}) obtained by Chevalier (1974) and McKee & Ostriker (1977), in the code only the gas particles within this size are affected by the SN feedback; cooling is turned off for $\Delta t < t_{\mathrm{hot}}$.

To avoid the instantaneous feedback of energy and metal, it takes place following a logarithmic time interval, and the energy and metals deposited in this interval are computed using the mentioned CELib library. The energy that an i-th gas particle takes from SN explosions in its surroundings is, then

$$\Delta E_{\rm i} = \frac{m_{\rm i} W(r_{\rm i}, R_{\rm bub})}{\sum_{\rm j=1}^{\rm N} m_{\rm j} W(r_{\rm j}, R_{\rm bub})} E_{\rm SN}, \tag{13}$$

where $E_{\rm SN}$ is the SN energy from the star particle, W is the SPH kernel function as defined above and $r_{\rm i}$ is the distance between i-th gas particle and the star particle; $R_{\rm bub}$ is used as the smoothing length for SN feedback. Gas and metals ejected by the SN are distributed in the same manner. SN II and SN Ia just have a different time delay in their treatment, being similar in all other ways. The wind velocity of the SN is computed as follows

$$V_{\rm wind} = \sqrt{\frac{2E_{\rm K}^{\rm SN}}{m_{\rm gas}}}, \tag{14}$$

to avoid energy conservation violation. $E_{\rm K}^{SN}$ is the kinetic SN feedback energy received by a gas particle (the energy of the SN is received by each particle fractionated in kinetic energy and thermal energy) , and $m_{\rm gas}$ is the mass of a gas particle. The direction of wind particles is randomly chosen.

Energy input from massive stars (stellar feedback) is also considered. UV radiation and stellar winds emitted by young, massive stars ionize and heat up the gas in its surroundings, suppressing star formation and changing the gas distribution because of the thermal pressure. This 'early stellar feedback' (ESFB) may be one of the solutions to overcome the overcooling problems in cosmological simulations. In order to reproduce the stellar-halo mass ratio, it has been stated that 10% of total stellar radiation from massive young stars is necessary (Stinson et al. (2013)); however, other authors argued it is not necessary to reproduce the observations (Dalla Vecchia & Schaye (2012); Keller et al. (2014, 2015, 2016)). So far, a model similar to Stinson et al. (2013) has been implemented in the code, without artificially turning off radiative cooling, and setting and small enough minimum time step to resolve ESFB and SN events.

The model presented above is successful in reproducing the main statistical properties of star formation and stellar content in galaxies, especially, the cosmic star formation rate density (SFRD) and the stellar mass function at $z = 0$ (Schiminovich et al. (2005); Bouwens et al. (2009); Reddy & Steidel (2009); Moustakas et al. (2013); Tomczak et al. (2014)), except at the very massive end of the stellar mass function; this might be resolved by including feedback of active galactic nuclei (AGN), that is currently not taken into account in the code.

The dust evolution model that is used in the code is the one presented in Hou et al. (2017). The whole grain population is represented by small grains (radius $a < 0.03\mu$) and large grains ($a > 0.03\mu$), according to Hirashita (2015). The typical radii are set as $0.1\,\mu$m and $5\times10^{-3}\,\mu$m, respectively. The abundances of each population are represented by the ratio between the mass of dust grain an the

mass of a gas particle. Time evolution of the two population's abundances is given by

$$\begin{aligned}\frac{\mathrm{d}\mathcal{D}_{\mathrm{L}(i)}(t)}{\mathrm{d}t} &= -\left(\frac{\mathcal{D}_{\mathrm{L}(i)}(t)}{\tau_{\mathrm{sh}}} - \frac{\mathcal{D}_{\mathrm{S}(i)}(t)}{\tau_{\mathrm{co}}}\right) - \frac{\mathcal{D}_{\mathrm{L}(i)}(t)}{\tau_{\mathrm{sp}}(a_{\mathrm{L}})} \\ &\quad + \left[\frac{\mathrm{d}\mathcal{D}_{\mathrm{L}(i)}(t)}{\mathrm{d}t}\right]_{\mathrm{Source}} - \left[\frac{\mathrm{d}\mathcal{D}_{\mathrm{L}(i)}(t)}{\mathrm{d}t}\right]_{\mathrm{SNe}} \\ &\quad - \frac{\mathcal{D}_{\mathrm{L}(i)}(t)}{m_{\mathrm{g}(i)}}\frac{\mathrm{d}m^{\mathrm{return}}_{\mathrm{g}(i)}}{\mathrm{d}t}, \end{aligned} \tag{15}$$

$$\begin{aligned}\frac{\mathrm{d}\mathcal{D}_{\mathrm{S}(i)}(t)}{\mathrm{d}t} &= \left(\frac{\mathcal{D}_{\mathrm{L}(i)}(t)}{\tau_{\mathrm{sh}}} - \frac{\mathcal{D}_{\mathrm{S}(i)}(t)}{\tau_{\mathrm{co}}} + \frac{\mathcal{D}_{\mathrm{S}(i)}(t)}{\tau_{\mathrm{acc}}}\right) \\ &\quad - \frac{\mathcal{D}_{\mathrm{S}(i)}(t)}{\tau_{\mathrm{sp}}(a_{\mathrm{S}})} - \left[\frac{\mathrm{d}\mathcal{D}_{\mathrm{S}(i)}(t)}{\mathrm{d}t}\right]_{\mathrm{SNe}} \\ &\quad - \frac{\mathcal{D}_{\mathrm{S}(i)}(t)}{m_{\mathrm{g}(i)}}\frac{\mathrm{d}m^{\mathrm{return}}_{\mathrm{g}(i)}}{\mathrm{d}t}, \end{aligned} \tag{16}$$

where $\mathcal{D}_{\mathrm{L}}$ and $\mathcal{D}_{\mathrm{S}}$ are the abundances of each population (large and small), (i) refers to the gas particle with which the mass is being compared, τ_{sh}, τ_{co}, and τ_{acc} are different time-scales associated with dust grains (shattering, coagulation and accretion), $\mathrm{d}m^{\mathrm{return}}_{\mathrm{g}(i)}/\mathrm{d}t$ is the gas ejection rate from stars, $\tau_{\mathrm{sp}}(a)$ is the sputtering time-scale as a function of grain size in the hot gas not associated with SNe, and the terms with 'Source' and 'SNe' account for the dust production in stars and dust destruction in SN. Every time-scale depends on gas density, dust abundance and metallicity in this model. Accretion and coagulation occur in scales that cannot be resolved in this simulation, in dense gas particles, so a subgrid model is adopted for these phenomena. Shuttering, on the contrary, occurs only in the diffuse gas particles. Finally, where the temperature is high enough ($\gtrsim 10^6$ K), dust grains are destroyed by sputtering; as dust destruction has been already counted in SN shocks, a density threshold is considered to avoid overcounting. To differentiate the dust associated to galaxies and that contained in the IGM, the code uses P-Star groupfinder (Springel et al., 2001).

2.4. Outputs of the simulation

With this configuration, the simulation that has been provided to work with consists of the following outputs, in the form of mesh data grids:

- Dark matter density field, in units of g/cm^3
- Gas density field, in units of g/cm^3
- Neutral hydrogen number density (nHI), in cm^{-3}
- Gas temperature, in K
- Optical depths (τ) with the line of sight from 3 different directions: x, y and z

Every field comes in three different resolutions: 128^3 particles, 256^3 particles and 512^3 particles. Also, the size of the side of the simulation is $100h^{-1}$Mpc, and the data corresponds to $z \sim 2$. Initial conditions have been generated with MUSIC (Hahn & Abel (2011)).

3. Methodology: Bias Assignment Method (BAM)

In this section, the methodology used to map the baryonic physics extracted from hydrodynamical simulations, as presented in section 2, onto a dark matter field is described.

In particular the Bias Assignement Method (`BAM`) developed by Balaguera-Antolínez et al. (2019, 2020) is used. As this computing code was envisaged to reproduce integer halo number counts per cell in a mesh, it had to be extended to deal with continuous quantities. In the following the `BAM` code will be recapped and the novel contributions of this work will be described in detail.

The former idea behind `BAM` is to extract a bias scheme from the dark matter halo distribution from a time costly, high resolution, reference N-body simulation, and use it to map it onto a dark matter field obtained by applying approximate gravity solvers, such as `FastPM` (Feng et al., 2016) or the Augmented Lagrangian perturbation theory (`ALPT,` Kitaura & Hess, 2013), to the same initial conditions of the reference simulation in a lower resolution. Previous works, such as `PThalos` (Scoccimarro & Sheth, 2002), `MoLUSC` (Sousbie et al., 2008), `PATCHY` (Kitaura et al., 2014), `QPM` (White et al., 2014), `EZMOCKS` (Chuang et al., 2015) and `HALOGEN` (Avila et al., 2015), have studied the idea of mapping the halo distribution onto a downgraded dark matter density field, but `BAM` does not assume any analytical form for the halo-bias relation, using a statistical interpretation of the halo-bias (e.g. Dekel & Lahav, 1999), and computing a parameter-free probability distribution, leading to the extraction of all possible unknown dependencies or non-linearities. The power of the method is that, then, the mapping can be applied onto any other initial conditions that contain the same cosmological parameters. This lays on the concept of the extension to a multi-dimensional bias relation of the rank ordering method proposed by Weinberg (1992).

`BAM` has been developed to generate mock catalogues with a power spectrum that matches the one of the high resolution halo distribution reference, based on the calibration of the halo bias obtained from one realisation of that reference, by producing both a kernel —whose main purposes are to correct any large scale contributions coming from non-local dependencies that may have not been taken into account inside `BAM` (see McDonald & Roy, 2009) and to correct any aliasing effects arisen when placing the dark matter field and the halo distribution on a mesh, instead of the original algorithm used on the reference— and a halo-bias that will be used on the production of that mocks. Here it will be shown how that calibration and mock generation processes work, which changes have been introduced to generalize this method from a discrete distribution (namely, halo number counts) to a continuous one (i.e. gas density fields, described in section 2), and the particularities of the reference simulation used. To illustrate the procedures involved in the method, a flow-chart describing the steps has been made and represented on Fig. 1 (credits to Balaguera-Antolínez et al., 2020); its notation will be used hereafter.

3.1. Calibration

The first step that is taken in a general process is to down-sample the original initial conditions of the reference into a lower resolution white noise. Then, the Fourier Transform is carried out onto this white noise, multiplying the result by the square root of the linear matter power spectrum in spherical shells ($P_{\rm lin}(|\mathbf{k}|)$) evaluated at the redshift of that initial conditions. The result will be evolved using approximate gravity solvers, as mentioned before, up to $z = 1$, generating an approximated dark matter density field $\delta^i_{\rm dm}(\mathbf{r})$, using a cloud-in-cell mass assignment scheme. Phase-space mapping (Abel et al., 2012; Hahn et al., 2013) is also used on this step, so three-point statistics are reasonably reproduced on the mocks obtained (see Pellejero-Ibañez et al., 2020). Particularly, this first step has not been carried out in the case of the reference simulation used in this work, as the dark matter field was already downgraded to a maximum resolution of a 512^3 mesh and evolved to $z = 2$ by default.

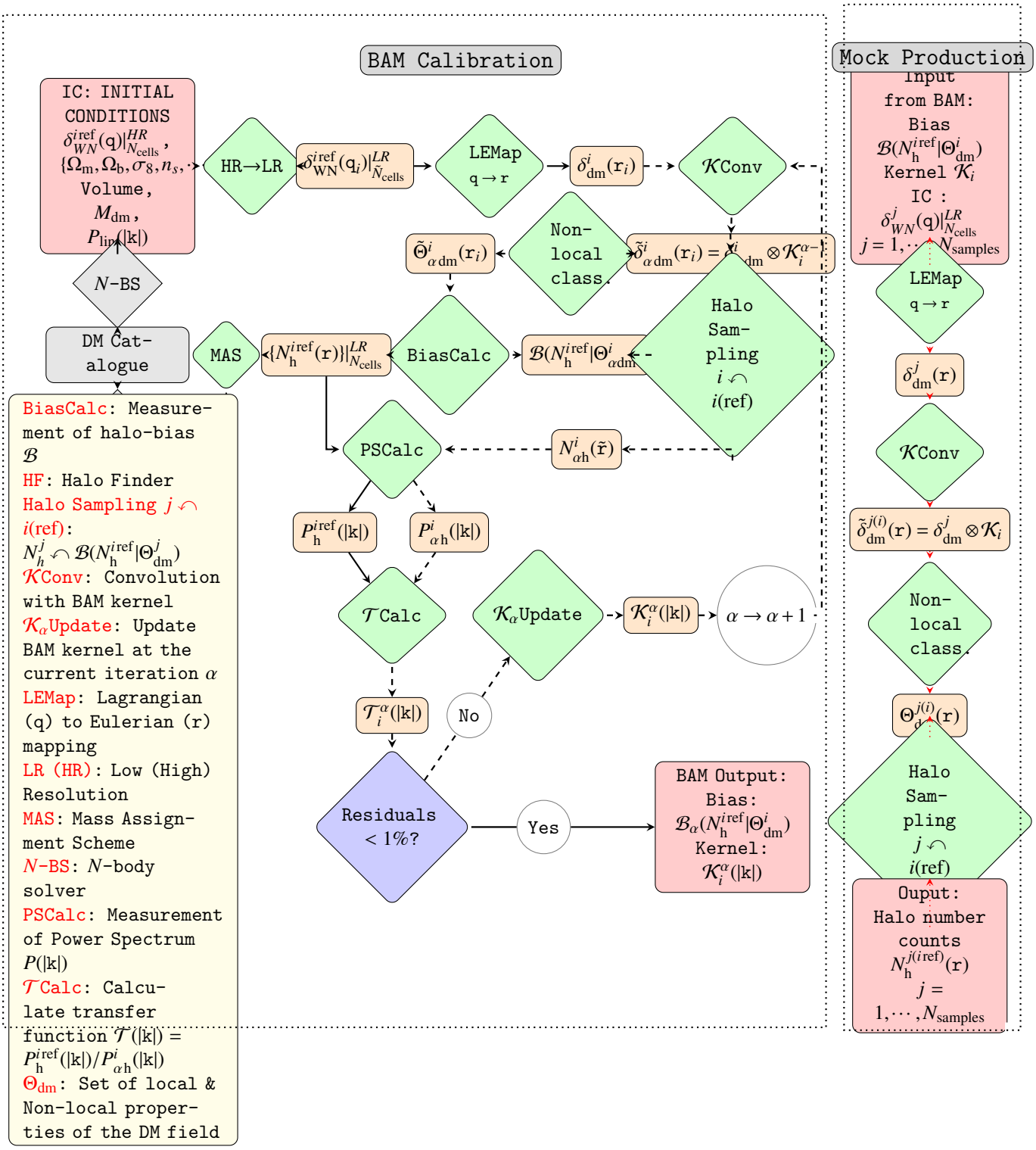


Figure 1: Credit to Balaguera-Antolínez et al. (2020). Flow-chart representing the operations performed within `BAM`. On the left-hand side box (`BAM Calibration`) the iterative process developed in order to calibrate the kernel $\mathcal{K}$ and the halo-bias using *one* realisation of a halo catalogue from the reference simulation, $\{\mathbf{r}_{\rm h}^{i\,\rm ref}\}$, in combination with a dark matter density field obtained from the approximated gravity solver with the corresponding initial conditions $\delta_{\rm WN}^{i\rm ref}(\mathbf{q})$ is shown. Solid line arrows denote the process involving the reference N-body simulation, performed once. Dotted-line arrows denote the iterative process, which finishes once the relative residuals (see text) are below 1%. The index α is assigned to quantities being updated within the iterative process, while i identifies a white noise from the reference simulation. For the first iteration the kernel is $\mathcal{K}_i^{\alpha=0}(|\mathbf{k}|)=1$. The right hand side box (`Mock Production`) depicts the process followed to generate mock halo density fields based on the outputs of `BAM`.

Iteration zero of the method starts with the measurement of the halo-bias (given its stochastic interpretation, see Dekel & Lahav, 1999; Sigad et al., 2000), calculated as the probability of finding a number of dark matter haloes, $N_{\rm h}$, within a cell of the approximated dark matter density field, conditional to number of local or non-local properties of that cell, $\mathcal{N}_{\rm p}$. To extend the analysis from just considering the dependence of the halo-bias with the dark matter density field, there are two non-local properties that are taken into account in the analysis if desired: the cosmic-web classification (CWC) into knots, sheets, filaments and voids, based on the eigenvalues of the tidal tensor of the dark matter field (Hahn et al., 2007; Paranjape et al., 2018), and the mass of collapsing regions —established using a friend-of-friends algorithm in the cells classified as knots—, M_k (Zhao et al., 2015). The kernel —a Dirac's delta function on this first iteration— is meant to take into account any other non-local dependencies that have not been considered, although adding them to the analysis could, of course, could improve the three-point statistics. The halo-bias is then computed considering the joint probability distribution of $N_{\rm h}$ and the properties of the dark matter density field mentioned above, and normalizing that distribution, as follows:

$$\mathcal{B}(N_{\rm h}|\Theta_{\rm dm}) = \frac{\sum_{i=1}^{N_{\rm cells}} \mathbf{1}_{N_h}(N_{\rm h}(\mathbf{r}_i)) \prod_{\kappa=1}^{\mathcal{N}_{\rm p}} \mathbf{1}_{\gamma_\kappa}(\{\Theta_{\rm dm}(\mathbf{r}_i)\}_\kappa)}{\sum_{i=1}^{N_{\rm cells}} \prod_{\kappa=1}^{\mathcal{N}_{\rm p}} \mathbf{1}_{\gamma_\kappa}(\{\Theta_{\rm dm}(\mathbf{r}_i)\}_\kappa)}, \tag{17}$$

where $\Theta_{\rm dm}$ is the set of $\mathcal{N}_{\rm p}$ used to characterize the dark matter cell (in particular, with the properties mentioned above, $\Theta_{\rm dm} \equiv \{\log_{10}(2+\delta_{\rm dm}), \mathrm{CWC}, M_K\}$, and being $\delta_{\rm dm}$ the dark matter overdensity), $\gamma_\ell \equiv [\{\Theta_{\rm dm}\}_\ell - \Delta_\ell/2, \{\Theta_{\rm dm}\}_\ell + \Delta_\ell/2)$ represents the bins in which the field is divided, with a width of Δ_ℓ, defined for each of the ℓ properties of the dark matter density field described in $\Theta_{\rm dm}$. $\mathbf{1}_A(x)$ is an *indicator function* that indicates membership of an element in a subset A of x, having the value 1 for all elements of A and the value 0 for all elements of x outside A. This kind of representation of the halo-bias does not account for different effects that appear in the bias, such as correlation between particles placed in different property bins; however, the iterative process that takes place afterwards captures those aspects.

The iterative process now begins: for each α-th iteration the bias $\mathcal{B}(N_{\rm h}^{i\,\rm ref}|\Theta^i_{\alpha\rm dm})$ is computed using Eq. (17), being $N_{\rm h}^{i\,\rm ref}$ the halo number counts in each cell, using a mass assignment scheme such as nearest-grid-point or cloud-in-cell, from the reference. Then, the dark matter density field is convoluted with the kernel obtained on the previous iteration, as shown in Eq. 18. Finally, a halo number count $N^i_{\alpha\rm h}(\mathbf{r})$ is assigned to each point of the mesh following the halo-bias extracted in the first step, therefore obtaining the halo density field; this process is represented on Eq. 19, and it is constructed on a way that the new halo density field maintain the statistics and the total number of counts as the reference.

$$\tilde{\delta}^i_{\alpha\rm dm}(\mathbf{r}) \equiv \left(\mathcal{K}_i^{\alpha-1} \otimes \delta^i_{\rm dm}\right)(\mathbf{r}) \tag{18}$$

$$\{N^i_{\alpha\rm h}(\mathbf{r})\} \curvearrowleft \mathcal{B}\left(N_{\rm h}^{i\,\rm ref} \mid \Theta^i_{\rm dm} = \Theta^i_{\alpha\rm dm}(\mathbf{r})\right), \tag{19}$$

A phase-independent transfer function is built using the power spectrum of the halo density field obtained at α-th iteration, $P^\alpha_{i\rm h}(k)$, following Eq. 20. Carrying out a Metropolis-Hasting (MH) algorithm in each spherical shell where the power spectrum is measured, computing $\min(1, \exp(\mathcal{H}^2_{0j} - \mathcal{H}^2_{1j}))$ with $\mathcal{H}_{ij} = (P_{j,\rm ref} - P_{i,j})/\sigma_j$, σ_j the variance associated to the reference power spectrum, assumed to be a gaussian (Feldman et al., 1994). Whether the MH algorithm accepts a certain value or not, a weight ω^α_i is defined, following Eq. 21.

$$\mathcal{T}^\alpha_i(k) \equiv \frac{P_{\rm h}^{i\,\rm ref}(k)}{P^\alpha_{i\rm h}(k)}, \tag{20}$$

$$\omega_i^{(\alpha)}(k) \equiv \begin{cases} \mathcal{T}_i^{\alpha}(k) & \text{if } MH = 1 \\ 1 & \text{if } MH = 0, \end{cases} \tag{21}$$

The kernel is then updated as the product of the α-th weigth and all the previous ones, basically assigning to the kernel its previous value if the α-th transfer function is not accepted.

$$\mathcal{K}_i^{(\alpha)}(k) = \prod_{\ell=1}^{\ell=\alpha} \omega_i^{(\ell)}(k) \tag{22}$$

Finally, Eq. 18 is applied again, convoluting the new obtained kernel with the dark matter density field to step into the next iteration. The method is considered to converge when the residuals of this transfer function over the number of spherical shells in Fourier space $N_{\rm F}$ used to measure the power spectra, namely $\sum_\kappa |T_i^{\alpha}(k_\kappa) - 1|/N_{\rm F}$, reach a value under 1%; after that, the outputs required for mock generation are computed.

3.2. Generation of mock density fields

The steps followed on the production of mock halo density field are presented on the right-hand side panel of Fig. 1. A set of new approximate dark matter density fields $\delta_{\rm dm}^{j}$ are generated by evolving a set of different realisations of the initial conditions of the reference simulation, after down-sampling them if necessary. These dark matter density fields are then covolved with the kernel obtained from the iterative process, following Eq. 18 (for each j on the set), $\tilde{\delta}_{\rm dm}^{j(i\,{\rm ref})}(\mathbf{r}) \equiv \left(\mathcal{K}_i \otimes \delta_{\rm dm}^{j}\right)(\mathbf{r})$. Non-local properties that are now implemented in the method, previously mentioned, are computed for the new $\tilde{\delta}_{\rm dm}^{i(j)}$. Then, on a similar approach as the one followed on the iterative process, halo number counts are assigned to each point of the dark matter density fields, recalling Eq. 19:

$$\{N_h^{j(i\,{\rm ref})}(\mathbf{r})\} \curvearrowleft \mathcal{B}\left(N_{\rm h}^{i\,{\rm ref}} \mid \{\Theta_{\rm dm}^{i}\}_\ell = \{\Theta_{\rm dm}^{j(i\,{\rm ref})}(\mathbf{r})\}_\ell\right), \tag{23}$$

Here $j(i\,{\rm ref})$ denotes that the i-th realisation has been used as the calibration for the j-th halo density field.

With this approach, `BAM` is able to account different contributions to the stochastic nature of the bias relation, such as the one that arises from relate number count of objects with a dark matter field, or the one that comes from the limited volume of the reference simulation (cosmic variance). As the whole bias relation is extracted from the reference simulation, the first contribution is completely taken into account. Regarding the second contribution, it has already been shown (Balaguera-Antolínez et al., 2020) that the effect of cosmic variance in the halo bias has little relative importance when they come from the calibration process described above. Also related to this contribution comes the fact that the mean large scale power spectrum are impacted significantly by the distribution of halo number counts; using the same initial conditions as the reference simulation and a proper-computed kernel take into account this effect.

3.3. Extension of the `BAM` method to a continuous distribution

As it has been previously mentioned, `BAM` was conceived as a method to extract a bias scheme from a density field populated with dark matter haloes. `BAM` measured the number counts of haloes associated to each point of the dark matter density field in a reference N-body simulation, that number belonging to the discrete distribution of the natural numbers, $\mathbb{N}$. For each new iteration, a

natural number had to be assigned to each cell following what the halo-bias was pointing out, and a new halo density field was generated.

However, the reference simulation that has been used in this work does not follow a discrete distribution, based on number counts, for the density fields of the tracers that are being taken into consideration, but a continuous one, based on physical quantities such as the gas, the neutral hydrogen density, the temperature or the optical depth, as it has been stated in section 2. These magnitudes cannot be represented by the set of natural numbers, but by the set of real numbers, $\mathbb{R}$, so `BAM` could not be applied with its original configuration to the reference simulation. Therefore, some changes had to be considered in the method to generalize it to this kind of distributions, while maintaining all of its features untouched, as it has demonstrated to be a powerful tool. In practice Eq. 23 is generalized to

$$\{\delta_{\mathrm{X}}^{j(i\mathrm{ref})}(\mathbf{r})\} \curvearrowleft \mathcal{B}\Big(\delta_{\mathrm{X}}^{i\mathrm{ref}} \mid \{\Theta_{\mathrm{dm}}^{i}\}_{\ell} = \{\Theta_{\mathrm{dm}}^{j(i\mathrm{ref})}(\mathbf{r})\}_{\ell}\Big), \tag{24}$$

with X being the target continuous quantity.

The approach behind the changes that have been implemented is to transform, in some way, the continuous distribution into a discrete one, so `BAM` could address its treatment maintaining all the procedures in which it relies on. The first step taken is to transform the continuous density field into an overdensity field, and then take the logarithm of base 10 out of it, following Eq. 25.

$$\delta = \frac{\rho(\mathbf{r}) - \rho_{mean}}{\rho_{mean}} \rightarrow Y = \log_{10}(2+\delta) \tag{25}$$

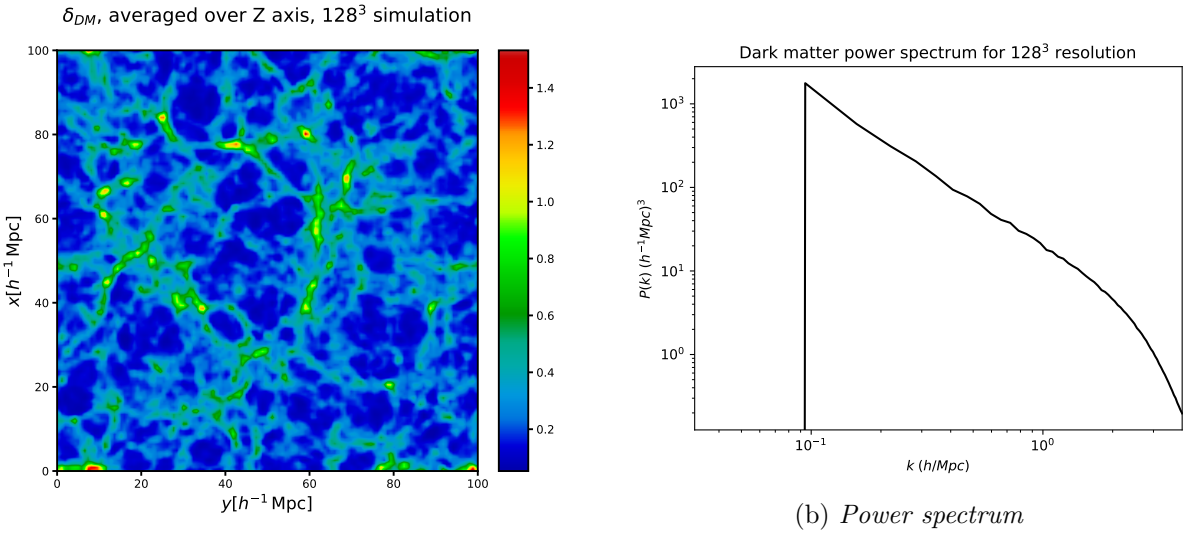
The next step is to define a binning between two numbers, one below $min(Y)$ and the other above $max(Y)$, as if a histogram of the overdensity field were to be made. A set of bins is therefore created, being the first being numbered as 1, and the last bin numbered as N, where N is the desired number of bins used for the tracer density field, each of them with its associated limits. The mesh of the tracer density field is then analyzed, assigning to the density on each point the bin where it is placed, getting as a result a new field that can be assimilated to a number count, discrete, density field, that `BAM` can use in its calibration procedure, as pointed in section 3.1. After the calibration has ended, a mock number count density field is generated, so it has to be turned into a continuous field again, using Eq. 26.

$$\rho(\mathbf{r}) = \rho_{mean}(1 + 10^{Y_{min} + (n_{bin} + \frac{1}{2}) * \Delta_{bin} - 2}) \tag{26}$$

Where Y_{min} is the minimum value of Y as calculated with Eq. 25, n_{bin} is the bin number where a certain point of the mesh is placed, and Δ_{bin} is the width of the bin. As it can be seen, the value assigned to the density field is the middle value of the limits of each bin.

4. Results and analysis

Here, the results from running the `BAM` process of calibration —as shown in section 3.1— to the reference hydrodynamical simulation, after modifying it as specified in section 3.3 to adapt it to the characteristics of the simulation, are presented. The evolution of the power spectra calculated by `BAM` in different steps of the calibration process, as well as that same picture for the so named kernel,



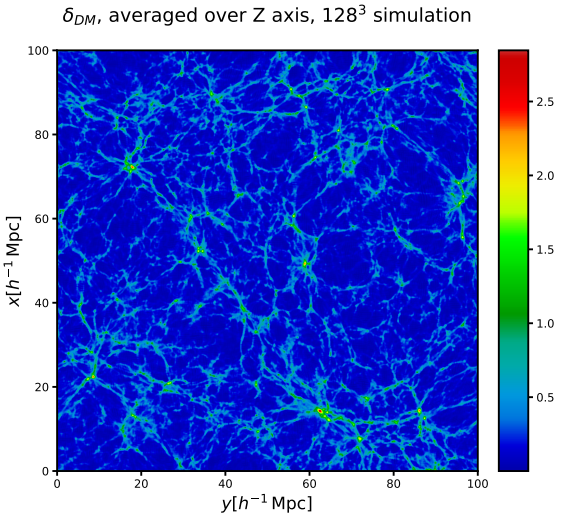
(a) *Density field*

(b) *Power spectrum*

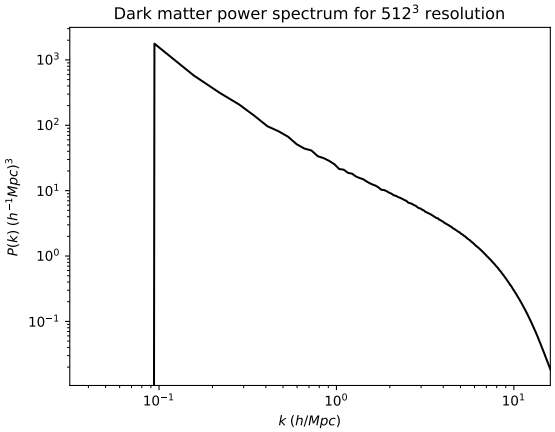
Figure 2: *Both images present the dark matter density field on its* 128^3 *resolution.* ***Left****: dark matter density field map; the original field has been converted to* δ*, and afterwards converted into a logarithmic scale using* $\log_{10}(2+\delta_{Temp})$*. The Z direction of the 3D field has been averaged over 5 .****Right****: dark matter power spectrum.*

will be shown. For every calibration done, the tracer density field generated by `BAM` will be compared to the one of the reference, so its differences and similarities can be appreciated. Along this, the steps taken into improving some of the non-successful initial results will be explained, showing the results brought by this process. As explained above the dark matter field defined on a mesh is taken as the basis (shown for the low and high resolution cases in Figures 2 and 3, respectively).

The method has been applied to map the gas density field in both the 128^3 and the 512^3 resolutions; being the first of them a way to approach how the method worked and understanding which changes had to be introduced in the code so that it could be applied to the reference simulation used, and the second of them a way to confirm that the modifications were effective and that `BAM` could be used in such a way that a continuous distribution is mapped correctly. After that, `BAM` was applied to the neutral hydrogen number density field on its 128^3 resolution; although the method was working, some problems that dealt to not satisfactory results arose, and different approaches were taken into solving them, before trying to use the non-local information of the dark matter field that have been accounted into `BAM`, the cosmic web classification and the mass of the collapsing regions. With this last approach, the results improved greatly; however, there are still some loose ends that will be solved with further —and ongoing— work. The performance has also been tried out with the gas density field, proving to be a powerful tool, as will be shown. Finally, `BAM` has been applied to the 128^3 resolution temperature field, showing how the method did not converge with its initial configuration, and leading to an attempt of a more efficient binning on the tracer distribution that improved the convergence up to $k \approx 2\,h\,\mathrm{Mpc}^{-1}$.



(a) *Density field*



(b) *Power spectrum*

Figure 3: *Both images present the dark matter density field on its* 512^3 *resolution.* ***Left****: dark matter density field map; the original field has been converted to* δ*, and afterwards converted into a logarithmic scale using* $\log_{10}(2+\delta_{Temp})$*. The Z direction of the 3D field has been averaged over 5 .****Right****: dark matter power spectrum.*

4.1. Gas density field

4.1.1. Low resolution case: mesh with 128^3 cells

The bias relation between the dark matter density field and the gas density field extracted from the reference is shown at Figure 4. There is a clear correlation between the gas density and the underlying dark matter field; although the scatter relation is not trivial, it will be mapped with `BAM`. The power spectrum of the gas density field on its 128^3 resolution configuration is firstly investigated. For this analysis, no non-local properties of the dark matter field have been taken into consideration within the process. Figure 5 shows the evolution of the power spectrum calculated by `BAM` during its iterative process (section 3.1), extracted from the gas density field obtained from mapping the bias relation onto each point of the mesh in each iteration. The process has been represented in different panels so it could be clearly seen how the evolution takes place. In the first panel (a) it can be seen how, for the very first iteration (iteration 0), the power spectrum generated by `BAM` (red, dashed line) does not fit well enough with the reference one (black, solid line), assigning approximately a 25% more power to every mode up to $k \approx 2.5h\mathrm{Mpc}^{-1}$; after that, it assigns up to a $\sim 60\%$ less power to small scales than they have in the reference. After only 20 iterations, its remarkable how the maximum deviation from the reference power spectra is just about a $\sim 15\%$ below its actual level, and just for a single mode; larger and smaller scales are all within a 5% deviation (light grey in the lower subpanel of every graphic), and just in the middle scales there are some modes whose deviation is above a 10%. After 40 iterations, every mode is below a 5% deviation from the reference, and most of them are contained below a 1% deviation (dark grey in the lower subpanel of every graphic). In approximately 60 iterations, the method seems to have converged, capturing every mode below a 1% deviation.

Every iteration that has occurred after this point does not change the picture very much, as the Metropolis-Hasting algorithm has ceased to accept new values for the transfer function defined in

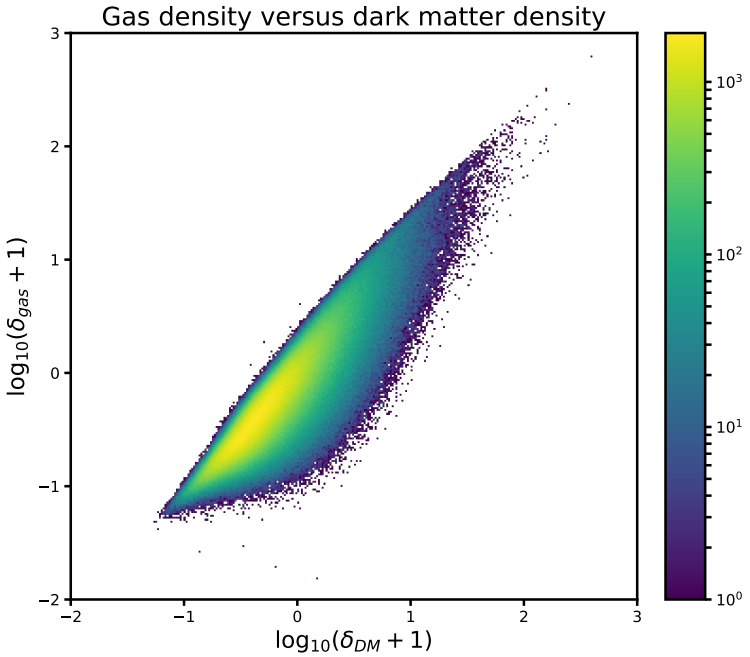
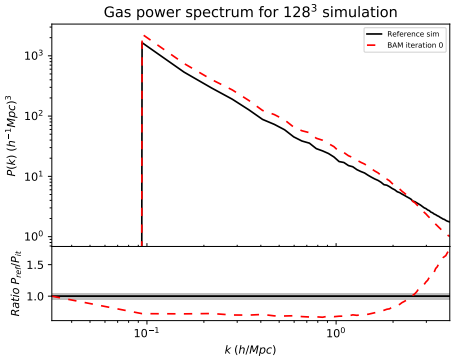


Figure 4: *Two-dimensional histogram relating dark matter density field with gas density field for the* 128^3 *resolution simulation. Both fields have been converted to* δ*, and afterwards converted into a logarithmic scale using* $\log_{10}(1+\delta)$*.*

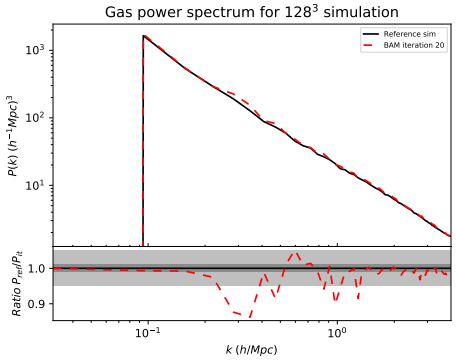
Eq. 20, assigning a value of 1 to almost every new weight that is computed, following Eq. 21. Of course, some fluctuations can be seen, as some values might get accepted, but the final result does not change very much, with almost every mode within a 1% deviation from the unity.

These results can now be complemented by the image shown in Figure 6, where the evolution of the kernel, computed following Eq. 22, is represented. Noticeably, and in well agreement with how the concept of kernel is conceived inside `BAM`, the one associated to the very first iteration resembles the ratio between the power spectra of the reference and that generated by the method —actually, they are the same, as the transfer function is precisely defined as that quotient, and the only term of the kernel is that one. On agreement with what has been previously stated, the difference between the kernel at iteration 40 and the one at iteration 100 is still perceptible, but after iteration 60, the kernel has converged to its final form, with only slight fluctuations amongst iterations 60, 80 and 100.

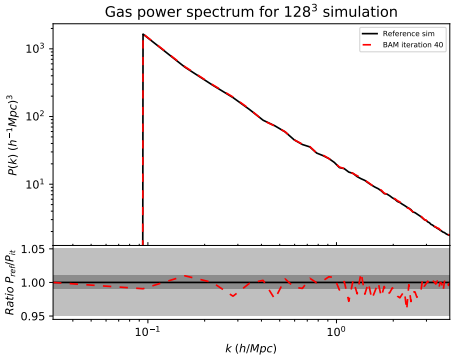
Finally, the density field of the reference and the one generated by `BAM` at it last iteration are presented on Figure 7. Both of the gas density fields have been converted to overdensity, and then to a logarithmic scale, following Eq. 25; also, in both cases, an average of the same five adjacent slices along Z axis have been done so the fields can be fairly compared. On the left panel, the gas density field of the reference is shown; on the right panel, the same for the mock generated by `BAM`. The exact same structures can be observed in both of them, showing the power of the method to generate mock density fields: a box of $100\,h^{-1}\,\mathrm{Mpc}$ of side, as this one, with a resolution of 128^3, shows how structures up to a scale of $\sim 1\,h^{-1}\,\mathrm{Mpc}$ can be reproduced without any important noticeable difference with respect to the reference. Voids might be slightly more populated than the reference's, however, but that could be an effect of not having used any non-local property in this analysis.



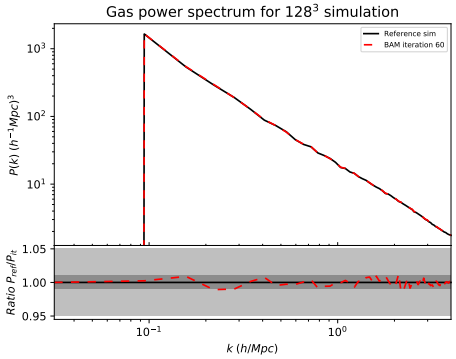
(a) *Iteration 0*



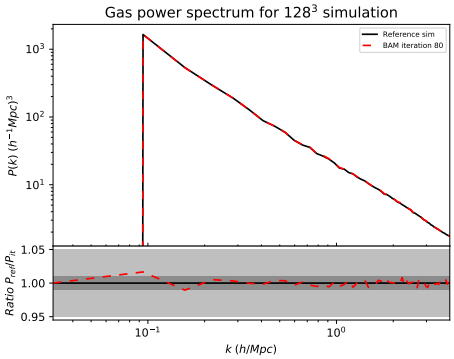
(b) *Iteration 20*



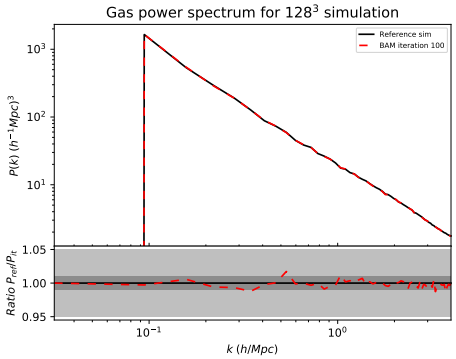
(c) *Iteration 40*



(d) *Iteration 60*



(e) *Iteration 80*



(f) *Iteration 100*

Figure 5: *Gas' power spectrum from the reference* 128^3 *simulation is represented with a black line. BAM's generated mock's power spectrum is represented with a red line. As it can be seen, as BAM advances in its iterations, updating its Kernel, it subsequently creates a mock whose power spectrum approaches the reference's until it is on top of it, with an accumulate residuals* $< 1\%$*. On the bottom panel corresponding ratios with respect to the reference are shown. The different shaded areas stand for 1 and 5% difference with respect to 1.*

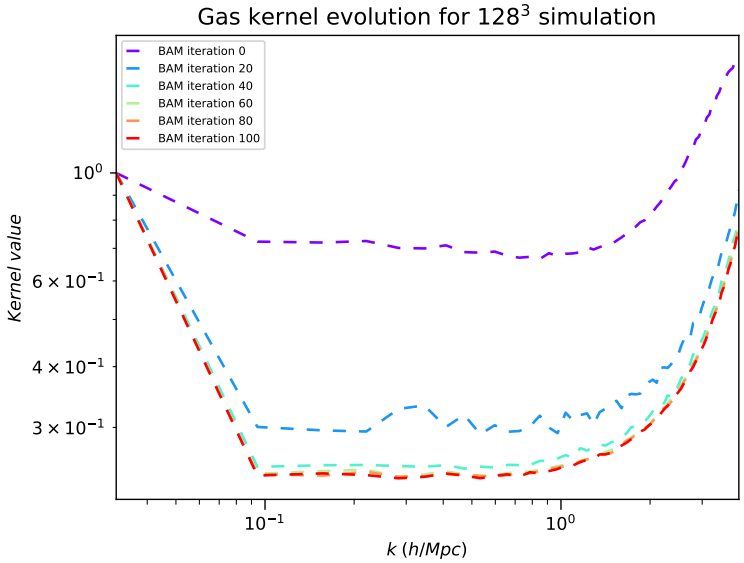
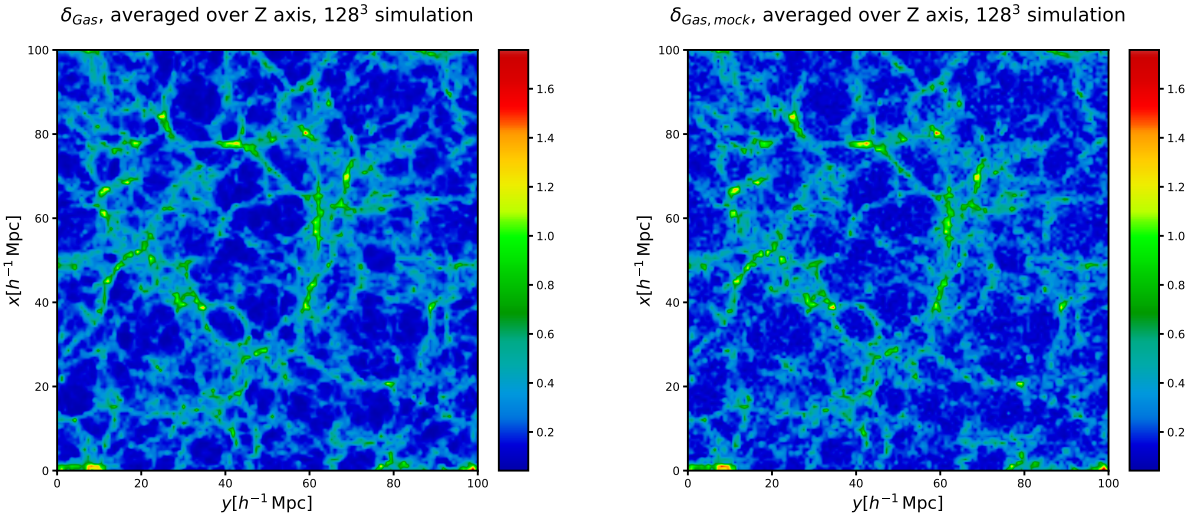


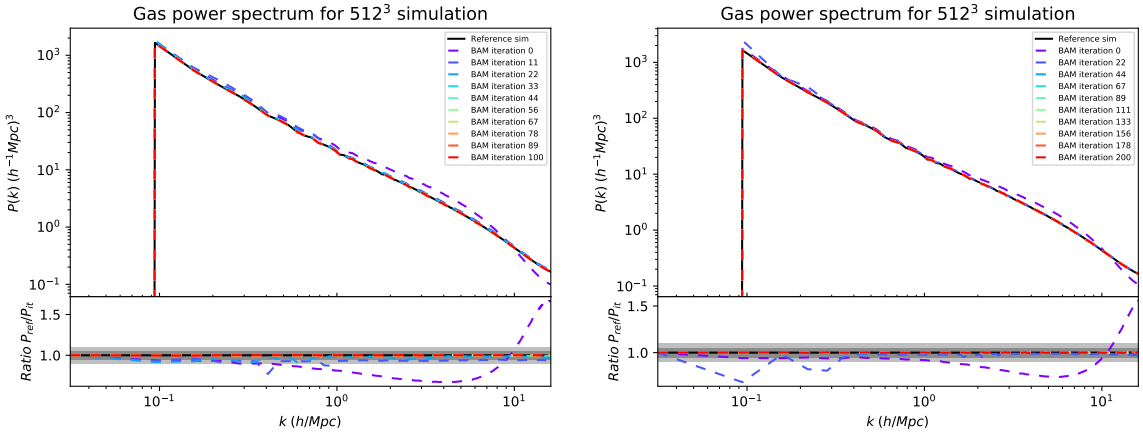
Figure 6: *Evolution of BAM's kernel as iterations advance, for the* 128^3 *gas reference. Kernel value is computed for each k mode with Eq. 22. As it can bee seen, it approaches a certain value, as the power spectrum generated by BAM approaches the reference's, and the Metropolis-Hasting algorithm ceases to accept new values, as specified in Eq. 21.*



(a) *Gas density field, reference*

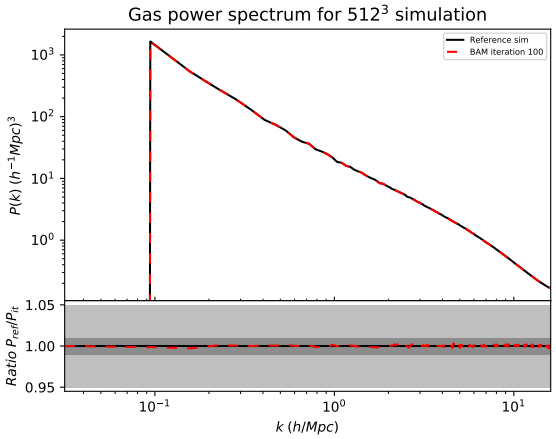
(b) *Gas density field, BAM*

Figure 7: ***Left****: reference gas density field for the* 128^3 *resolution simulation.* ***Right****: gas density field generated by BAM for the* 128^3 *resolution simulation. For the generation of this mock, no cosmic web classification has been used. In both cases, the original field has been converted to* δ*, and afterwards converted into a logarithmic scale using* $\log_{10}(2+\delta_{gas})$*. The Z direction of the 3D field has been averaged over the same 5 slices in both fields.*

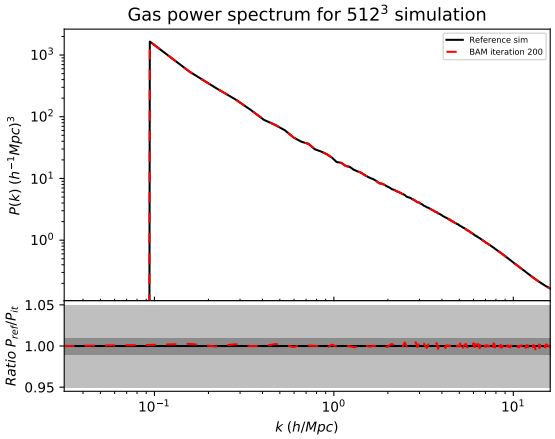


(a) *Power spectrum evolution, no CWC*

(b) *Power spectrum evolution, CWC*

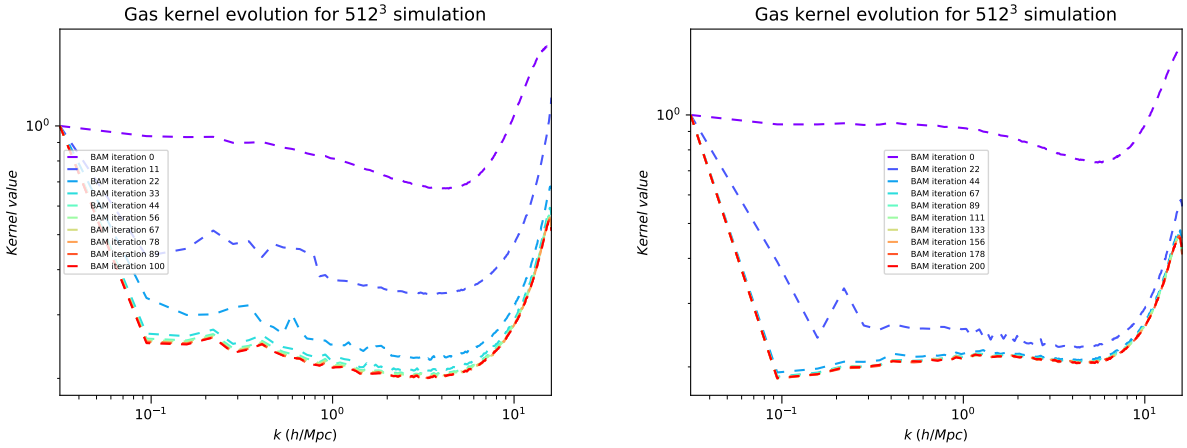


(c) *BAM's last iteration, no CWC*



(d) *BAM's last iteration, CWC*

Figure 8: *Every image presents results from the gas density field,* 512^3 *resolution.* ***Upper left****: on the top panel, power spectrum generated by BAM with respect to the reference's (black, solid line) as iterations advance with no cosmic web classification applied.* ***Upper right****: on the top panel, same as (a), but applying cosmic web classification.* ***Bottom left****: on the top panel, power spectrum generated by BAM with respect to the reference's for the last iteration with no cosmic web classification applied.* ***Bottom right****: on the upper panel, power spectrum generated by BAM with respect to the reference's (black, solid line) for the last iteration using cosmic web classification. The bottom panel of the four images shows the corresponding ratios with respect to the reference; the different shaded areas stand for 5 and* 10% *difference with respect to 1 on the left panel, 1 and 5% difference with respect to 1 on the right one.*



(a) *Kernel evolution, no CWC*

(b) *Kernel evolution, CWC*

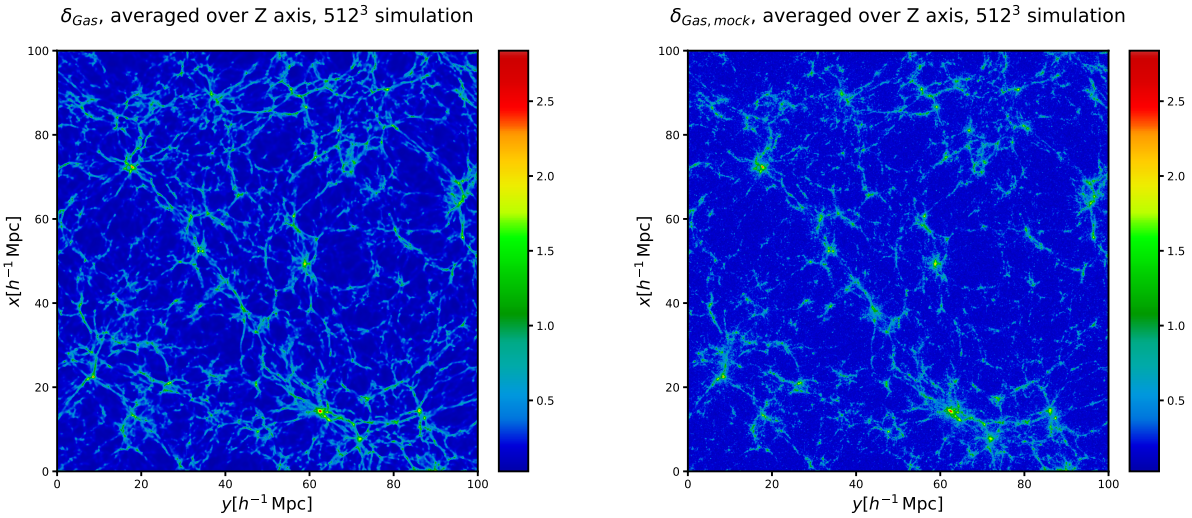
Figure 9: *Evolution of BAM's kernel as iterations advance, for the* 512^3 *gas reference. Kernel value is computed for each k mode with Eq. 22. As it can bee seen, it approaches a certain value, as the power spectrum generated by BAM approaches the reference's, and the Metropolis-Hasting algorithm ceases to accept new values, as specified in Eq. 21.*

4.1.2. High resolution case: mesh with 512^3 cells

The power spectrum of the gas density field generated by `BAM` in its calibration is shown on Figure 8. The results of the method not using the cosmic web classification should be analyzed in the first place: again, the very first iteration generated a power spectrum that is above the reference's on most of the possible modes until $k \sim 1\,h\,\mathrm{Mpc}^{-1}$, and although the difference is smaller for larger scales than for the case presented in section 4.1.1, it reaches about the same $\sim 25\%$ of deviation from the reference before the generated power spectrum takes values below the reference's, reaching a $\sim 60\%$ deviation at the smaller scales. However, after ~ 60 iterations the method seems to have converged, generating a power spectrum that is on top of the reference's. On panel (a), the evolution throughout the iterations is represented, being the last iteration shown on panel (c). As it can be seen in the lower subpanel of (c), the ratio between the generated and the reference's power spectrum is below 1% for every mode. On Figure 9, panel (a), the evolution of `BAM`'s kernel through the iterative process can be seen: as stated before, after about 60 iterations the kernel converges to a certain value.

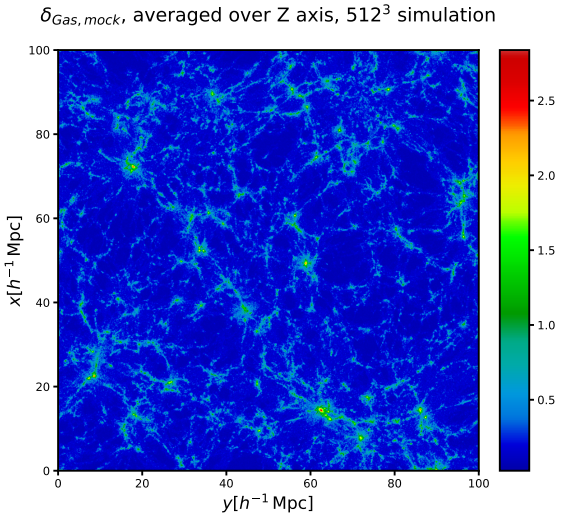
The gas density field generated by `BAM` after this calibration is presented on panel (b) of Figure 10, and can be compared with the reference's (on panel (a) of the same Figure). Both of the gas density fields have been converted to overdensity, and then to a logarithmic scale, averaging the same five adjacent slices along Z axis. The same structures are observed in both fields, but there are some differences on small scales, were structures such as filaments and sheets seem to be slightly broken, and voids seem to have been somewhat more populated than those on the reference. For this reason, and although the results are still pretty good, an analysis including the cosmic web classification into knots, sheets, filaments and voids has been carried out.

On panels (b) and (d) of Figure 8 the evolution of the power spectrum through the iterative process using the cosmic web classification is presented. As it can be appreciated, the deviation from the reference's power spectrum is slightly smaller on the first iteration when compared to that in the previous analysis, with the large scale values of k being within a 5% difference with unity, and growing *only* to $\sim 50\%$ on small scales. The iterative process rapidly solves that discrepancies,



(a) *Gas density field, reference*

(b) *Gas density field, BAM, no CWC*



(c) *Gas density field, BAM, CWC*

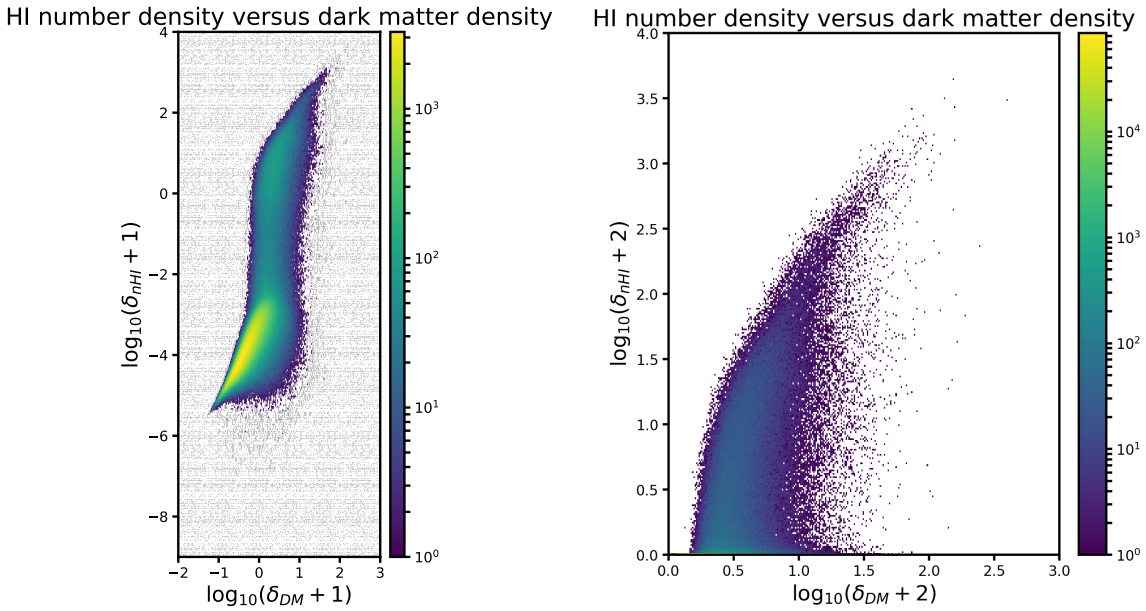
Figure 10: ***Upper left****: reference gas density field for the* 512^3 *resolution simulation.* ***Upper right****: gas density field generated by BAM for the* 512^3 *resolution simulation. For the generation of this mock, no cosmic web classification has been used.* ***Bottom****: gas density field generated by BAM for the* 512^3 *resolution simulation. For the generation of this mock, no cosmic web classification has been used.In the three cases, the original field has been converted to* δ*, and afterwards converted into a logarithmic scale using* $\log_{10}(2+\delta_{gas})$*. The Z direction of the 3D field has been averaged over the same 5 slices in both fields.*

and after only 20 iterations the small scales have been captured under the 1% deviation threshold. Between iterations 40 and 60, the power spectrum generated by `BAM` is placed on top of the reference's, not changing after it reaches its final iteration, 200 in this case. This analysis can be confirmed by taking a look at panel (b) on Figure 9, where it can be noticed that after iteration 67 the kernel does not change at all, reaching convergence. It is remarkable how, using the cosmic web classification, the shape of the kernel changes slightly for the modes associated with larger scales, assigning lower values than in the previous result.

The classification into the cosmic web type associates 0.76% of the particles to knots, 20.59% to filaments, 60.30% to sheets and 18.35% to voids. On panel (c) of Figure 10 it can be seen how, including this non-local property in the analysis, the filamentary structure of the gas density field is conserved in a more obvious way, besides the voids being somewhat more populated than those in the reference. Finally, it is necessary to notice how, both using the cosmic web classification or not using it, the structures generated by `BAM` are slightly less smooth that those for the reference, as well as the appearance of a granulated pattern on the smaller scales, but this is considering that structures on a scale of $100\,h^{-1}\,\mathrm{Mpc}/512$ are to be reproduced here.

4.2. Neutral hydrogen: HI number density field

4.2.1. Low resolution case: mesh with 128^3 cells

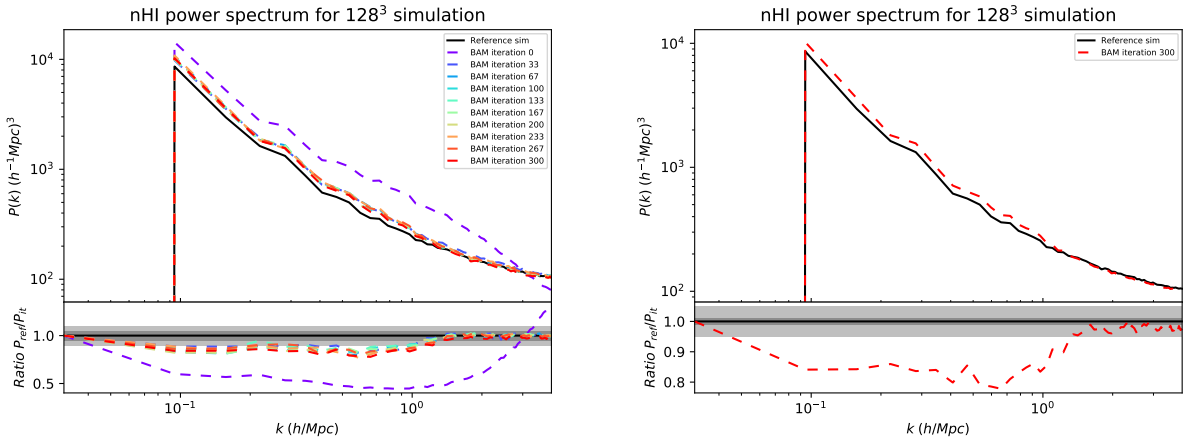


(a) *Bias relation between nHI and dark matter*

(b) *Same, but with a 2 in the log*

Figure 11: *Two-dimensional histogram relating dark matter density field with HI number density field for the* 128^3 *resolution simulation. Both fields have been converted to* δ*, and afterwards converted into a logarithmic scale, using* $\log_{10}(1+\delta)$ *in panel (a) and using* $\log_{10}(2+\delta)$ *in panel (b).*

The bias relation between the dark matter density field and the neutral hydrogen density field extracted from the reference follows a much more complex distribution than the gas', with two



(a) *Power spectrum evolution* (b) *BAM's last iteration*

Figure 12: *Every image presents results from the neutral hydrogen number density field,* 128^3 *resolution.* ***Left****: on the top panel, power spectrum generated by BAM with respect to the reference's (black, solid line) as iterations advance with no cosmic web classification applied.* ***Right****: on the top panel, same as (a), but applying cosmic web classification. The bottom panel of both images shows the corresponding ratios with respect to the reference; the different shaded areas stand for 5 and* 10% *difference with respect to 1 on the left panel, 1 and 5% difference with respect to 1 on the right one.*

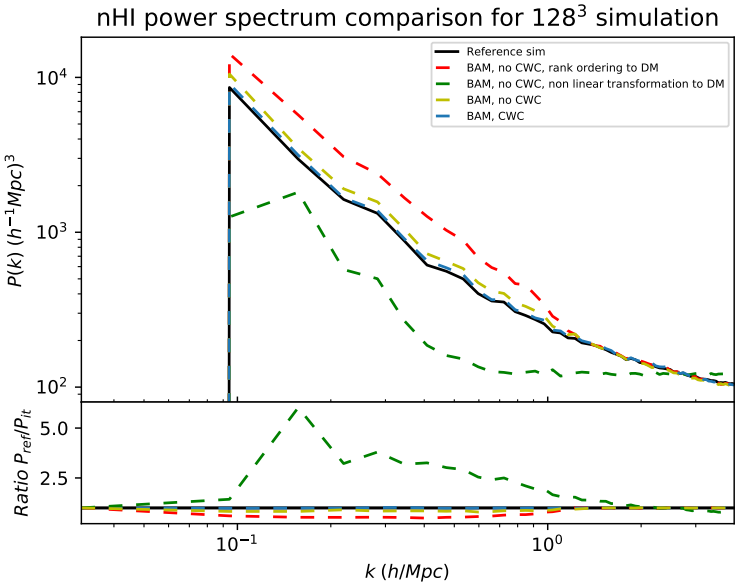
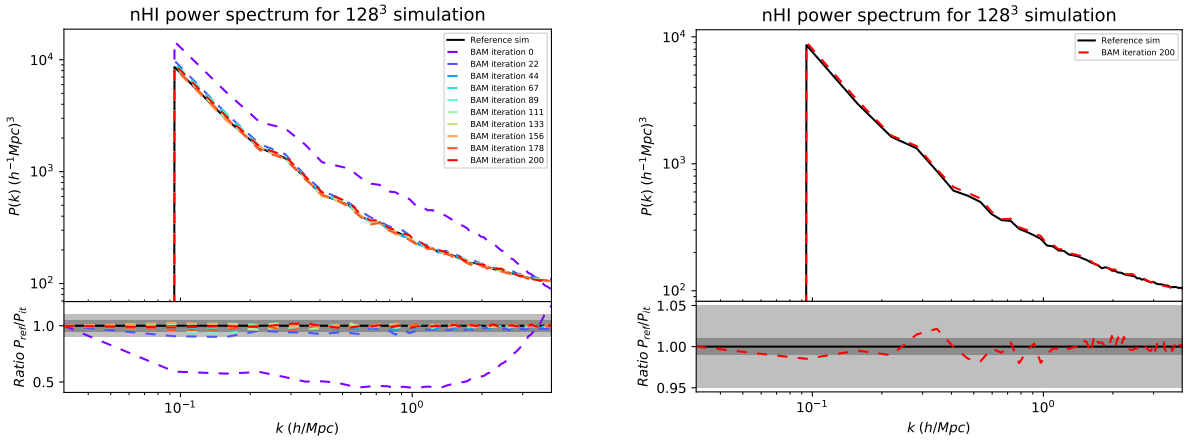


Figure 13: *BAM's last iteration power spectrum for different approaches trying to improve the performance of the one showed on Fig. 12, explained in the text. The only method that provided a good result was including the cosmic web classification and the mass of collapsing regions in the analysis (blue, dashed line). The bottom panel shows the corresponding ratios with respect to the reference; the different shaded areas stand for 5 and* 10% *difference with respect to 1.*



(a) *Power spectrum evolution*

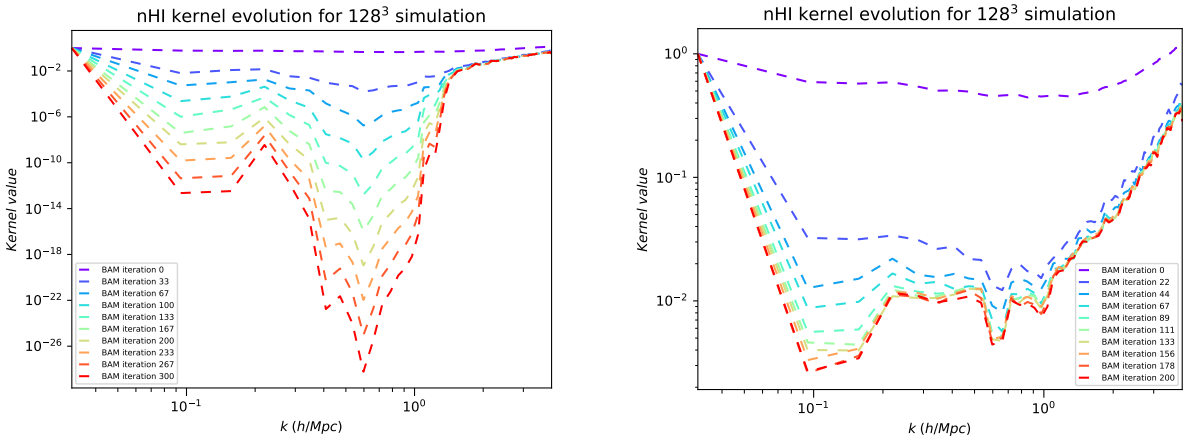
(b) *BAM's last iteration*

Figure 14: *Every image presents results from the neutral hydrogen number density field,* 128^3 *resolution, using the cosmic web classification.* ***Left****: on the top panel, power spectrum generated by BAM with respect to the reference's (black, solid line) as iterations advance.* ***Right****: on the top panel, same as (a), but applying cosmic web classification. The bottom panel of both images shows the corresponding ratios with respect to the reference; the different shaded areas stand for 5 and* 10% *difference with respect to 1 on the left panel, 1 and 5% difference with respect to 1 on the right one.*

different lobes, one of them associated to very low densities of neutral hydrogen, as can be seen in Figure 11, panel (a). The power spectrum of the neutral hydrogen number density field on its 128^3 resolution configuration has caused some troubles during the calibration of `BAM`, and the process followed in trying to solve them and improve the results will be set forth here. As a first approach, the calibration was done without considering the cosmic web classification. This dealt to a first generated power spectrum that was above the reference's on almost a 50% for every mode up to $k = 2\,h\,\mathrm{Mpc}^{-1}$, much more than when the method was applied with this configuration to the gas density field in any of its setups (see section 4.1); after that mode, the power spectra generated by `BAM` followed a similar trend as those calculated for the gas, turning its value below the reference's. As opposed with what happened in the previous calibration processes, the iterative scheme did not achieved a satisfactory result, as can be seen in Figure 12: the power spectrum that were being generated seemed to converge to a certain shape, but it systematically failed to be settled on top of the reference's, as it can be seen in panel (a). After even 300 iterations, the deviation from the reference on most of the modes was around 20%, although for small scales the power spectrum of the reference and the obtained from the calibration differed on less than a 5%, as it can be seen on panel (b). As a confirmation of the non convergence of the method, one can notice that the kernel was not converging for any value of k except for those greater than $k \approx 1\,h\,\mathrm{Mpc}^{-1}$, that is, the same smaller scales that seemed to reach a better sampling on the power spectrum; the kernel did not stop decreasing its value elsewhere as iterations go on, as can be seen on panel (a) of Figure 15.

To solve the problem, and as it seemed that it caused that the power spectra were to high on large scales, an attempt was made to substract some power from that scales, applying transformations to the dark matter density field that dealt on a reduction of its power. The first test consisted on applying a non-linear transformation to the dark matter density field, as follows:

1. $\delta_{DM} = (1+\delta_{DM})^{\alpha}$, $0 < \alpha < 1$



(a) *Kernel evolution, no CWC*

(b) *Kernel evolution, CWC*

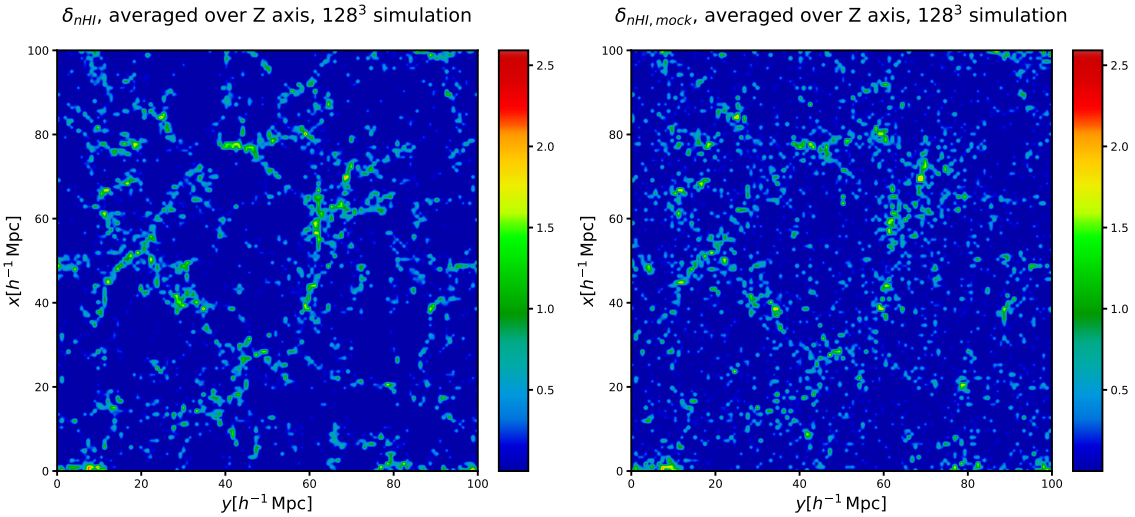
Figure 15: *Evolution of BAM's kernel as iterations advance, for the* 128^3 *neutral hydrogen number density reference. Kernel value is computed for each k mode with Eq. 22. As it can bee seen, it approaches a certain value, as the power spectrum generated by BAM approaches the reference's, and the Metropolis-Hasting algorithm ceases to accept new values, as specified in Eq. 21.*

2. Get overdensity of the new δ_{DM}
3. $\delta_{DM} = \exp(\beta\, \delta_{DM})$
4. Get overdensity of the new δ_{DM}
5. $\delta_{DM} = (1 + \delta_{DM})^{\gamma}$, $\gamma > 1$

The results obtained from this method are presented on a dashed, green line on Figure 13. As it can be seen, not only it did not solve the problem, but it got it worse: the ratio between the reference and the generated power spectrum grows above 5 for some modes, and in smaller scales a white noise pattern began to appear, as the kernel made the dark matter density field grow uncontrollably, dealing to the fact that the points of the mesh that already had high values grew above the limits that were defined. Different values for α, β and γ were tried, but none of them turned out to be satisfactory.

As a second test, a rank ordering was applied to the probability distribution function of the dark matter density field, so that it resembles that of the neutral hydrogen number density field. The results obtained are represented on a dashed, red line on Figure 13. Although the problem occurring with the previous method is avoided, the power spectrum generated is worse than the original method, being above the reference's until $k \approx 1\, h\mathrm{Mpc}^{-1}$, and having greater values than the first attempt for every mode before that one.

Finally, the two non-local properties of the dark matter that are included in `BAM` were taken into consideration: the cosmic web classification and the mass of the collapsing regions. With this approach, the results are much more convincing, as can be seen on the dashed, blue line of Figure 13, and will be analyzed separately. On Figure 14, panel (a), the evolution of the power spectra generated by `BAM` can be seen: for the first iteration, the difference with the reference goes up to $\sim 50\%$, reaching a deviation of less than a $\sim 10\%$ in 20 iterations. The main difference between this configuration and the previous one with no cosmic web classification is that the method seem



(a) *Neutral hydrogen number density field, reference*

(b) *Neutral hydrogen number density field, BAM, CWC*

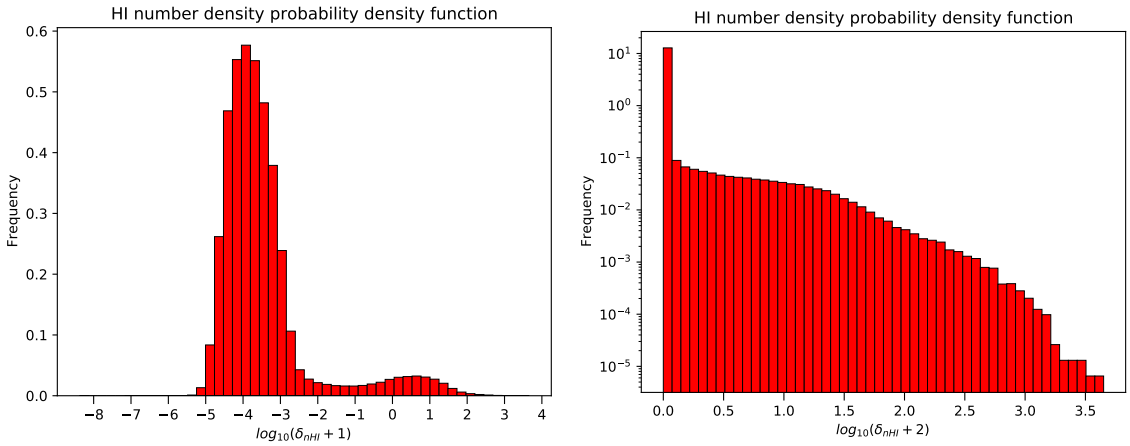
Figure 16: ***Left**: reference neutral hydrogen number density field for the* 128^3 *resolution simulation.* ***Right**: neutral hydrogen number density field generated by BAM for the* 128^3 *resolution simulation. For the generation of this mock cosmic web classification has been used. In both cases, the original field has been converted to* δ*, and afterwards converted into a logarithmic scale using* $\log_{10}(2+\delta_{nHI})$*. The Z direction of the 3D field has been averaged over the same 5 slices in both fields.*

to converge and place its power spectrum on top of the reference at about iteration 60, reaching deviations of around 1% for almost every mode, as can be seen on the lower subpanel in panel (b) of the figure. The classification into the cosmic web type associates 0.024% of the particles to knots, 0.058% to filaments, 0.293% to sheets and 99.625% to voids.

To confirm this analysis, one can take a look at Figure 15. As it has been said, the kernel generated by `BAM` when the cosmic web is not taken into account does not converge to any value except for the modes corresponding to the smaller scales, as panel (a) shows. On the contrary, as shown in panel (b), the kernel converges to a certain value, although the speed with which it occurs is much slower than those that took place for the gas density field, as expounded on section 4.1.

Finally, on Figure 16 the neutral hydrogen number density field of the reference (on panel (a) of the figure) can be compared with the density field generated by `BAM` after the calibration has ended, with cosmic web classification and mass of the collapsing regions taken into account in the analysis (on panel (b) of the figure). Although at a large scale structures seem to be conserved, it is pretty obvious that filamentary structures get broken and that mock voids have some structures that in the reference do not exist.

The solution to this might come from considering a more efficient binning after applying Eq. 25 to the density field. As it can be seen in Figure 17, panel (a), the probability distribution function of the reference density field has two distinguishable lobes, one around $log_{10}(\delta_{nHI}+1)=-4$ —which will be associated to regions with a very low density of neutral hydrogen, almost null— and the other around $log_{10}(\delta_{nHI}+1)=1$ —which corresponds to regions where there is actually high densities of neutral hydrogen—. This leads to the fact that most of the field is *empty* and there are only some points where all the neutral hydrogen is located. When dealing with this fact inside `BAM`, where it is



(a) *HI number density PDF, 1 in the log*

(b) *HI number density PDF, 2 in the log*

Figure 17: ***Left**: probability distribution function of the neutral hydrogen number density for the* 128^3 *resolution simulation, adding 1 inside the logarithm when transforming the overdensity into a logarithmic scale.* ***Right**: probability distribution function of the neutral hydrogen number density for the* 128^3 *resolution simulation, adding 2 inside the logarithm when transforming the overdensity into a logarithmic scale.*

necessary to use Eq. 25 (adding a 2 inside the logarithm when transforming the overdensity into a logarithmic scale), the probability distribution function turns into the one in panel (b) of the same figure: the great majority of the points of the field, up to a factor 100, are contained in the very first bin of the distribution. These issue can also be seen in panel (b) of Figure 11. What this means is than, when using a linear binning between the minimum and the maximum of that distribution, most of points (actually, about a 95%) will fall in the first bin, being the density field produced by the method, therefore, poorly sampled for most of the possible values of the reference density field.

4.3. Temperature field

4.3.1. Low resolution case: mesh with 128^3 cells

The bias relation between the dark matter density field and the temperature field extracted from the reference also follows a complex relation, with two different populations associated to the same dark matter density, as it is shown at Figure 18. The power spectrum of the temperature field on its 128^3 resolution configuration has caused major troubles during the calibration of `BAM`, as the method did not converge at all with the previous configurations. As it is shown on Figure 19, panels (a), the power spectra were always below the reference one and the code stopped its calibration process only after 6 iterations, causing a core dump. This can be explained by panel (b) on that same figure: the kernel is growing above 1 in each iteration on an attempt to provide more power to the spectrum for almost every k, and this leads to an uncontrolled growth of the density of the dark matter density field that is being mapped, causing that `BAM` has to abort its calibration. However, applying a different binning for the bias extraction (that was originally conceived to solve the problem mentioned on the neutral hydrogen distribution, but did not work well) seems to solve the problem, at least for every mode up to $k \sim 2h\mathrm{Mpc}^{-1}$. This specific binning is defined as follows: taking the first $\sim 95\%$ smaller values of the field, define a lower threshold as their minimum, and an upper threshold as their maximum —note that those first values where gathered together into the

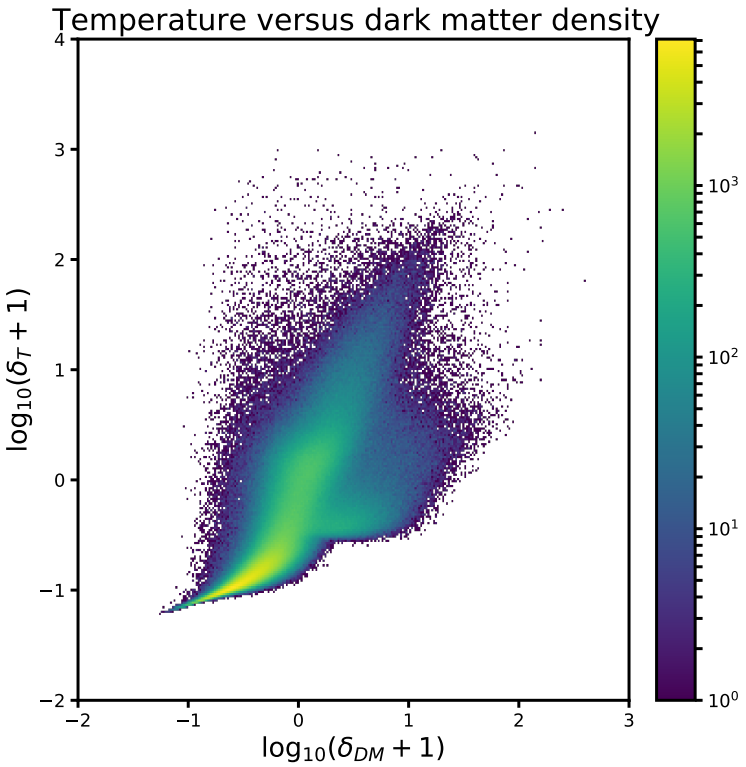
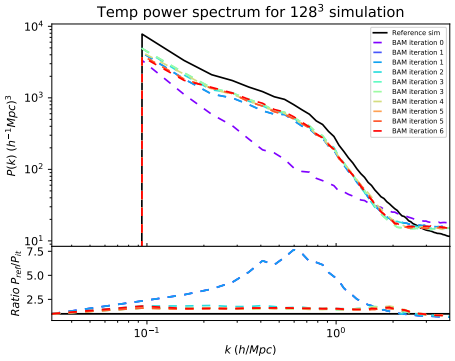


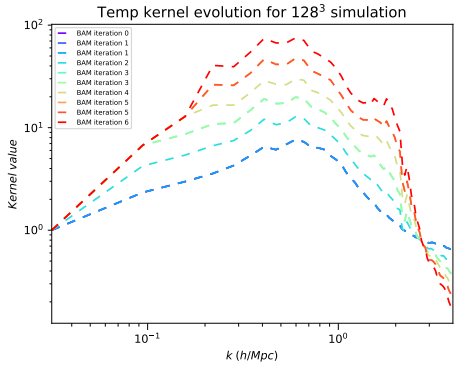
Figure 18: *Two-dimensional histogram relating dark matter density field with temperature field for the* 128^3 *resolution simulation. Both fields have been converted to* δ*, and afterwards converted into a logarithmic scale using* $\log_{10}(1+\delta)$*.*

first bin with the basic configuration. Force that half of the bins that are to be used will be placed there, and fill each one of them with a fixed number of cells; the limits of each bin are defined, then, by the lower and higher value of the first and last cell placed in the bin. The second half of the bins will be defined as they were previously, but only contain a rough $\sim 5\%$ of the cells. This leads to an irregular distribution of the binning, where the first half of them have a different width for each bin, but the same amount of cells in everyone, and the second half of them has the same width for each bin, but with a different amount of cells in each of them.

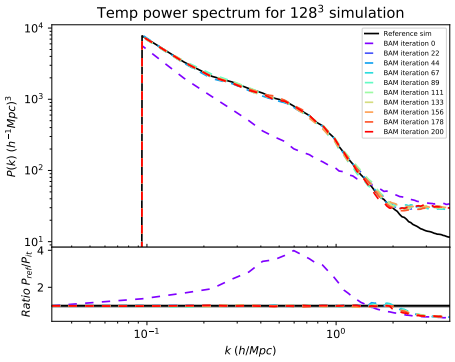
As it is shown on panel (c) of Figure 19, the power spectra generated by `BAM` are on top of the reference's for every mode up to $k \sim 2\,h^{-1}$ Mpc, being above the reference for smaller scales; on panel (e) it can be seen how the deviation between the reference and the power spectrum generated by the method is well below 5% at those scales. On panel (d), the kernel using this binning is presented, showing a clear convergence at the mentioned scales, and a clear problem on smaller scales, where it does not reach any stable value and keeps decreasing. In Figure 20 the temperature field is represented, both the reference's and the generated by `BAM` after its calibration, with this bin configuration. As it can be seen, the granulation appears on the mock field at Mpc scales, but at large scales the picture is maintained between both fields.



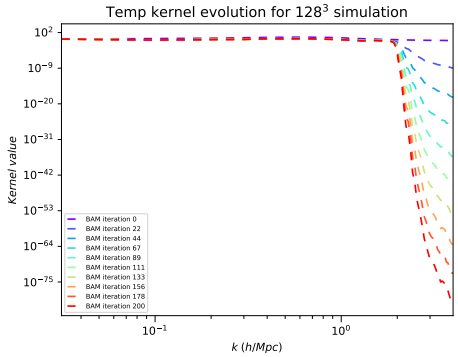
(a) *Power spectrum evolution*



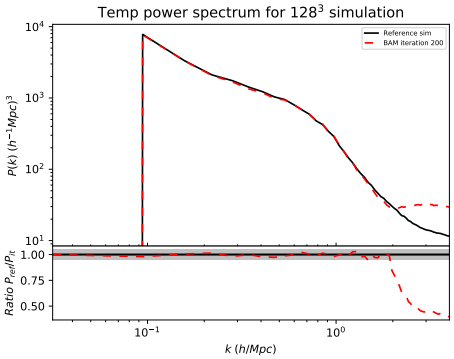
(b) *Kernel evolution*



(c) *Power spectrum evolution, different binning*

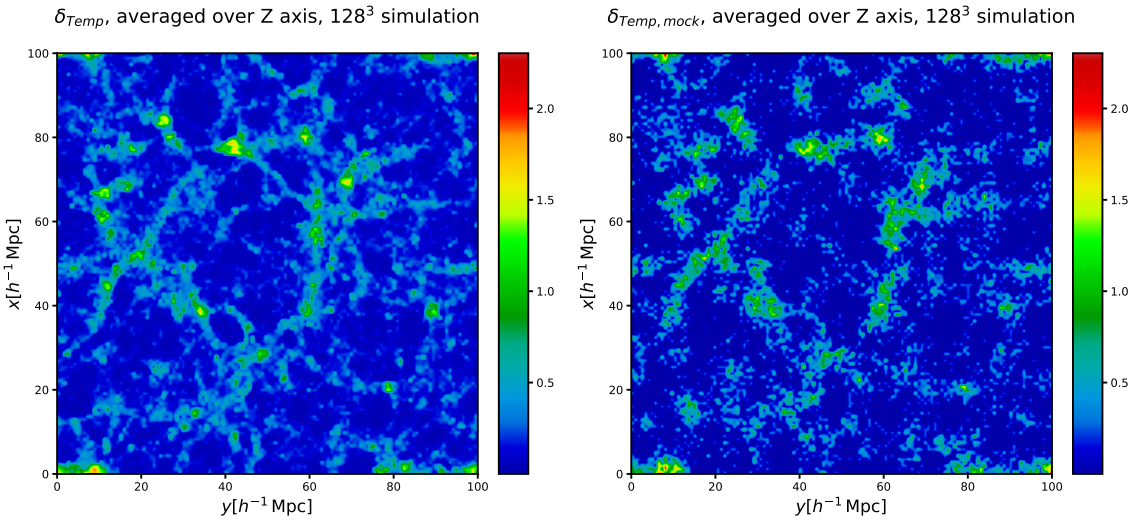


(d) *Kernel evolution, different binning*



(e) *Power spectrum last iteration, different binning*

Figure 19: ***Panel (a)****: power spectrum generated by BAM with respect to the reference's (black, solid line) as iterations advance, applying cosmic web classification, for the* 128^3 *resolution.* ***Panel (b)****: Evolution of BAM's kernel as iterations advance.* ***Panel (c)****: power spectrum generated by BAM with respect to the reference's as iterations advance, applying cosmic web classification and the binning described in the text, for the* 128^3 *resolution.* ***Panel (d)****: Evolution of BAM's kernel as iterations advance, using the binning described in the text.* ***Panel (e)****: Power spectrum of the last iteration of BAM, applying cosmic web classification and the binning described in the text.*



(a) *Temperature field, reference* (b) *Temperature field, BAM, different binning*

Figure 20: ***Left**: reference temperature field for the* 128^3 *resolution simulation.* ***Right**: temperature field generated by BAM for the* 128^3 *resolution simulation. For the generation of this mock, cosmic web classification has been used, along with the binning defined in the text. In both cases, the original field has been converted to* δ*, and afterwards converted into a logarithmic scale using* $\log_{10}(2+\delta_{Temp})$*. The Z direction of the 3D field has been averaged over the same 5 slices in both fields.*

5. Summary and conclusions

In this work a mapping technique which allows to paint the dark matter field defined on a cubical mesh with baryonic properties obtained from detailed hydrodynamical simulations has been presented. This can be of major importance in cosmological studies which require hydrodynamical baryonic physics.

The distribution of baryonic matter in the Universe has a wealth of cosmological and astrophysical information. It results from structure formation dominated by dark matter in an expanding background Universe within the interplay of a series of complex physical processes. State-of-the-art hydrodynamical simulations include the majority of the involved physics to produce a fair description of the baryonic matter in the Universe. However, they are extremely expensive to run, and thereby neither the cosmic volumes, which will be provided by forthcoming astronomical observational campaigns, nor the required number of simulations to perform robust statistical analysis are available. This calls for the development of techniques which permit us to learn extract the information from such expensive calculations and map them onto a far less expensive dark matter only simulation. The present master thesis explores this possibility by focusing on a series of baryonic quantities such as the ionised gas density, and the neutral hydrogen.

In particular, the ability of the Bias Assignment Method (BAM), described in Balaguera-Antolínez et al. (2019, 2020), to surpass its original idea of mapping a halo distribution onto a dark matter density field using a high resolution N-body simulation as a reference and produce mock catalogues out of it, has been assessed. To that aim, the code has had to be adapted from its previous configuration, which computed the number counts of a discrete density field distribution, to a new one, which is able to compute continuous fields that stand for baryonic matter distributions, such as gas

densities, neutral hydrogen densities, temperatures of gas or optical depths. A high resolution hydrodynamical simulation produced with the GAGDET3-OSAKA code (Aoyama et al., 2018; Shimizu et al., 2019), whose basic features have been discussed throughout this text, has been used to fulfill that purpose, taking advantage of the provided 128^3, 256^3 and 512^3 resolution dark matter, gas and neutral hydrogen density fields.

The approach of the method is based on the extraction of the non-linear and non-local bias relations from a N-body simulation (extended to a cosmological SPH simulation here) using an explicit, parameter-free model. Among the non-local quantities, `BAM` currently takes into account the cosmic web classification, based on the eigenvalues of the tidal field, and the mass of the collapsing regions (namely, the knots). The method is not able to mirror the power spectrum from the references in its first measure, as a set of additional non-local relations are not included in the analysis, such as the missing power of Lagrangian perturbation theory based methods (see Munari et al., 2017) or third order contributions (see McDonald & Roy, 2009). The treatment given, for instance, to the computation of the eigenvalues, could also lead to some aliasing effects on small scales, even only considering the second order bias description which `BAM` uses. However, the iterative process used to compute the bias relation and compute a kernel (following Equations (17) and (22) respectively) handles with all the unknown dependencies, leading to the generation of power spectra in, roughly, a $\sim 1\%$ agreement with the reference's, for every density field analyzed, showing how the effects that are not modelled are absorbed in the calibration process of the kernel.

This process can be interpreted as a machine learning method where the power spectrum represents the *cost function*. However, deep learning techniques usually require large number of samples as a training set, and for purposes like the one presented here that represents a problem, as N-body simulations are highly demanding in terms of computation time and power, even restricting to volumes smaller than the one needed to cover the current galaxy surveys, as the one used in this work. In order to loosen this need, more physical and statistical relations between the tracer distribution and the dark matter have been included into the method. Also, limiting the cost function to second order statistics functions such as the power spectrum could be problematic, as the bias relation is degenerated at the three-point statistics level (Kitaura et al., 2015; Vakili et al., 2017). However, as `BAM` completely extracts the bias relation from the reference, it is acceptable to use second order statistics.

The method, as has been shown, is able to reproduce the two-point statistics and main structures present in a cosmological volume of $100\,h^{-1}$ Mpc for the gas density field in both a resolution of 128^3 and 512^3 particles, extracting the non-linear and non-local information from the reference simulation, with the unknown dependencies contained in a kernel. In terms of three points stadistics, it is very likely that these two cases would reproduce the three-point statistics on a fairly good manner. However, the results were not so good when dealing with the neutral hydrogen number density field, and although the two-point statistics where reproduced with deviations of only a $\sim 1\%$, it is clear that three-point statistics would not bring good results. To solve this problem, it has been shown that a more efficient binning might be necessary, as most of the particles of the density field are contained in the first bin of the distribution with the current configuration. Along this, another problem will need to be faced when dealing with the temperature and optical depths fields. `BAM` is not able to reproduce the power spectrum from the reference, as can be seen in Figure 19, panel (a), as the power spectrum generated by the method is always below the reference's, and the kernel grows above 1 (panel (b)), leading to the uncontrolled increase on the values of the density fields for almost every scales except the small ones. On an aim to solve these problems, some different transformations and binnings are being tried to alleviate the loss of power at certain scales, and new analysis are being done with the dark matter field obtained with an NGP scheme, instead a CIC one. Using a new approach based on the tidal tensor invariants (Kitaura et al., 2020), instead of some restricted combinations of its eigenvalues, as it is usually done in cosmic web classification

studies, could also improve the performance of the method.

In summary, it has been shown that the BAM method can be extended to be applied to map baryonic physics onto a dark matter density field, extracting the relation between that field and a continuous one measuring different physical properties from a reference simulation and taking advantage of some non-local properties of the dark matter field such as the eigenvalues of the tidal tensor or the mass of the collapsed regions. Some problems have arisen, though, but there is an ongoing, promising effort on solving them, and produce precise and accurate effective hydrodynamical simulations on cosmological scales in large numbers. This will enable us to analyse upcoming surveys mapping the inter-galactic medium.

6. Acknowledgements

I would like to thank Francisco Kitaura for his encouragement and guiding in some difficult moments during my master thesis. I want to especially thank Andrés Balaguera-Antolínez for sharing and instructing me in the usage of the BAM code, and for his help and time dedication throughout every different stage of this work. Also special thanks to Kentaro Nagamine for providing me his hydrodynamical simulations, which made this work possible.

References

Abel T., Hahn O., Kaehler R., 2012, MNRAS, 427, 61

Amendola L., et al., 2018, Living Reviews in Relativity, 21, 2

Aoyama S., Hou K.-C., Hirashita H., Nagamine K., Shimizu I., 2018, MNRAS, 478, 4905

Avila S., Murray S. G., Knebe A., Power C., Robotham A. S. G., Garcia-Bellido J., 2015, MNRAS, 450, 1856

Balaguera-Antolínez A., Kitaura F.-S., Pellejero-Ibáñez M., Zhao C., Abel T., 2019, MNRAS, 483, L58

Balaguera-Antolínez A., et al., 2020, MNRAS, 491, 2565

Bautista J. E., et al., 2017, Astr.Astrophy., 603, A12

Benitez N., et al., 2014, arXiv e-prints, p. arXiv:1403.5237

Blot L., et al., 2019, MNRAS, 485, 2806

Bouwens R. J., et al., 2009, ApJ, 705, 936

Chabrier G., 2003, Pub.Astr.Soc.Pac., 115, 763

Chevalier R. A., 1974, ApJ, 188, 501

Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, MNRAS, 446, 2621

Colavincenzo M., et al., 2019, MNRAS, 482, 4883

Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140

Dawson K. S., et al., 2016, AJ, 151, 44

Dekel A., Lahav O., 1999, ApJ, 520, 24

Dodelson S., Schneider M. D., 2013, Phys.Rev.D, 88, 063537

Eisenstein D. J., Hu W., 1998, ApJ, 496, 605

Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23

Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, MNRAS, 463, 2273

Haardt F., Madau P., 2012, ApJ, 746, 125

Hahn O., Abel T., 2011, MNRAS, 415, 2101

Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489

Hahn O., Abel T., Kaehler R., 2013, MNRAS, 434, 1171

Hirashita H., 2015, MNRAS, 447, 2937

Hopkins P. F., Quataert E., Murray N., 2011, MNRAS, 417, 950

Hopkins P. F., Narayanan D., Murray N., Quataert E., 2013, MNRAS, 433, 69

Hou K.-C., Hirashita H., Nagamine K., Aoyama S., Shimizu I., 2017, MNRAS, 469, 870

Howlett C., Percival W. J., 2017, MNRAS, 472, 4935

Kaiser N., 1987, MNRAS, 227, 1

Keller B. W., Wadsley J., Benincasa S. M., Couchman H. M. P., 2014, MNRAS, 442, 3013

Keller B. W., Wadsley J., Couchman H. M. P., 2015, MNRAS, 453, 3499

Keller B. W., Wadsley J., Couchman H. M. P., 2016, MNRAS, 463, 1431

Kim J.-h., et al., 2016, ApJ, 833, 202

Kitaura F. S., Hess S., 2013, MNRAS, 435, L78

Kitaura F. S., Yepes G., Prada F., 2014, MNRAS, 439, L21

Kitaura F.-S., Gil-Marín H., Scóccola C. G., Chuang C.-H., Müller V., Yepes G., Prada F., 2015, MNRAS, 450, 1836

Kitaura F.-S., et al., 2016, MNRAS, 456, 4156

Kitaura F.-S., Balaguera-Antolínez A., Sinigaglia F., Pellejero-Ibáñez M., 2020, arXiv e-prints, p. arXiv:2005.11598

Larson R. B., 1981, MNRAS, 194, 809

Levi M., et al., 2013, arXiv e-prints, p. arXiv:1308.0847

Lippich M., et al., 2019, MNRAS, 482, 1786

McDonald P., Roy A., 2009, J.Cosm.Astr.Phys., 2009, 020

McKee C. F., Ostriker J. P., 1977, ApJ, 218, 148

Monaghan J. J., 1997, Journal of Computational Physics, 136, 298

Morris J. P., 1996, PASA, 13, 97

Moustakas J., et al., 2013, ApJ, 767, 50

Munari E., Monaco P., Koda J., Kitaura F.-S., Sefusatti E., Borgani S., 2017, J.Cosm.Astr.Phys., 2017, 050

O'Connell R., Eisenstein D. J., 2019, MNRAS, 487, 2701

Paranjape A., Hahn O., Sheth R. K., 2018, MNRAS, 476, 3631

Paz D. J., Sánchez A. G., 2015, MNRAS, 454, 4326

Pearson D. W., Samushia L., 2016, MNRAS, 457, 993

Pellejero-Ibañez M., et al., 2020, MNRAS, 493, 586

Percival W. J., et al., 2014, MNRAS, 439, 2531

Planck Collaboration et al., 2016, Astr.Astrophy., 594, A13

Reddy N. A., Steidel C. C., 2009, ApJ, 692, 778

Saitoh T. R., 2016, CELib: Software library for simulations of chemical evolution (ascl:1612.016)

Saitoh T. R., 2017, AJ, 153, 85

Saitoh T. R., Makino J., 2009, ApJ, 697, L99

Saitoh T. R., Makino J., 2013, ApJ, 768, 44

Schiminovich D., et al., 2005, ApJ, 619, L47

Scoccimarro R., Sheth R. K., 2002, MNRAS, 329, 629

Sedov L. I., 1959, Similarity and Dimensional Methods in Mechanics

Shimizu I., Todoroki K., Yajima H., Nagamine K., 2019, MNRAS, 484, 2632

Sigad Y., Branchini E., Dekel A., 2000, ApJ, 540, 62

Smith B. D., et al., 2017, MNRAS, 466, 2217

Sousbie T., Courtois H., Bryan G., Devriendt J., 2008, ApJ, 678, 569

Springel V., 2005, MNRAS, 364, 1105

Springel V., Hernquist L., 2002, MNRAS, 333, 649

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Stinson G. S., Brook C., Macciò A. V., Wadsley J., Quinn T. R., Couchman H. M. P., 2013, MNRAS, 428, 129

Sutherland R. S., Dopita M. A., 1993, Rev.Astr.Astrphys., 88, 253

Taylor G., 1950, Proceedings of the Royal Society of London Series A, 201, 159

Taylor A., Joachimi B., Kitching T., 2013, MNRAS, 432, 1928

The Dark Energy Survey Collaboration 2005, arXiv e-prints, pp astro–ph/0510346

Tomczak A. R., et al., 2014, ApJ, 783, 85

Vakili M., Kitaura F.-S., Feng Y., Yepes G., Zhao C., Chuang C.-H., Hahn C., 2017, MNRAS, 472, 4144

Weinberg D. H., 1992, MNRAS, 254, 315

White M., Tinker J. L., McBride C. K., 2014, MNRAS, 437, 2594

Zhao C., Kitaura F.-S., Chuang C.-H., Prada F., Yepes G., Tao C., 2015, MNRAS, 451, 4266