Exploring the Epoch of Reionization with Line-Intensity Mapping

DISSERTATION
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Benedetta Spina
aus
Conegliano – Italien

Bonn, 2023

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Cristiano Porciani
2. Gutachterin:	Prof. Dr. Andrina Nicola

Tag der Promotion:	22.09.2023
Erscheinungsjahr:	2023

To my mom and my dad,
to Francesco, Lorenzo and Margherita,
and to you, for staying with me until the very end

Abstract

The epoch of reionization (EoR) stands as a key phase in the history of the Universe, where the neutral elements that pervaded the cosmos gradually became ionized. This process was driven by the photons emitted by the the first luminous sources and it strongly depends on their formation and evolution. Understanding the progression of the EoR means constraining the growth of structures and the interplay between galaxies and the intergalactic medium. In this work, we shed light on several aspects of this phenomenon, in order to get an insight of the general picture.

This thesis is also motivated by the large amount of on-going and up-coming experiments that will finally unveil the core of the EoR. In particular, we make use of the line-intensity mapping (LIM) technique which allows us to study the large-scale structure of the Universe through the detection of the hyperfine transition in neutral hydrogen (HI) and single-ionized helium (HeII) atoms.

The first project focuses on the connection between HI and the dark-matter counterpart within haloes, the HI-halo mass relation (HIHMR). We simulate mock HI intensity maps parametrizing the HIHMR and implementing systematic observational effects. We analyze the geometry and topology of the mock HI maps with the Minkowski functionals and we explore the dependence of the HIHMR on halo mass and galaxy properties. The results demonstrate the potential of LIM experiments, such as the Square Kilometre Array, to constrain the HIHMR.

We then proceed to study helium reionization, an epoch for which we currently lack strong constraints regarding its drivers and endpoint. We investigate the imprint of different reionization models (early, driven by quasars at high redshift $z > 5$, and late, where active-galactic nuclei contribute from $z = 5$) on the power spectrum of HeII mock intensity maps, forecasting for radio surveys. We find that constraints on the timeline of the EoR and therefore on the ionizing sources can be achieved with such instruments.

At last, we address the quenching of dwarf galaxies by investigating the impact of the reionization front on their star-formation rates (SFRs). Observations of the local Universe suggest different behaviours for galaxies with same features (e.g. mass, age). We assess the significance of galaxies' main characteristics using the THESAN simulations and we parametrize the probability of quenching with a simple analytical form. Our findings reveal how the ionizing radiation from early sources affects the gas reservoirs and the SFRs in these systems. We provide a valuable tool to connect the observations with the theoretical framework.

With this thesis, we have explored several sides of the EoR. We have investigated the HI and the HeII evolution, the sources of reionization itself and its impact on galaxy evolution and we have provided forecasts for next generation instruments. We have paid attention to consider both sides of the phenomenon we study, observations and theory. Our findings refine the theoretical models on reionization processes and guide the interpretation of future observations, contributing to our understanding of the EoR and paving the way for further advancements.

Acknowledgements

A journey sometimes is more about the people one meets along the way, rather than the destination itself. And if this work has finally seen an end, it is because of the many people who contributed to it.

I owe my thanks to Cristiano, for supervising this work during the past (almost) four years, for the scientific discussions and for the many lessons I have learned working together. I am very grateful to all the committee members, Professors Andrina Nicola, Thomas Reiprich and Daniel Huybrechts for the time devoted reading this thesis. I especially want to thank Enrico, for all the scientific (and not) discussions, for supervising one of these projects and for always being there when I needed it; this thesis would not have been completed without his constant support. Even if I am not baking many cakes these days, I want to thank Carlo for making me passionate about science and radio-astronomy, for treating me always as his equal. And finally, this journey would not have started without the support of Sabino Matarrese, to whom I can not express my gratitude enough.

I want to thank my parents, Francesca e Diego, for always supporting me during these years of studying and working far from home; for the questions they never asked and for the ones they asked too often (ma quando finisci? stai scrivendo? e la casa nuova?); for the love, the care, the patient and for teaching me to never give up and the importance of always keeping some cash with me. I want to thank my nonna as well, her polpette made all problems more bearable and I deeply regret not having learned how to crochet together. During these years, I had to joy and the pleasure to see my siblings growing and becoming amazing people. I hope Francesco will keep calling me when in trouble; I can not wait to see what there will be next for Margherita; and I should stop being surprised when Lorenzo does something nice for me. Vi voglio bene, and I would not be here without you. And as I have defended this thesis, I am ready to fight for all of you, no matter what.

I acknowledge the support and care from my aunt Gloria, my uncle Domenico and my cousins Chiara, Davide and Matteo, during these years apart.

Apart from my family, there is one person who has always been by my side, from the first Chemistry lecture, through moving in different Countries, up to now. Anna, my Turk, my person, and my best friend. Thank you for making me see the light in the darkest days, sharing this journey with you made me feel all the weights lighter, all the problems smaller, all the knitting patterns easier. And along with Anna, Norma, the prettiest and strongest walking cow I was lucky to find.

Many things have changed in my life since I started this Ph.D. and the pandemic played a huge part in it. Being alone in a new Country was not easy, but I was lucky enough to meet some friends along the way. Diana and Luca, I miss already our chats, our gourmet dinners and the tons of pizza we ate together. Elena and Claudia, our coffee breaks gave me the strength to face many difficult mornings and taught me to appreciate every shared moment. Crossing my path with yours was as unexpected as great. My hypochondria would have killed me long ago if it was not for Gere, the kindest-hearted person I have the pleasure to know. Being your friend is proof that I have done something right in life. And I would not have passed Covid-times without the support of Shashwata, thank you for introducing me to poetry.

Dato, Mandar, Prachi, Yousry, Yvonne (along with the plushes) I enjoyed our time as PhD mates and I hope you will think about me the next time you start gossiping about something. I want to thank all the group members, in particular Emilio, Alex, Matteo, for the shared laugh and scientific discussion. A special mention to Giovanna, for making a polenta that tasted like home, and to Yulia, for the patient and the care, for helping with the complicated German bureaucracy.

Growing sometimes means apart, sometimes closer, and I am glad the distance did not affect the relationship with my Italian friends. In particular Noah, I wish I had more time for our phone calls, and I can not express how grateful I am you have always been by my side. And the same for Federica, I am glad our friendship grew with us and I want to thank you for being there when I needed, no questions asked. Care can be expressed in many forms, and sharing cute dogs photos is definitely one of them, so thank you Davide for sharing the good and the bad moments of the PhD with me; thank you for giving me the strength to keep going when the darkness was winning over. And so, Chiara, despite the ups and downs, our friendship means the world to me, and you as well. I was not even 10 y.o. when I first realized I wanted to be a scientist, as Andrea knew she wanted to be a doctor; thank you for the many medical advices, for not letting time changing our friendship, for always understanding and never judging.

Weird enough, it seems like there is life outside academica. And for this realization I need to thank Cristina, for opening my eyes to the world and for the many hours spent talking about patriarchy, feminism and most importantly, bouldering. I will learn Spanish, sooner or later, I promise, as I promise I will prepare you a risotto any time you will crave for one. I want to thank Marcello as well, I know I can count on you when the zombie apocalypse happens.

I want to deeply thank all the people I met at the AIfA in these years: Sven, Cosima, Christos, Lucas, Sabine and Ellen. And a mention to my therapist as well, I literally would not be here if it was not for him.

Finally, for all the amazing dishes cooked together, for the never-ending support, love and care, for the many hours spent driving through Europe and the Balkans, for introducing me to Genova (very beautiful but not as much as Venezia), my last thanks is for Maurizio (and Gocciole). Perchè viaggiamo insieme, in direzione ostinata e contraria.

I sincerely hope I did not forget anyone in these pages. But if I did, please forgive me, I am just very tired.

List of Publications

Relevant First-Author Publications

The following list indicates the relevant first-author publications and in preparation projects that are discussed and presented in this thesis. Along with one peer-reviewed and published research publication, two more manuscripts are in preparation for submission to the Monthly Notices of the Royal Astronomical Society.

1. **Spina B.**, Porciani C. and Schimd C. (2021), *The HI-halo mass relation at redshift $z \simeq 1$ from the Minkowski functionals of 21-cm intensity maps*, Monthly Notices of the Royal Astronomical Society, 505, 3492, https://doi.org/10.1093/mnras/stab1555

   Author's contribution: B.S. simulated the mock 21 cm intensity maps, performed the statistical analysis by the means of the Minkowski Functionals and forecasted with a Fisher matrix formalism, under the supervision of C.P. and C.S. The co-authors helped to interpret the results and provided comments and edits to the final manuscript.

2. **Spina B.** and Porciani C. (2023), *When HeII goes to HeIII: detecting signature of Helium reionization through line-intensity mapping*, in prep.

   Author's contribution: B.S. simulated the mock 3.46 cm intensity maps, performed the statistical analysis by the means of the power spectrum and forecasted for future radio experiments, under the supervision of C.P. C.P. helped to interpret the results and provided comments on the manuscript's draft.

3. **Spina B.**, Garaldi E. and Porciani C. (2023), *How to identify galaxies quenched by reionizaton*, in prep.

Author's contribution: B.S. computed all the relevant quantities for galaxies and the IGM, performed the statistical analysis and produced the scientific results focusing on understanding the quenching mechanisms affecting dwarf galaxies, under the supervision of C.P. and E.G. E.G. provided the simulations used in the work. The co-authors helped to interpret the results and gave comments on the manuscript's draft.

Relevant Co-Authored Publications

The following list indicates in preparation works that are complementary and related to results and analysis presented in this thesis. The author of this dissertation has contributed as co-author to these papers throughout the duration of the thesis project.

1. Garaldi E., Kannan R., Smith A., Borrow J., Vogelsberger M., Pakmor R., Springel V., Hernquist L., Shen X., Yeh J. Y.-C., Galárraga-Espinosa D., Xu C., Neyer M., Almualla M. and **Spina B.** (2023), *The* THESAN *project: public data release*, 2023, https://doi.org/10.48550/arXiv.2309.06475

   Author's contribution: B.S. contributed with the data analysis and editing the manuscript, discussing in-depth the results and interpretation with the lead-author.

Contents

CHAPTER 1

Introduction

The attempt of understanding our Universe started at the dawn of humankind. We passed from observing the sky with our bare eyes to more and more complex instruments. We have been exploring what is around us in many different ways, always with an unique point of view. But only observations are not enough to describe the underlying drivers of the Universe's evolution, we also need to interpret those observations. How to do this? Developing and testing theories from motivated cosmological principles, building and refining our theories in synergy with the observations. We develop our description of the Universe starting from the assumption that the Universe is homogeneous (matter is uniformly distributed) and isotropic (wherever an observer looks, the Universe appears the same). This principle holds only on the large scales, considering the Universe as a whole. Evidences of the validity of the cosmological principle are found in the way galaxies are distributed (Bundy et al., 2015) and in the observations of the Cosmic Microwave Background radiation (CMB, Planck Collaboration et al., 2020).

The cosmological principle has an important impact on the role the Earth and the humankind have in the Universe's framework: we are not special at all. And if we are not special, then what are we? We will leave this question open for the reader to answer.

We can set the beginning of modern cosmology around 1929, when Edwin Hubble, while observing the light coming from the galaxies around us, noticed that their recession velocities were increasing the further they were from us (Hubble, 1929). Since we are not special at all and we are not at the center of our Universe, the only explanation is that the Universe as at whole, the fabric of the Universe itself, is expanding and driving the galaxies embedded within in this expansion. If the Universe is expanding with the passing of time, let us rewind the cosmic history to its beginning. We conclude that there is a moment in time where *everything* is in the same point, all at once, in what we can call a

singularity. Way before us, this was noticed by George Lemaître (Lemaître, 1931b). So, how has everything started? And what happened from that moment on?

In the following, we present the foundational concepts necessary to comprehend modern cosmology. Starting from an overview of the fabric of the Universe itself we define the notion of distance in an expanding Universe in Section 1.1. We discuss in Section 1.2 the laws governing such expansion and their dependence on the composition of the Universe in Section 1.3. The post-recombination Universe is described in Section 1.4. The formation and evolution of structures under the influence of gravity are displayed in Section 1.5. We describe the processes governing the evolution of the Epoch of Reionization in Section 1.6, along with the techniques we use in this thesis to probe it. Finally, we give a brief overview on how to simulate the Universe and the Epoch of Reionization in particular in Section 1.7.

1.1. The fabric of the Universe

Cosmologically speaking, of all the forces acting in nature, gravity is the most interesting one. We can not describe it through a quantum-formulation, it is the only one dominant on the cosmological scales and it connects space and time by expressing them as a four-dimensional manifold, merging the concepts of energy and matter in a whole. Moreover, it is beautifully described by the general relativity field equations (EFE) in Einstein (1915)

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} - \Lambda g_{\mu\nu} \,, \tag{1.1}$$

where the indices μ and ν run over the four coordinates of space-time, conventionally $[0, 1, 2, 3] \equiv [ct, x, y, z]$. The left-hand side of Eq. (1.1) describes the geometry of the space-time fabric, through the Einstein tensor $G_{\mu\nu}$, given by the Ricci tensor $R_{\mu\nu}$, the Ricci scalar R and the metric tensor $g_{\mu\nu}$. The right-hand side of Eq. (1.1) represent the energy-matter content of a universe, via the stress-energy tensor $T_{\mu\nu}$ and the cosmological constant Λ (which appears in the case of an accelerated expansion of a universe). The constant quantities G and c are respectively the Newtonian gravitational constant and the speed of light in vacuum. In summary, EFEs show how the space-time fabric changes as a function of its mass (and thus energy) content.

Gravity is described by EFE and it is thanks to gravity if stars, galaxies, clusters and the large-scale structures (LSS) of the Universe are the ones we see today, and probably it has all started from tiny imperfections in the (almost) constant density field of the primordial Universe. It is actually funny when we think about it: physics and astrophysics aim to find general and abstract description of the reality, but the real Universe is quite a messy place. Anyway, these small perturbations originated non-uniform gravitational potentials, the non-linear collapses of matter and some time later, the LSS.

1.1.1. The Space-Time metric

We have been referring often to the space-time fabric, let us now explore this concept in a more detailed way. We can think our Universe as a 4-dimensional manifold, where one dimension represents time and the remaining three dimensions space. A metric is a mathematical object that specifies the distance between points in a manifold, it provides a way to measure intervals between events in space and time, determining how distances and angles are related between different points. For a completely flat space-time, the metric will be the Minkowski one and it can be expressed in spherical coordinates as

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}\chi^2 + \chi^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right), \tag{1.2}$$

where t is the cosmic time, χ the radial distance coordinate and θ and ϕ the angular coordinates.

If the cosmological principle holds, we can solve the EFE for comoving observers by generalizing the Minkowski metric in case of an expanding universe. We introduce the scale factor $a(t)$, such that comoving observers have constant spatial coordinates but the proper distances increases with the expansion of the Universe, and the curvature parameter K, which value is negative for a negative curvature (saddle-like space), null for a flat space and positive for a positive curvature (3-sphere-like space). The most generic form for such an universe is given by the Friedmann-Lemaître-Robertson-Walker metric (FLRW, Friedmann, 1924; Lemaître, 1931a; Robertson, 1933; Walker, 1937), as

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\left[\mathrm{d}\chi^2 + f_K^2(\chi)\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right], \tag{1.3}$$

where

$$f_K(\chi) = \begin{cases} |K|^{-1/2}\sinh^{-1}\sqrt{|K|}\chi, & \text{if } K < 0 \\ \chi, & \text{if } K = 0 \\ |K|^{-1/2}\sin^{-1}\sqrt{|K|}\chi, & \text{if } K > 0 \end{cases}. \tag{1.4}$$

1.1.2. Redshift

Let us consider a particle at initial (comoving) position $(\mathbf{x}, t)$ on the surface of an expanding sphere with uniform density $\rho(t)$. At time t, the particle is at (the physical) position

$$\mathbf{r}(t) = a(t)\mathbf{x}, \tag{1.5}$$

with the scale factor normalized so that $a(t_0) = 1$ at the current cosmic time t_0. The physical velocity $\mathbf{u}(t)$ follows as

$$\mathbf{u}(t) = \dot{a}(t)\mathbf{x} + a(t)\dot{\mathbf{x}} = \tag{1.6}$$

$$= \left(\frac{\dot{a}(t)}{a(t)}\right)a(t)\mathbf{x} + a(t)\dot{\mathbf{x}} = \tag{1.7}$$

$$\equiv H\mathbf{r} + \mathbf{v}, \tag{1.8}$$

where the derivatives are with respect to the cosmic time t, $\mathbf{v}$ is the peculiar velocity (without considering the expansion of the Universe) and $H = \frac{\dot{a}(t)}{a(t)}$ is the Hubble parameter and account for the expansion of the Universe. It is convenient to define the parameter H_0 as the Hubble constant, i.e. the Hubble parameter at present time. Historically, this parameter is furthermore parametrized as $H_0 = 100\,h\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Observations so far (e.g. Planck Collaboration et al., 2016a) suggest $h \approx 0.7$, but with large uncertainties. In particular, measurements obtained from high-redshift probes suggest $h \simeq 0.67$ (e.g. Planck Collaboration et al., 2016a), while observations at low-redshift suggest $h \simeq 0.73$ (e.g. Freedman et al., 2012; Riess et al., 2019). This disagreement is known as the Hubble tension (Di Valentino et al., 2021).

The first proofs of the expansion of the Universe were found by Vesto Slipher (Slipher, 1917) and some years later by Edwin Hubble, both observing galaxies with recessional velocities. Recessional velocities refer to the apparent motion of galaxies moving away from an observer, caused by the expansion of the Universe. When light from these distant galaxies reaches us, its wavelength appears stretched or "redshifted" due to the expansion of the space through which the light travels. This redshift is directly proportional to the recessional velocity of the galaxy. The concept is based on the idea that the Universe is expanding uniformly, and as a result, galaxies located at greater distances from us appear to be moving away at faster speeds. This shift towards the red end of the electromagnetic spectrum is called redshift

$$1 + z = \frac{\lambda_{\mathrm{obs}}}{\lambda_{\mathrm{em}}} = \frac{1}{a(t)}, \tag{1.9}$$

where λ_{em} is the wavelength emitted by the source in the source's rest-frame and λ_{obs} is the wavelength observed.

1.1.3. Distance measures

The redshift of an object is a good first measure of its distance from us (the higher the redshift, the larger the strecth of the emitted wavelength, the further from us is the object) but it gives us no information about the physical distance between the object and us. The concept of distance for an expanding universe with non-zero curvature has to be addressed carefully and indeed we can refer to several definition of distance.

Comoving distance

For the purposes of this work, we will refer often to the comoving distance. Let us consider a photon with only radial velocity (i.e. moving towards the observer). It follows a null-geodesics ($\mathrm{d}s^2 = 0$) and it has no angular component ($\mathrm{d}\theta = \mathrm{d}\phi = 0$). Therefore Eq. (1.3) simplifies as

$$c\mathrm{d}t = -a(t)\mathrm{d}\chi. \tag{1.10}$$

A photon emitted at a time t_0 and observed at a time t, is at comoving distance

$$\chi(t) = \int_t^{t_0} \frac{c\mathrm{d}t'}{a(t')} = \int_{a(t)}^{a(t_0)} \frac{c\mathrm{d}a}{a^2 H(a)}. \tag{1.11}$$

In terms of redshift, Eq. (1.11) reads

$$\chi(z_1, z_2) = \int_{z_1}^{z_2} \frac{c\mathrm{d}z'}{H(z')}. \tag{1.12}$$

Proper distance

The comoving distance, by definition, factors out the contribution from the expansion of the Universe. When taking this contribution into account, it is convenient to define the proper distance $d_{\mathrm{pr}}(t)$ as

$$d_{\mathrm{pr}}(t) = a(t)\chi(t). \tag{1.13}$$

Luminosity distance

Let us consider an object with absolute luminosity L and observed flux $F(r) = \frac{L}{4\pi r^2}$ at a radial distance χ. The luminosity distance is defined as

$$d_{\mathrm{L}} = \sqrt{\frac{L}{4\pi F}}. \tag{1.14}$$

Accounting for the expansion of the Universe, the flux observed is reduced by a factor $(1+z)^2$ (because the energy emitted by the source is reduced of a factor $1+z$ and the actual time of arrival is increased by a factor $1 + z$). Therefore, the relation between luminosity distance and proper distance is

$$d_{\mathrm{L}} = (1 + z)d_{\mathrm{pr}}. \tag{1.15}$$

Angular diameter distance

Let us consider an object with on observed angular size $\delta\theta$ and a physical transverse size l. The angular diameter distance is defined as the ratio between these two quantities, namely

$$d_{\mathrm{A}} = \frac{l}{\delta\theta}. \tag{1.16}$$

Under these assumptions, Eq. (1.3) reduces to $l = a(t)\chi\delta\theta$, therefore the angular diameter distance is given by

$$d_{\mathrm{A}}(z) = \frac{1}{(1 + z)^2} d_{\mathrm{L}}(z). \tag{1.17}$$

1.2. The Friedman equations

The set of GR field equations defined in Eq. (1.1) can be solved in some cases, in particular, if the cosmological principle holds and if the content of the Universe can be described as a perfect fluid (with energy density ρ and isotropic pressure p). The solutions describe the evolution of the scale factor along cosmic time for such a fluid. When combining Eq. (1.1) with Eq. (1.3), the two independent differential equations resulting are the Friedmann equations (Friedmann, 1922),

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3}\,, \tag{1.18}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}\,. \tag{1.19}$$

In this framework, the physical interpretation for the cosmological constant Λ is the vacuum (or dark energy) density, a fluid-like quantity with negative pressure that is driving the accelerated expansion of the Universe. The combination of the Friedmann equations gives

$$\frac{\mathrm{d}\left(a^3\rho c^2\right)}{\mathrm{d}t} = -p\frac{\mathrm{d}a^3}{\mathrm{d}t}\,, \tag{1.20}$$

which is nothing else then the analogue of the first law of thermodynamics, i.e. the internal energy of a closed system changes with its volume.

In order to solve Eq. (1.20), an assumption has to be made on the equation of state (EOS) $p(\rho)$, which characterize the matter-energy component. The common assumption is that $p = w\rho c^2$, where w is a constant value that changes based on the fluid considered. Solving Eq. (1.20) accordingly, the density as a function of time for each component is

$$\rho_{\mathrm{x}}(a) = \rho_{\mathrm{x},0}a^{-3(1+w)}\,. \tag{1.21}$$

The main components of the Universe, as far as we know, are the following:

- Cold, non-relativistic matter, with negligible pressure when compared with its energy density (Amendola & Tsujikawa, 2010) and $w_{\mathrm{m}} = 0$. Its energy density is $\rho_{\mathrm{m}} \propto a^{-3}$.
- Radiation (in general, anything with relativistic velocities), with w_{r}. Its energy density is $\rho_{\mathrm{m}} \propto a^{-4}$.
- Dark energy, a fluid-like quantity with negative pressure and therefore $w = -1$. Its energy density is constant, $\rho_{\mathrm{m}} = \mathrm{const}$.

Since their energy density depends on the scale factor, these three components dominate the Universe at different times. For this reason, it is convenient to define today's density parameter $\Omega_{\mathrm{x},0} = \rho_{\mathrm{x},0}/\rho_{\mathrm{crit}}$, where ρ_{crit} is the critical density at redshift $z = 0$, i.e.

$$\rho_{\mathrm{crit}} = \frac{3H_0^2}{8\pi G}. \tag{1.22}$$

The first Friedman equation reads

$$\left(\frac{H(a)}{H_0}\right)^2 = \frac{\Omega_{\mathrm{r},0}}{a^4} + \frac{\Omega_{\mathrm{m},0}}{a^3} + \frac{1-\Omega_{0,0}}{a^2} + \Omega_{\Lambda,0}, \tag{1.23}$$

where $\Omega_{0,0}$ is the total energy density parameter (today) and it is related to the curvature of the Universe and to the other parameters as $\Omega_{0,0} = -Kc^2/H_0^2 = \Omega_{\mathrm{r},0} + \Omega_{\mathrm{m},0} + \Omega_{\Lambda,0} = 1$.

The standard theory describing the Universe composition and growth is called the ΛCDM model, and it is the subject of the next section.

1.3. The Universe composition

In the previous section, the basics to describe the fabric of the Universe are laid, starting from Einstein's theory of general relativity. The components of the Universe follow such equations and dominate at different epochs in its expansion. But what are exactly these components? The purpose of this section is to give a brief summary of the content of the Universe (matter, radiation and energy) and the state-of-the-art model describing it, the ΛCDM model.

1.3.1. Radiation

After the Big Bang, the hot, dense Universe was dominated by its radiation content, i.e. all particles with relativistic energies. In a first approximation, these are described as a pressured fluid. With the expansion of the Universe and its subsequent cooling, the massive relativistic particles started to cool down as well and the number of relativistic particles to decrease. After this radiation-dominated phase, the matter-dominated epoch started.

The relativistic particles accounted in the radiation energy density are mainly photons, with a contribution from neutrino (depending on their rest mass), and its best estimate (today) is $\Omega_{\mathrm{r}} \approx 10^{-5}$ (Planck Collaboration et al., 2020).

Neutrinos are subatomic particles that are electrically neutral, have very low mass, and interact only weakly with matter. They are produced in many astrophysical environments, including the sun, supernovae, and cosmic ray interactions with the Earth's atmosphere. Neutrinos come in three different flavors - electron, muon, and tau - and can oscillate between these flavors as they travel through space. This phenomenon was first observed in the late 1990s and confirmed the existence of non-zero neutrino masses (Kajita, 2016; McDonald, 2016), which had previously been assumed to be zero. Neutrino oscillations have important implications for our understanding of fundamental physics, including the Standard Model of particle physics and the nature of neutrino masses. As for the boundary on neutrino masses, current experimental measurements suggest that the mass for the three flavors is $\Sigma m_\nu = 0.12$ eV at 95% C.L. (Planck Collaboration et al., 2020).

Neutrinos play an important role in astrophysics and cosmology. They are one of the most abundant particles in the Universe and can help us understand processes such as supernova explosions, the formation of large-scale structures in the Universe, and the nature of the early Universe.

CMB

Around 380,000 years after the Big Bang, the Universe had cooled down enough for electrons and protons to combine and form neutral atoms. Before this event, the Universe was filled with a hot, dense plasma of charged particles that effectively scattered and absorbed photons, preventing them from traveling freely. As the Universe continued to expand and cool, the first atoms formed, and the photons were no longer tightly coupled to the matter. This moment is known as recombination since neutral atoms formed by combining electrons and protons. At this point, the photons decoupled from the matter and began to travel freely through space. The photons that were emitted at this last scattering surface and have been traveling through space ever since. This radiation is known as the CMB, an electromagnetic radiation that permeates the entire Universe, with a temperature of approximately $T_{\mathrm{CMB}} = 2.73$ K (Penzias & Wilson, 1965). Moreover, it exhibits small fluctuations in temperature across the sky, known as anisotropies, at the level of $\approx 10^{-5}$, as shown in Figure 1.1. These anisotropies are thought to have arisen from quantum fluctuations in the early Universe, and provide valuable information about the composition, structure, and evolution of the Universe. The CMB radiation was discovered in 1964 by Arno Penzias and Robert Wilson, two astronomers working at Bell Labs in New Jersey, while conducting experiments for satellite communications. They found a persistent background noise that they could not eliminate, coming from all directions in space, corresponding to a faint glow of radiation left over from the Big Bang.

1.3.2. Baryonic matter

The matter we experience in - almost - everyday life is the baryonic matter, which includes protons, neutrons, electrons and the atoms they form. Baryonic matter is sensitive to all the known form of interaction (electromagnetic, strong, weak, gravity), via different mediators, and can therefore be detected by the current telescopes: the photons emitted in this interactions are the basis of today's observational astronomy. The today energy density parameter is estimated around $\Omega_{\mathrm{b}} h^2 = 0.0224 \pm 0.0001$ (Planck Collaboration et al., 2020).

It is clear that the combination of only radiation and baryonic matter is not enough to reach the total amount of energy density indicated by the observations. What is it missing?

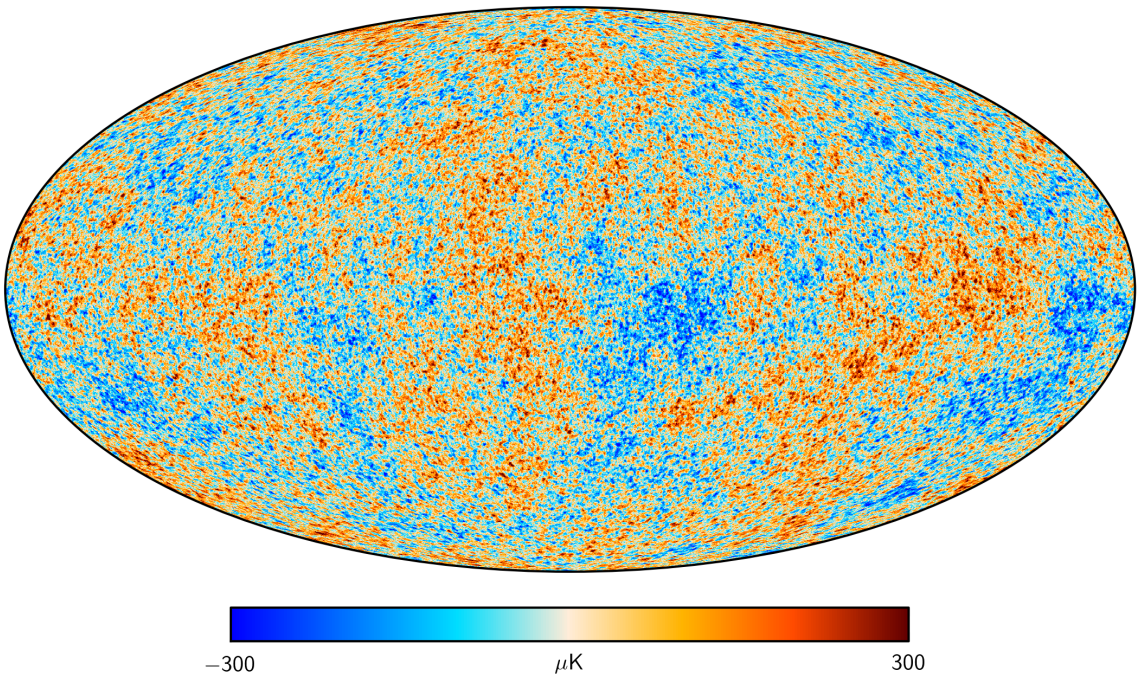

Figure 1.1. – The anisotropies of the Cosmic microwave background (CMB) as observed by Planck. Courtesy of Planck Collaboration.

1.3.3. Dark Matter

The first hints of what we call today dark matter (DM) dates back to the 30s, with the studies of the Coma Cluster by F. Zwicky (Zwicky, 1937). When applying the virial theorem to this cluster, he noticed a ($\tilde{4}$00 times) larger amount of matter than the one optically visible. In particular, the velocities of the galaxies near the edge of the cluster were too large for the amount of matter detected. Zwicky therefore postulated the existence of some dark - not visually observable - matter.

During the 60s and the 70s, more evidences of the DM existence were provided by Vera Rubin, Kent Ford, and Ken Freeman, once again studying the rotation curves of spiral galaxies (Rubin & Ford, 1970; Freeman, 1970). In the same period, the first measurements of the atomic hydrogen distribution (through the 21 cm emission line) were obtained with radio telescopes (Begeman, 1987). Neutral hydrogen (HI) is also distributed on the outskirt of galaxies and it extends to larger radii than the optically visible ones, extending the limit of the rotation curves and proving the total mass distribution.

More and more evidences supporting the existence of DM were obtained in the 80s, for example via the gravitational lensing measurements (i.e. the amount of shape and size deformation of background objects by galaxy clusters and DM distribution, Walsh et al., 1979), the anisotropies in the CMB distribution (and the position of the first peak of its angular power spectrum, Smoot et al., 1992) and the distribution of hot gas within galaxies and galaxy clusters (McKee & Ostriker, 1977).

This kind of matter only interacts via gravity with baryonic matter, but it seems not to be affected by the other known forces present in Nature. Constraints on the DM particles, both from astrophysics and particle physics (Leauthaud et al., 2012), suggest a *cold* dark matter (CDM), i.e. matter with a non-relativistic velocities disentangled from radiation. Conversely, we talk about *hot* and *warm* dark matter (Bond et al., 1984, WDM,) depending on their relativistic (or weakly non-relativistic) velocities. A cold DM model leads to a bottom-up scenario in the formation of structure: galaxies formed first and galaxy clusters at a latter stage. The opposite is expected for an hot dark matter scenario (HDM), where these free-streaming particles prevents the formation of small scales structures. While an HDM scenario has already been excluded from high-z observations, WDM particles with masses $m_{\mathrm{WDM}} \geq 1$ keV have not been ruled out yet. Experiments (especially on the particle physics side) for the direct and in-direct detection of a DM signal have been performed using high-energy proton-proton collisions at the LHC, and through the study of cosmic rays in location where the ambient radiation is shielded. For an extensive review, refer to (Schumann, 2019).

After the Big Bang, the Universe was dominated by its radiation component. With the expansion and the cooling down of the Universe, the matter component started to play a more and more important role, until it became dominant at $z_{\mathrm{eq}} \approx 3400$ (following from Friedman's equation). DM dominated the expansion of the Universe until $z \approx 0.3$, when the negative pressured fluid we refer at as dark energy began to dominate and keeps dominating today as well.

Recent constraints on the abundance of DM suggest $\Omega_{\mathrm{m}} = \Omega_{\mathrm{b}} + \Omega_{\mathrm{DM}} = 0.3111 \pm 0.0056$, meaning that the energy density parameter relative to dark energy has to be around $\Omega_{\Lambda} = 0.6889 \pm 0.0056$ for a flat Universe (Planck Collaboration et al., 2020).

1.3.4. Dark Energy

Dark energy dominates the current epoch of the evolution of the Universe, driving its accelerated expansion. Let us briefly summarize the probes indicating the existence of this pretty much unknown quantity and their contribution in constraining cosmological parameters (Figure 1.2).

Accurate measurements of the CMB radiation from Planck Collaboration et al. (2020) are in agreement with the ΛCDM model (Hu & Dodelson, 2002). In particular, the shape and the position of the angular power spectrum gives us information about cosmological parameters. A spatial curvature $K = 0$ is consistent with such measurements, indicating that the energy densities of matter (both baryonic and dark matter) and radiation are not enough to meet the total energy density required from the Friedman's equations.

The distribution of galaxies is another strong indicator of the existence of dark energy. Assuming that the position of galaxies is a good tracer of the underlying DM distribution, the density peaks of such distribution has a separation distance r_s (corresponding to the position of the baryonic acoustic oscillations) in the monopole of the correlation function.

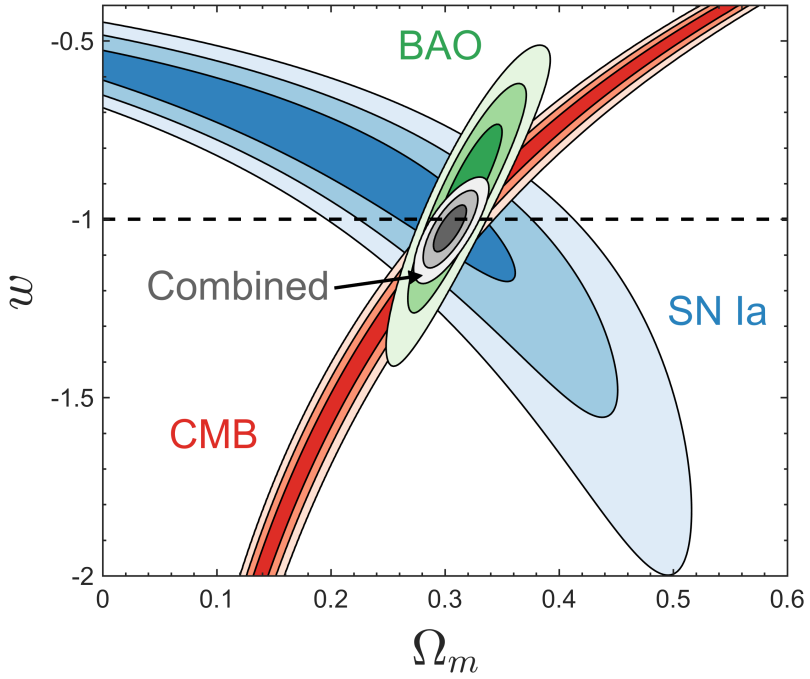

Figure 1.2. – Constraints on w and Ω_m parameters from Huterer & Shafer (2018), from measurements of the CMB (in red, Planck Collaboration et al., 2016b), BAO (in green, Betoule et al., 2014) and SN Ia (in blue, Alam et al., 2017). Contours indicate model C.I. of 68.3%, 95.4%, and 99.7%.

This quantity has been used to set a distance-redshift relation and it is consistent with a $w_\Lambda = -1$ cosmological constant (Alam et al., 2017).

As already mentioned in the previous section, distances are a complicated subject in astronomy. This is why it is important to consider objects as Supernovae Type Ia as a standard ruler to set a distance-redshift relation. These objects have a characteristic luminosity function, from which a luminosity distance can be obtained. They originate from white dwarf stars reaching a critical mass and undergoing a consistent and explosive fusion process. This uniform luminosity makes them valuable "standard candles" in cosmology. In the late 90s, independent measurements of Supernovae Type Ia from Riess et al. (1998) suggested an accelerated expanding Universe (not compatible with a Universe filled only with matter and radiation). Current constraints using SN Ia are set by Betoule et al. (2014).

1.4. The post-recombination Universe: the formation of structures

After the radiation-dominated epoch, matter started dominating the energy content of the Universe. This non-relativistic component can be described as a pressureless fluid, with density $\rho = \rho(\mathbf{r}, t)$ and velocity $\mathbf{u}$ in a Newtoninan framework (as a first approximation)

in a gravitational field with potential ϕ. The coordinate $\mathbf{r}$ is the proper spatial coordinate. The time evolution is described by the continuity equation (expressing mass conservation), the Euler equation (linked to the conservation of momentum) and the Poisson equation (including the cosmological constant),

$$\frac{\partial \rho}{\partial t} = -\rho \nabla_r \cdot \mathbf{u}, \tag{1.24}$$

$$\frac{\partial \mathbf{u}}{\partial t} = -\left(\mathbf{u} \cdot \nabla_r\right) \mathbf{u} - \nabla_r \phi, \tag{1.25}$$

$$\nabla_r^2 \phi = 4\pi G \rho - \Lambda. \tag{1.26}$$

The density ρ can be written in terms of the density perturbation contrast against the background, as

$$\rho(\mathbf{x}, t) = \bar{\rho}(t)[1 + \delta(\mathbf{x}, t)], \tag{1.27}$$

where $\bar{\rho}(t)$ is the spatially averaged density. When converting the fluid equations to comoving coordinates and expanding to first order in the density field and velocity, the set of equations describing the fluid motion reads

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla_x \cdot \mathbf{v} = 0, \tag{1.28}$$

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a} \mathbf{v} = -\frac{1}{a} \nabla_x \Phi, \tag{1.29}$$

$$\nabla_x^2 \Phi = 4\pi G \bar{\rho} a^2 \delta, \tag{1.30}$$

where $\Phi(\mathbf{x}, t) \equiv \phi\left(a\mathbf{x}, t\right) + a\ddot{a}/2|\mathbf{x}|^2$. It is worth mentioning that the cosmological constant seems not to play a role in this set of equations. However, the evolution of the scale factor depends on it, making these equations sensitive to the dark energy content as well. When combining these equation, we obtain a second-order differential equation describing the time evolution of the density contrast

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho} \delta. \tag{1.31}$$

This equation admits a generic solution of type $\delta(x, t) = D_+(t)A(\mathbf{x}) + D_-(t)B(\mathbf{x})$, where $A(\mathbf{x}) and B(\mathbf{x})$ are comoving density fluctuation fields. If the late-time Universe is dominated by a cosmological constant, the decreasing mode is $D_-(t) \propto H(t)$ and the corresponding growing mode is given by

$$D^+(t) \propto \frac{\sqrt{\Omega_\Lambda a^3 + \Omega_k a + \Omega_m}}{a^{3/2}} \int \frac{a^{3/2} \mathrm{d}a}{(\Omega_\Lambda a^3 + \Omega_k a + \Omega_m)^{3/2}}. \tag{1.32}$$

We have been discussing how structures form under the influence of gravity. This process is dominated by DM, but baryonic matter plays an important role as well. The high temperature (and therefore velocity) that electrons had before recombination made it

possible its coupling with photons via Thompson scattering, and with baryonic matter via the Coulomb scattering. Two opposite forces drive baryons in the growth of structures: gravity from the dominating DM component clusters the baryons into their potential well, pressure from the interaction with photons drives the baryons outside the well, originating oscillating perturbations called baryonic-acoustic-oscillations (BAO). After recombination, the photon's pressure stops being effective and baryons definitely fall into the DM potential well, became DM halos in the meantime. At this point, the growth of DM structures and baryon structures evolves at the same speed.

1.5. Clustering of structures

We require a statistical formalism to quantitatively describe the LSS observations we conduct, which then allows us to compare observational data with theoretical models. One-point statistics are powerful tools to describe the properties of an ensemble, but give little information about the spatial distribution of galaxies. For this reason, it is useful to consider two-point statistics, such as the two-point correlation function.

For an overdensity $\delta(\mathbf{x})$, the two-point correlation function is then defined by

$$\xi(d) \equiv \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{d}) \rangle \tag{1.33}$$

where $\mathbf{d}$ defines the separation between two points in the density field, and due to homogeneity and averaging over all directions, ξ only depends on the modulus $d = |\mathbf{d}|$. The correlation function is a measurement of the excess probability of objects (galaxies in this case) being separated by $\mathbf{d}$ compared with a randomly distributed density field. It is often convenient to work with the Fourier space equivalent of the correlation function $P(k)$, referred to as the power spectrum, which is related to the Fourier-transformed overdensity field $\delta(\mathbf{k})$ by

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = (2\pi)^3 P(k)\delta^3(\mathbf{k}-\mathbf{k}') \tag{1.34}$$

where $\delta^3(\mathbf{k}-\mathbf{k}')$ is the Dirac delta function, and the Fourier transform of the overdensity is given by

$$\delta(\mathbf{k}) = \int e^{-i\mathbf{k}\cdot\mathbf{x}}\delta(\mathbf{x})\mathrm{d}^3x \tag{1.35}$$

and the inverse is given by

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3}\int e^{i\mathbf{k}\cdot\mathbf{x}}\delta(\mathbf{k})\mathrm{d}^3k. \tag{1.36}$$

As for the correlation function, the power spectrum is only dependent on the modulus of the wavenumber $k = |\mathbf{k}|$. Accounting for the evolution of the overdensity field, the power spectrum in linear theory reads

$$P(k,z) = D_+^2(z)P(k,0). \tag{1.37}$$

A complete description of the power spectrum is obtained when fixing its overall amplitude: this is not predicted a priori since there is not a defined theory for the origin of cosmic perturbations, so it has to be fixed by observations. Historically, the prescription for normalizing a theoretical $P(k)$ involves the variance of the galaxy distribution when sampled with randomly placed spheres of radii R. These quantities are related by

$$\sigma^2(R,z) = \frac{1}{2\pi^2}\int P(k,z)W_R^2(k)k^2\mathrm{d}k, \tag{1.38}$$

where $W_R(k)$ is the Fourier transform of a window function $W_R(r)$. For the top-hat window function,

$$W_R(r) = \begin{cases} \left(\frac{4\pi}{3}R^3\right)^{-1} & \text{if } r \leq R, \\ 0 & \text{otherwise,} \end{cases} \tag{1.39}$$

corresponding in Fourier space to

$$W_R(k) = \frac{3}{(kR)^3}\left[\sin(kR) - kR\cos(kR)\right]. \tag{1.40}$$

The window size can be described by the mean mass contained in it

$$M(R) \equiv \overline{\rho}(t_0)\frac{4\pi}{3}R^3. \tag{1.41}$$

The normalization of the amplitude of the power spectrum is often specified by the value of $\sigma_8 \equiv \sigma(R = 8h^{-1}\mathrm{Mpc})$.

The structure formation in CDM models proceeds hierarchically: at early times, the majority of DM is situated in low-mass halos and high-mass halos are formed from the merging of the former. We saw in the previous sections that the majority of the matter content in the Universe is in the form of DM. In order to trace the DM distribution, we rely on its tracers, which most commonly is emission from galaxies. In the linear regime, we can describe the relation between these two fields by a single linear factor referred to as the bias b_g, where

$$\delta_g = b_g\delta \Rightarrow P_g(k) = b_g^2 P(k). \tag{1.42}$$

We therefore assume a relation between the baryonic, observable, matter and the DM distribution in our Universe. Studying the distribution of galaxies means studying the underlying DM halos distribution. And so we ask ourselves: how densely is the Universe populated? And how common are halos with a given mass?

The Press-Schechter halo mass function (Press & Schechter, 1974) provides a statistical description of the abundance of halos as a function of their mass. The idea behind the Press-Schechter formalism is to relate the growth of density perturbations in the early Universe to the formation and evolution of DM halos. It assumes that the density fluctuations follow a Gaussian distribution, and it provides a probabilistic framework to describe the collapse of these perturbations into halos. The key quantity in the Press-Schechter formalism is

the critical density contrast δ_c, which represents the threshold for collapse. Halos form in regions where the linear density contrast exceeds this critical value. The halo mass function can be expressed as:

$$n(M) = \frac{\bar{\rho}}{M}\frac{d\nu}{dM}f(\nu), \tag{1.43}$$

where M is the mass of the halo, ν is the peak height parameter, $\frac{d\nu}{dM}$ represents the mass derivative of ν, and $f(\nu)$ is the multiplicity function (which characterizes the fraction of collapsed regions that form halos with a given ν).

The peak height parameter ν is defined as the ratio of the critical density contrast to the root mean square of the density fluctuations in the smoothed field:

$$\nu = \frac{\delta_c}{\sigma(R)}, \tag{1.44}$$

where $\sigma(R)$ is defined in Eq. (1.38).

Now that we have set the formalism describing the growth of structures from the initial density perturbations, we can take a closer look at the baryonic content and its evolution.

1.6. Epoch of Reionization

The moment of decoupling between matter and radiation marks the beginning of the dark ages (which are not referring to some obscure medieval times, just to the complete absence of light in the whole Universe). No sources of radiation, and therefore no light, was present in the Universe. With the formation of the first hydrogen atoms, more and more complex objects began to form, until the complex molecules and structures we observe today. As stars and galaxies were forming, the intergalactic medium (IGM) filling the space in between started to be ionized by the high-energy photons emitted by these objects. It is the purpose of this section to describe the processes which drove the EoR and how to probe those. We refer the reader to Barkana & Loeb (2001); Furlanetto et al. (2006); Furlanetto & Oh (2008a); Pritchard & Loeb (2012); Zaroubi (2013) for a complete review.

1.6.1. And there was light

With the progressive ionization of the IGM, the Epoch of Reionization (EoR) started. Photons emitted by star-forming galaxies at high redshift ionize HI (whose ionization potential $E_P^{\rm HI}$ is 13.6 eV) and neutral helium ($E_P^{\rm HeI}$ = 24.6 eV), but do not contribute significantly to helium reionization ($E_P^{\rm HeII}$ = 54.4 eV). Photons with higher energies are required to fully ionize helium, as the ones emitted by Active Galactic Nuclei (AGNs). These objects are the extremely luminous centers of galaxies, powered by accretion of mass onto supermassive black holes. As matter falls toward the black hole, it forms an

accretion disk that emits vast amounts of energy across various wavelengths, from radio waves to X-rays. AGNs exhibit highly variable brightness and often emit powerful jets of particles and radiation. Quasars (QSOs) are a subclass of AGNs, representing the most energetic and distant members. They emit intense radiation, often outshining their host galaxies. Quasars were more common in the early Universe, playing a significant role in ionizing intergalactic gas during the epoch of reionization. The history of hydrogen helium reionization is closely connected to the properties of the QSOs and AGNs population (e.g. their luminosity function).

The sources of the ionizing photons are the first galaxies which formed after the dark ages, but their properties are still mainly unknown. Groundbreaking results in this direction are expected from experiments like the James Webb Space Telescope (JWST, Gardner et al., 2006) which will unveil the dark ages and the reionization sources. Objects like the AGNs would be energetic enough to contribute to the ionizing photons budget, but their abundance at high redshift ($z \geq 4$) is too small too produce a significant contribution. This standard picture might be questioned if a larger population of QSOs than expected is detected at high redshift, as claimed by Giallongo et al. (2015). An abundance of faint AGNs would boost the reionization photon budget at high redshift, but it won't contribute significantly to the evolution of the EoR at lower redshift. However, it would be a game changer for the second-ionization of helium (hereafter referred as helium reionization).

The process of forming molecular hydrogen H_2 in the IGM is hampered by the recombination happening in the IGM itself, preventing the growth of the ionized bubble surrounding the first stars. In this first pre-overlap phase, the Universe is mostly neutral, with the exception for the isolated HII regions. The recombination process mainly depends on the density of the gas, becoming less effective with the expansion of the Universe and the decrease of the gas density. At the same time, the gravitational collapse is responsible for the formation of new stars and galaxies, and the associated increase of ionizing photons. In the central phase of the EoR (starting at $z \approx 20$), the HII bubbles start growing and overlapping, thanks to the larger number of ionizing photons available and the Universe becomes more and more ionized. Within regions denser in gas, like protogalaxies and protoclusters, the balance between recombination and photo-ionization remains unaffected from the growth of HII bubbles. Gas neutral regions are therefore located within self-shielded, high-density clouds, where the radiation field can not enter. As galaxy formation proceeds, the gas high-density regions become the only neutral regions in the IGM.

The evolution of the EoR is strictly connected to the overdensities at high redshift, where the first objects started forming and ionizing the surrounding environment. The topology of the ionized bubbles and of their overlapping depends on the clustering properties of the sources and on the cosmological parameters overall.

1.6.2. Probing the EoR

The evolution and the topology of the EoR are still not deeply understood. The most promising tool to study the EoR is via the detection of the photons emitted in the hyperfine transition of HI. This transition is between the two hyperfine states of the $1^2S_{1/2}$ ground level of hydrogen and it is induced by the coupling between the proton and electron spins. Either the two spins are aligned (maximum energy configuration, called the triplet state), or anti-aligned (minimum state configuration, called the singlet state). The transition between the two configurations is equivalent to a spin-flip of the electron, with an emitted energy of $\Delta E = h\nu = 5.9 \times 10^{-6}$ eV (corresponding to a wavelength of $\lambda_{\rm HI} = 21$ cm and a frequency of $\nu_{\rm HI} = 1420$ MHz). A similar transition happens for the single-ionized helium-3 atom (HeII), with a wavelength of $\lambda_{^3{\rm HeII}} = 3.46$ cm and a frequency of $\nu_{^3{\rm HeII}} = 8.66$ GHz.

The excitation temperature of these emission lines, the spin temperature T_S, depends on the number densities of neutral hydrogen (or single-ionized helium) in the singlet level n_0 and in the triplet level n_1,

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} \exp\left(-\frac{T_\star}{T_S}\right), \tag{1.45}$$

where $T_\star \equiv \dfrac{hc}{k_{\rm B}\lambda}$ $\left(= 0.068\,{\rm K}$ for HI, $0.416\,{\rm K}$ for ^{3}HeII$\right)$, and g_0 and g_1 are the statistical weights of the two configurations.

The emission line corresponding to this hyperfine transition is a powerful tool to probe the hydrogen (or helium) distribution accounting for background sources, as the CMB one. The brightness temperature $\delta T_b(z)$ associated to the hyperfine transition is computed with respect to the CMB as

$$\delta T_b(z) = \frac{T_S(z) - T_{\rm CMB}(z)}{1+z}\left[1 - e^{-\tau(z)}\right]. \tag{1.46}$$

Following Furlanetto et al. (2006), the corresponding optical depth $\tau(z)$ is

$$\tau(z) = \frac{g_1}{g_0+g_1}\frac{hc^3}{k_b}\frac{A_{10}}{8\pi\nu_{\rm emitt}^2}\frac{n_{\rm H,^3He}(z)x_{\rm HI,^3HeII}(z)}{T_S(z)}\frac{1}{(1+z)\frac{{\rm d}v_\parallel}{{\rm d}r_\parallel}} \tag{1.47}$$

where the constant quantities h and k_b are respectively the Planck constant and the Boltzman constant, A_{10} is the spontaneous emission coefficient of the 21 cm (3.46 cm) transition, $\nu_{\rm emitt}$ the rest-frequency of the emission line, $n_{\rm H,^3He}(z)$ is the number density of either H or ^{3}He, $x_{\rm HI,^3HeII}(z)$ the neutral fraction of either hydrogen or 3-helium, $\frac{{\rm d}v_\parallel}{{\rm d}r_\parallel}$ is the velocity gradient along the line-of-sight, including both the Hubble expansion and the peculiar velocity.

Let us write the number density $n_Y(z)$ for a species Y as a function of the the baryon density field δ_b:

$$\begin{aligned} n_Y(z) &= \frac{\rho_Y}{m_Y} = \\ &= \frac{(1+\delta_b)\bar{\rho}_b(z) f_Y}{m_Y} = \\ &= \frac{(1+\delta_b)\Omega_{b,0}\rho_{b,0}(1+z)^3 f_Y}{m_Y} = \\ &= \frac{(1+\delta_b)\Omega_{b,0}\frac{3H_0^2}{8\pi G}(1+z)^3 f_Y}{m_Y}, \end{aligned} \tag{1.48}$$

where f_Y is the mass fraction of the species (e.g. $f_{\rm H} = 0.76$, $f_{^4{\rm He}} = 0.24$), $\rho_Y(z)$ its density and $\bar{\rho}_b(z)$ the baryon mean density. The abundance of ^{3}He with respect to H is (Steigman, 2007)

$$10^5(^3{\rm He}/{\rm H}) = 1.1, \tag{1.49}$$

while the mass fraction is given by

$$f_{^3{\rm He}} = f_{\rm H}\,(^3{\rm He}/{\rm H})\,\frac{m_{^3{\rm He}}}{m_{\rm H}} = 2.28\times10^{-5}. \tag{1.50}$$

Combining Eqs. from (1.45) to (1.50) and assuming that in a matter-dominated Universe the Hubble parameter can be approximated to $H(z) = H_0\Omega_m^{0.5}(1+z)^{3/2}$, the brightness temperature is

$$\delta T_b(z) = \overbrace{\left[\frac{g_1}{g_0+g_1}\frac{hc^3}{k_b m_{\rm H,^3He}}\frac{A_{10}}{8\pi\nu^2_{\rm emitt}}\frac{\Omega_{b,0}}{\Omega_m^{0.5}}\frac{3H_0}{8\pi G}f_{\rm H,^3He}\right]}^{C} x_{\rm HI,^3HeII}(z)(1+\delta_b)(1+z)^{1/2} \left[1-\frac{T_{\rm CMB}(z)}{T_S(z)}\right]\left[\frac{H(z)/(1+z)}{{\rm d}v_\parallel/{\rm d}r_\parallel}\right]. \tag{1.51}$$

Equation (1.51) holds both for the 21 cm emission line from the hyperfine transition for HI and for the 3.46 cm emission line for HeII. The constant quantities in the first term of the equation, labeled as C, depend on the transition considered, and they are summarized in Table 1.1. The emission from HeII is several order of magnitudes lower than the one for HI, but still interesting for studying the drivers of helium reionization and the QSOs population features. The evolution of the brightness temperature for the 21 cm line is shown in Figure 1.3.

Depending on the value of the spin temperature with respect to the CMB temperature, three cases are possible:

- $T_S > T_{\rm CMB} \longrightarrow \delta T_b > 0$, emission line signal,

	g_0/g_1	m $[10^{-27}$ kg]	A_{10} $[10^{-12}$ s$^{-1}]$	$\nu_{\rm emitt}$ [GHz]	f	C [mK]
HI	1/3	1.67	$2.84 \cdot 10^{-3}$	1.42	0.76	8.8
^{3}HeII	3/1	5.01	1.92	8.66	$2.28 \cdot 10^{-5}$	$5.3 \cdot 10^{-4}$

Table 1.1. – Constant quantities in Eq. (1.51) for the hyperfine transitions for HI and ^{3}HeII

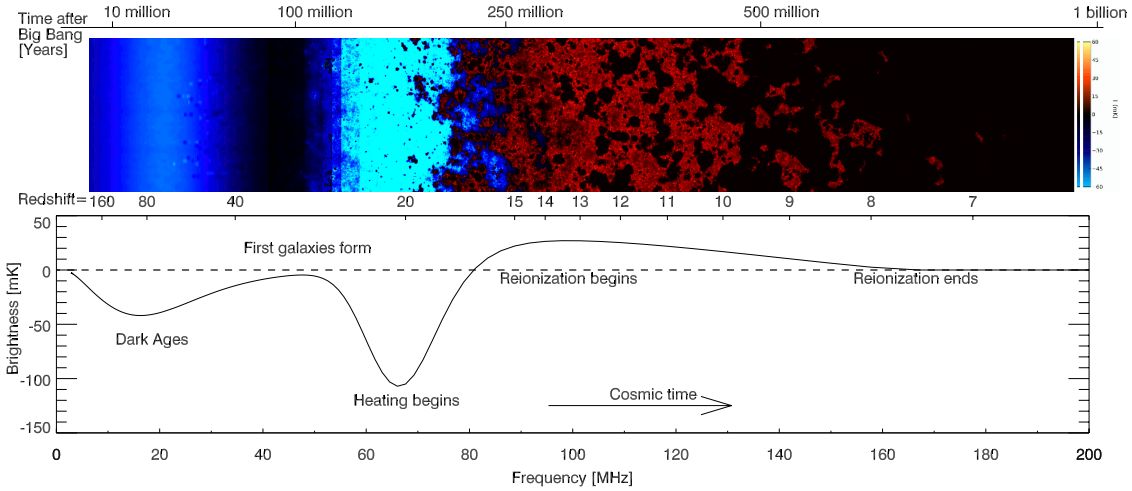

Figure 1.3. – Evolution of the 21cm brightness temperature. In the top panel, the absorption phase of the 21cm radiation is indicated in blue, the emission in red. In the bottom panel, the brightness temperature along the different epochs (from the beginning of stars formation to the end of the reionization epoch) is shown. From Pritchard & Loeb (2012).

- $T_S = T_{\rm CMB} \longrightarrow \delta T_b = 0$, no signal,
- $T_S < T_{\rm CMB} \longrightarrow \delta T_b < 0$, absorption line signal.

For a comprehensive review on the determination of the spin temperature, we refer to Furlanetto et al. (2006). In summary, the spin temperature depends on three processes: *(i)* 21 cm photons are absorbed or emitted by background radio sources, such as the CMB, *(ii)* other hydrogen atoms and free electrons collides with the interested photons, *(iii)* the scatter with Ly-α photons might induce an intermediate excited state and a spin flip.

As a first approximation, the spin temperature is given by the equilibrium of these processes,

$$T_S = \frac{1 + x_\alpha + x_c}{\dfrac{1}{T_{\rm CMB}} + \dfrac{x_\alpha}{T_\alpha} + \dfrac{x_c}{T_c}}, \tag{1.52}$$

where T_α the temperature of the Ly-α radiation, T_K is the gas kinetic temperature, closely coupled with T_α by recoil during repeated scattering, x_c is the coupling coefficient regarding atomic collisions and x_α is the coupling coefficient due to the scattering of Ly-α photons.

The Lyman-α line is one of the transitional lines in the Lyman series, which result in ultraviolet emission for any one-electron atom. For an hydrogen atom the wavelength of the line is 1.22×10^{-7} m, while for a single-ionized helium atom s of 3.04×10^{-8} m. This transition is sensitive to small column densities of the gas and therefore a powerful tool to constrain the end of reionization. The peak produced by this transition is very sharp and it is optimal to identify the redshift position of the background source responsible for it. The Ly-α peak in the spectra of high-redshift bright sources, as QSOs, inform us about the evolution and topology of the reionization front.

1.6.3. Line-Intensity Mapping

The 21 cm emission line for HI and the 3.46 cm emission line for HeII are powerful tools to prove the EoR. The IGM, consisting mainly of hydrogen and helium, permeate the Universe and makes the detection of these emission lines possible on large scales. Instead of highly-resolving a small region of the IGM, it is more convenient and more informing to track the diffuse distribution of the emission lines. The Line-Intensity Mapping (LIM) technique consists in the measurement of the spatial fluctuations of the brightness temperature defined in Eq. (1.51), with a poor angular resolution but surveying large - cosmological - volume. Moreover, the redshift range accessible from LIM measurements has never been explored before with such large coverage (galaxy surveys usually explore the Universe for $z \lesssim 2$) and it is sensitive to all sources of emission in the line, allowing to access faint and extended emission sources. This technique promises to be very efficient in probing the large scale structures and the evolution of the IGM during the EoR (for a review see Bernal & Kovetz, 2022).

We have mentioned here the 21 cm and 3.46 cm emission lines, but other emission lines are accessible for LIM (in particular for more astrophysical purposes). In particular, the forbidden emission line [CII] is used to estimate the star-formation-rate in galaxies, informing about galaxy formation and evolution (e.g. Padmanabhan, 2019). Likewise, the [OIII] line is proper to the diffuse and highly ionized regions near young O-type stars and provide information about low-metallicity environments where photodissociation regions are not dominating the interstellar medium (e.g. Gong et al., 2017).

The main three goals of LIM studies are the EoR, galaxy formation and evolution, and the large scale structures. Several experiments are pointing in these directions and many more are expected to be developed in the future. Probing the (end of the) EoR at redshift $z \lesssim 6$ is possible through the HI distribution within galaxies and galaxy clusters, which can be achieved via experiments as CHIME (Canadian Hydrogen Intensity Mapping Experiment, Bandura et al., 2014), HIRAX (Hydrogen Intensity and Real-time Analysis eXperiment, Newburgh et al., 2016a) or SKA-MID (Square Kilometre Array Phase 1 MID, Dewdney et al., 2009) for higher spatial distribution. At higher redshifts, the optimal radio telescopes are HERA (Hydrogen Epoch of Reionization Array, DeBoer et al., 2017) and SKA-LOW (Dewdney et al., 2009), which will extend the detection of the 21 cm emission line up to redshift $z \approx 30$. Lines associated to [CII] and CO transitions will be detected by e.g. TIME

(Crites et al., 2014), CCAT-prime (Cherro Chajnantor Atacama Telescope-prime, Stacey et al., 2018), CONCERTO (CarbON CII line in post-rEionisation and ReionisaTiOn epoch, CONCERTO Collaboration et al., 2020) and COMAP (CO Mapping Array Project, Cleary et al., 2022), at different redshift and resolution, from the ground, and from EXCLAIM (EXperiment for Cryogenic Large-Aperture Intensity Mapping, Pullen et al., 2023) from the sky. A summary of the main intensity mapping experiments is shown in Figure 1.4, displaying the redshift range, the angular extent of the survey, the line(s) targeted and the funding status. We refer the reader to Bernal & Kovetz (2022) for a comprehensive review on LIM techniques and experiments.

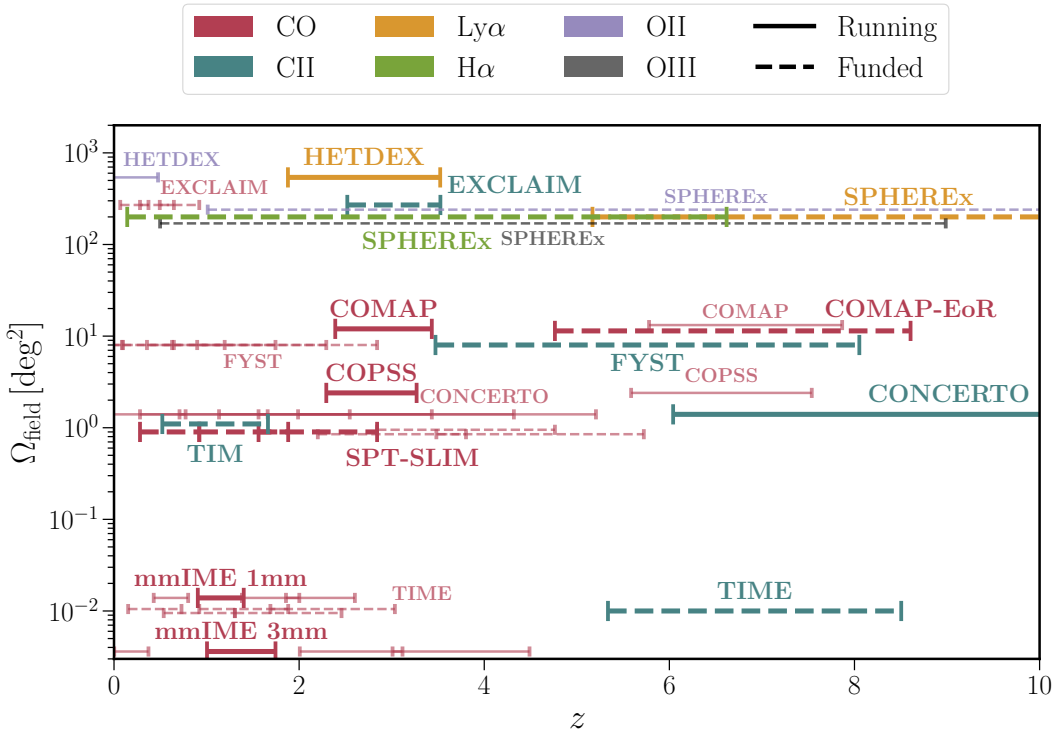

Figure 1.4. – A summary of ongoing and up-coming intensity mapping experiments. For each experiment, the targeted emission line(s), the redshift range and the total sky coverage are reported. From Bernal & Kovetz (2022).

1.7. Simulating the EoR

Observations of the EoR are not enough to understand the processes that governed it. Once observations are available, we need to extract all the possible physical information from it. In order to do this, we need to set up the theoretical framework and the simulations counterpart to test it, confronting theory with observations.

We can divide cosmological simulations in two main categories, N-body simulations and hydrodynamical simulations. N-body simulations (Bagla & Padmanabhan, 1997) focus on the gravitational interactions between DM particles, computing the motion of particles based on the gravitational forces between them. They simulate the dynamics of DM on large scales and are well-suited for studying the formation and evolution of cosmic structures like galaxies and galaxy clusters. Pros of N-body simulations include their computational efficiency, which allows them to simulate large volumes of the Universe over long time spans. However, they do not directly model gas and other baryonic components, limiting their ability to study processes like star formation and feedback. On the other hand, hydrodynamical simulations (Springel, 2005) incorporate the effects of gas dynamics alongside gravity. They model both DM and baryonic matter, making them more comprehensive for studying the interplay between gas, galaxies, and the cosmic environment. This allows them to explore a broader range of physical processes, such as cooling, star formation, and feedback mechanisms. However, hydrodynamical simulations are computationally more demanding and require higher resolution to accurately capture small-scale processes. This section provides an overview of the current state of EoR simulations, emphasizing the methodologies employed and key considerations.

Let us summarize here the main ingredients to run a cosmological simulation. Cosmological initial conditions have to be set with the appropriate cosmological model, specifying the initial density field, DM distribution, and velocity perturbations. Simulations track the hierarchical growth of DM halos through gravitational collapse and mergers; halo properties are essential for understanding the formation and evolution of the first luminous sources responsible for reionization. Hydrodynamic equations, such as the Euler and continuity equations described in Section 1.4, have to be solved to capture the evolution of cosmic gas. These equations incorporate fluid dynamics, gravity, and radiative transfer, accounting for processes like adiabatic cooling, shock heating, and pressure forces.

The simulation of radiative transfer (Mishchenko, 2003) plays a critical role in capturing the ionization process during the EoR. Several techniques are employed to model the propagation of ionizing photons. In Monte Carlo methods (Noebauer & Sim, 2019), ionizing photons are emitted from radiation sources (e.g., galaxies, quasars) and traced as photon packages through the simulated volume. Interactions with neutral gas (e.g., absorption, ionization) are stochastically determined, allowing the calculation of the ionization field. Instead of solving the complex radiative transfer equations that govern the propagation of ionizing photons through the intergalactic medium, semi-numerical methods (Janett et al., 2019) employ approximate algorithms to estimate the ionization state of different regions in the Universe. This technique provides a rapid estimation of large-scale ionization fields while sacrificing detailed small-scale information. An example is the 21CMFAST code described in Mesinger et al. (2011). Achieving accurate and reliable EoR simulations requires addressing numerical challenges and achieving sufficient resolution. Eulerian grid-based methods, such as adaptive mesh refinement (AMR), use fixed or moving grids to discretize the computational domain and enable higher resolution in regions of interest (as an example, see Gnedin, 2000). Lagrangian methods, such as

smoothed particle hydrodynamics (SPH) follow the evolution of individual particles and estimate their physical quantities by interpolating the properties of nearby particles. This smoothing process ensures a continuous representation of the fluid or matter distribution (see the SPH GADGET-2 code and the implementation of radiative transfer are presented in Iliev et al., 2006; Di Matteo et al., 2008).

It is worth mentioning the THESAN suite of simulations (Kannan et al., 2022), radiation-magneto-hydrodynamic simulations that model the large-scale statistical properties of the IGM targeting the EoR. The simulations use the AREPO-RT (Kannan et al., 2019) radiation hydrodynamics solver to solve the interaction between ionizing photons and gas and implement the ILLUSTRIS-TNG (Vogelsberger et al., 2014) galaxy formation and dust models to predict the properties of galaxies.

CHAPTER 2

The HI-halo mass relation at redshift $z \sim 1$ from the Minkowski functionals of 21 cm intensity maps

Benedetta Spina[1], Cristiano Porciani[1] & Carlo Schimd[2]
[1]*Argelander Institut für Astronomie , Auf dem Hügel 71, 53121 Bonn, Germany,*
[2]*Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France*
2021, MNRAS, 505, 3492, https://doi.org/10.1093/mnras/stab1555

2.1. Introduction

Galaxies rely on HI as a crucial cold-gas reservoir for star formation, making it essential to understand the variations in HI content across different galaxy populations and environments. The HI content within DM halos serves as a descriptive statistic, with the expectation that it primarily depends on halo mass, regulating gas accretion rates and depletion processes. Extensive studies using hydrodynamical simulations and semi-analytical models have explored the HI-halo mass relation (HIHMR) at different redshifts (Davé et al., 2013; Villaescusa-Navarro et al., 2014; Crain et al., 2017; Villaescusa-Navarro et al., 2018; Ando et al., 2019; Kim et al., 2017; Zoldan et al., 2017; Baugh et al., 2019; Spinelli et al., 2020), and observational constraints have been obtained from local to intermediate and higher redshifts (Giovanelli et al., 2005; Barnes & Haehnelt, 2014). The HIHMR is a crucial ingredient for predicting the signal in such experiments, and understanding it is essential for probing cosmology and the epoch of reionization. This chapter focuses on constraining the HIHMR using intensity maps of the redshifted 21 cm line in the post reionization Universe (Barkana & Loeb, 2005; Loeb & Wyithe, 2008; Switzer et al., 2013; Padmanabhan et al., 2015; Monsalve et al., 2019; Yue & Ferrara, 2019). The study presents a proof of concept by considering mock data at a redshift of approximately $z \simeq 1$

but emphasizes the potential for comprehensive analysis across various cosmic epochs. We utilize the geometry and topology of brightness-temperature isocontours, quantified by Minkowski functionals (MFs), to understand the spatial distribution of the HI signal (Gleser et al., 2006; Chen et al., 2019). Synthetic maps based on N-body simulations are generated, incorporating observational noise and foreground contamination. The Fisher information formalism is employed to forecast parameter constraints on the HIHMR. The chosen instrument is the Square Kilometre Array Phase 1 (SKA-1) MID observatory (Square Kilometre Array Cosmology Science Working Group et al., 2018), and the survey design considers a large survey area and a channel bandwidth of 2 MHz. Integration time and independent measurements of mean HI density are found to enhance the precision of parameter constraints. By combining data from multiple frequency channels, the study aims to provide comprehensive characterization of the HIHMR and contribute to our understanding of the epoch of reionization and the formation of cosmic structures.

I have conducted the analysis described in this paper, receiving valuable advice and input from the co-authors to enhance the methodology and the statistical counterpart. The following sections offer a comprehensive discussion of the scientific questions addressed by this project.

The Chapter is organized as follows. In the introduction section 2.1, we outline the state-of-the-art theoretical and observational background and our motivation to investigate the HIHMR and the HI content in general. In section 2.2 the HIHMR is presented, along with the procedure used to obtain mock intensity maps from N-body simulation. We continue in section 2.3 with the definition of the MFs, in particular for 2-dimensional fields, and we illustrate the their application for our purposes. In section 2.4 we present the Fisher matrix formalism we used to forecast the constraints on the HIHMR via SKA-1 MID results. The results of the Fisher matrix analysis are shown in section 2.5 and discussed in section 2.6. The main findings are outlined and commented in section 2.6.

2.2. Mock 21 cm intensity maps

To assess the potential of 21 cm intensity-mapping experiments, it is necessary to simulate the large-scale structure of the Universe and model the distribution of HI within DM halos. One approach involves combining high-resolution N-body simulations with a statistical framework to generate synthetic HI intensity maps. These maps need to account for instrumental noise, foreground contamination, and other observational artifacts. By constructing mock maps that incorporate the statistical properties of HI within halos and relevant noise sources, we can simulate expected observational data and evaluate the constraining power of 21 cm experiments for cosmology and astrophysics.

N-body simulation. For our analysis, we utilize the MultiDark-Planck (MDPL) simulation (Klypin et al., 2016), which evolves $N = 3840^3$ DM particles within a periodic

cubic box with a comoving side length of $L = 1\ h^{-1}$, Gpc.

HI-halo mass relation. We adopt the commonly used assumption that the mass of a DM halo is the primary factor influencing its HI content. Under this working hypothesis, it is reasonable to introduce the HIHMR $M_{\rm HI}(M)$ that gives the mean HI mass found in a DM halo of mass M (Pontzen et al., 2008; Bagla et al., 2010; Barnes & Haehnelt, 2014; Seehars et al., 2016; Padmanabhan & Refregier, 2017; Padmanabhan et al., 2017; Villaescusa-Navarro et al., 2018; Baugh et al., 2019). We adopt the following functional form,

$$M_{\rm HI}(M) = M_0 \left(\frac{M}{M_{\rm min}}\right)^{\alpha} \exp\left[-\left(\frac{M_{\rm min}}{M}\right)^{\gamma}\right] . \tag{2.1}$$

The HIHMR exhibits a scale-invariant power-law behavior with a positive slope, α, indicating that the HI mass scales with the halo mass, M, as $M_{\rm HI} \propto M^{\alpha}$. However, at small halo masses ($M \lesssim M_{\rm min}$), the HIHMR shows an exponential cutoff, resulting in a significant suppression of HI content. This cutoff reflects the inability of low-mass halos to self-shield from the UV background and inhibit gas cooling, limiting the formation and accumulation of HI (Rees, 1986; Efstathiou, 1992). The parameters γ and M_0 control the sharpness and normalization of the HIHMR, respectively. For our forecasts at redshift $z = 1$, we adopt fiducial values for the HIHMR parameters: $M_0 = 1.5 \times 10^{10} h^{-1} M_{\odot}$, $M_{\rm min} = 6.0 \times 10^{11} h^{-1} {\rm M}_{\odot}$, $\alpha = 0.53$, and $\gamma = 0.35$ (Spinelli et al., 2020). The HI content of a DM halo is influenced by factors beyond mass, including hydrodynamic processes, radiative and mechanical feedback, and halo spin. To account for these effects, we introduce scatter around the HIHMR parametrized by σ, by assuming a lognormal distribution for the HI mass $M_{\rm HI}$ at fixed halo mass M. We set $\sigma = 1$.

From halos to brightness temperature. To assign HI masses to each DM halo, a random sampling is performed from the corresponding lognormal distribution, incorporating scatter around the HIHMR. The HI mass of each halo is distributed in a Cartesian grid consisting of 210^3 cells, accounting for redshift-space distortions. The brightness temperature associated with the HI distribution is computed using the equation (1.51). See the upper left panel in Figure 2.1.

Frequency bandwidth and angular resolution. The receiver bandwidth is divided into frequency channels to obtain spectroscopic information about the observed sky. Narrower channels provide better spectral resolution but lower signal-to-noise ratios and vice versa. Synthetic maps are generated by averaging the brightness temperature over the channel bandwidth, resulting in non-overlapping intensity maps that represent integrated signals within specific frequency ranges. The instrumental blurring effect due to finite angular resolution is addressed by convolving the intensity maps with a Gaussian telescope beam. See Figure 2.1 for an illustration of the process. See the upper middle panel in Figure 2.1 for the resulting effects.

Thermal noise. The radio telescope's output is affected by thermal noise, typically approximated as Gaussian for single-dish observations. The rms noise fluctuation, determined by the radiometer equation, considers the system temperature, channel bandwidth and the integration time per pixel $t_{\rm pix}$. The system temperature includes contributions from various sources like spill-over, cosmic microwave background, galactic emission, and receiver noise. In the case of the 21 cm line at $z = 1$, the rms noise exceeds the HI signal's brightness temperature fluctuations, necessitating long integration times for reliable HI detection. Mapping the large-scale structure of the Universe requires multiple antennas, and the total observation time depends on the integration time per pixel, the number of antennas and feedhorns, and the solid angle coverage per pixel. See the center middle panel and the center right panel in Figure 2.1.

Foreground removal. Mapping the large-scale structure of the Universe in 21 cm encounters difficulties due to foreground contamination, where foreground emissions like synchrotron and extragalactic point sources can be much brighter than the HI signal (e.g. Shaver et al., 1999; Di Matteo et al., 2002; Oh & Mack, 2003; Zaldarriaga et al., 2004). However, foregrounds exhibit frequency-dependent smooth variations, while the 21 cm signal is expected to lack correlation (Wolz et al., 2015). This distinction allows for foreground separation techniques that statistically differentiate foregrounds from the 21 cm signal based on their spectral characteristics. In this approach, foreground subtraction eliminates Fourier modes of the brightness temperature distribution below a specific comoving radial wavenumber cutoff, typically denoted as $k_{\parallel}^{\rm FG}$ (Alonso et al., 2017; Cunnington et al., 2019). It effectively eliminates the monopole and largest-scale fluctuations, setting the mean value of the brightness temperature distribution over the entire simulation box to zero. See the center left panel in Figure 2.1.

Smoothed maps. The presence of thermal noise adds additional fluctuations to the smoothed 21 cm signal, making it challenging to detect and analyze the faint cosmological signal. To mitigate the impact of fine-grained noise on the measurement of the MFs, the observed map undergoes additional smoothing using a two-dimensional Gaussian filter with a width matching the telescope beam. This smoothing operation reduces the impact of noise on the MF measurements by averaging pixel-level fluctuations over a larger scale. See the lower left panel in Figure 2.1.

2.3. Minkowski Functionals

Basics. The Minkowski Functionals (Minkowski, 1903) are morphological descriptors that quantify the size and connectivity of subsets in a space, such as digital images or cosmological data. In d dimensions, there are $d + 1$ MFs denoted as $V_0, V_1, \ldots, V_d$, which capture different aspects of subset morphology. In cosmology, MFs are used to characterize the large-scale structure of the Universe, detect topological features,

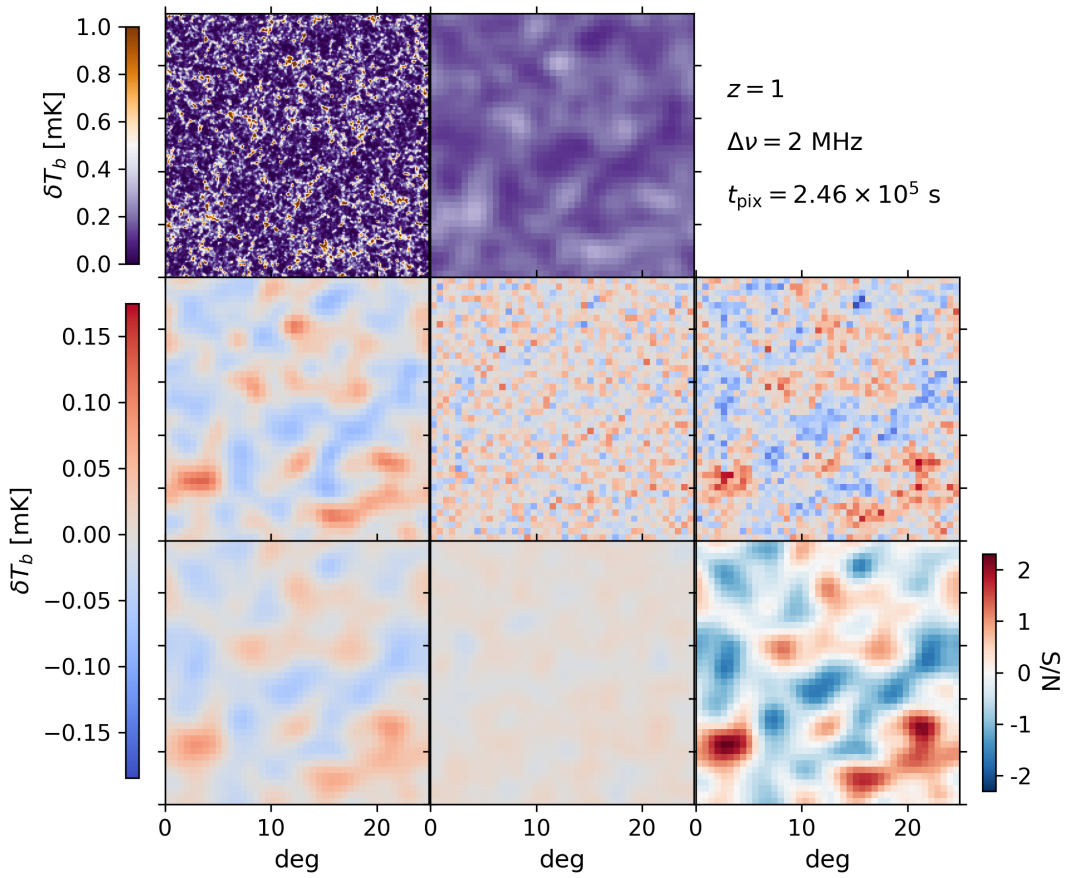

Figure 2.1. – The process of building mock 21 cm intensity maps. The top-left panel shows the original brightness-temperature distribution. To simulate the finite angular resolution of radio telescopes, the brightness-temperature distribution is smoothed (top-right panel). During the foreground removal process the mean temperature of the signal is lost (center-left panel). Thermal noise is generated at the pixel level and added to the signal (center-right panel). To enhance the signal-to-noise ratio, a smoothing technique is applied as a data processing step (bottom-left panel). The bottom-central panel displays the smoothed noise map, and the bottom-right panel shows the signal-to-noise map. From Spina et al. (2021).

and investigate statistical properties like non-Gaussianity. The application of MFs in cosmology was pioneered for analyzing the large-scale distribution of galaxies (Mecke et al., 1994). They have been used to study temperature anisotropies in the CMB, revealing information about primordial fluctuations and cosmic structure formation (e.g. Schmalzing & Gorski, 1998; Novikov et al., 1999). MFs have also been applied to study cosmic reionization, shedding light on the ionization history and early Universe evolution (Gleser et al., 2006; Chen et al., 2019).

Application to 21 cm maps. We focus on computing the MFs of two-dimensional maps of the 21 cm signal rather than analyzing the full three-dimensional data cube,

facilitating the estimation of the covariance matrix. The analysis is performed on maps corresponding to a single frequency channel without constructing a mock light cone. The maps are obtained at $z = 1$ from the MDPL simulation. The corresponding covariance matrix, which characterizes statistical uncertainties and correlations, is also derived by considering 105 independent mock 21 cm maps. The analysis is based on the excursion set formalism, representing the regions where the brightness temperature exceeds a specified threshold. The MFs of the excursion set, namely V_0, V_1, and V_2, are computed. V_0 represents the surface area, V_1 captures the perimeter or boundary, and V_2 is equivalent to the Euler characteristic, which provides insights into the connectivity and structure. The publicly available code MINKFNCTS2D[1] is used to compute the MFs. The MFs are analyzed for different threshold temperatures and observation times per pixel. The presence of thermal noise introduces smaller-scale structures, altering the values of the MFs. It reduces the range of V_0, shifts V_1 towards larger values, and increases the extreme positive and negative values of V_2. The sensitivity of the MFs to the HIHMR is explored by varying the parameters of the relation. The MFs exhibit distinct changes in shape and extremal points with parameter variations, highlighting their sensitivity to the underlying HI-to-halo mass relation. However, variations in the parameter σ have minimal impact on the MFs.

Selected observables. When using the MFs for measurements, a limited number of threshold values are chosen due to practical constraints in estimating the covariance matrix. After considering many combinations, two threshold values are chosen for V_0, two for V_1, and five for V_2, resulting in nearly optimal constraints on the HIHMR. All measurements are combined into a single seven-dimensional data vector $\mathbf{M}$, which represents the selected threshold values of the MFs. The mean HI density, $\bar{\rho}_{\rm HI}(z)$, determined by the parameters in equation (1.51), affects the overall level of the fluctuations, δT_b. However, due to foreground subtraction, the mean brightness temperature of the 21 cm fluctuations cannot be directly measured. This missing information may weaken the constraints that the MFs can impose on the HIHMR. Taking this into account, the measurements of the MFs are combined with observational constraints on the cosmic abundance of HI, parameterized as $\Omega_{\rm HI}(z)$. The constraint $\Omega_{\rm HI}(z = 1) = (6.1 \pm 0.4) \times 10^{-4}$ is used in this study to set constraints on the parameters of the HIHMR (Rao et al., 2006; Crighton et al., 2015; Rao et al., 2017; Hu et al., 2019). See Figure 2.2.

2.4. Fisher matrix

Minkowski functionals. In this study, the MFs measurements are fitted with a theoretical model $\mathbf{M}_{\rm mod}(\boldsymbol{\theta})$ that depends on adjustable parameters $\boldsymbol{\theta}$. The Fisher matrix $F_{i,j}$ is

[1] https://github.com/moutazhaq/minkfncts2d

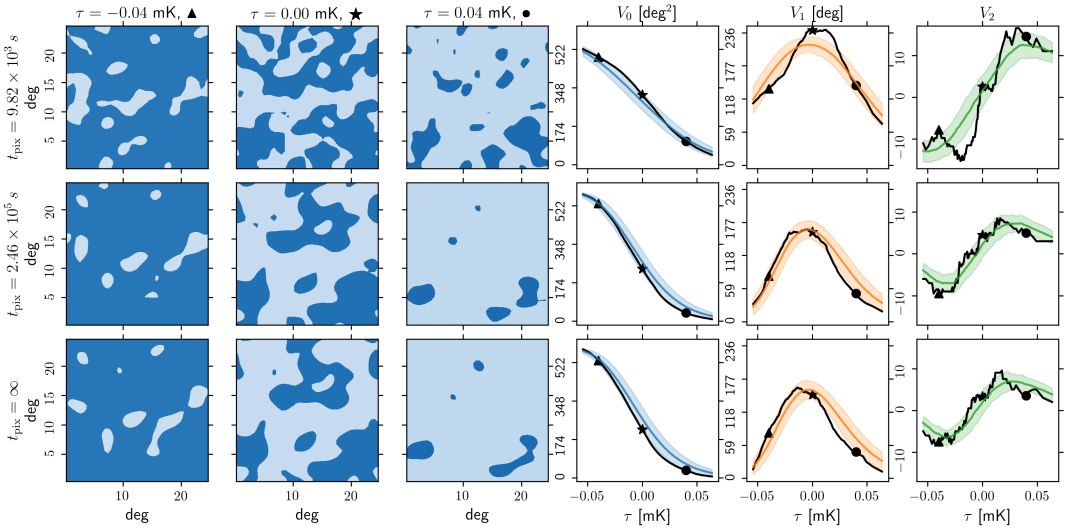

Figure 2.2. – Comparison of intensity maps and MFs for different integration times and brightness-temperature thresholds. Three intensity maps are considered, obtained from the same sky region as shown in Figure 2.1, but with varying integration times per pixel. The integration times, from top to bottom, are $t_{\rm pix} = 9.82 \times 10^3$ s, 2.46×10^5 s, and ∞. The left set of figures displays regions above (dark blue) and below (light blue) three brightness-temperature thresholds ($\tau = -0.04$ mK, 0 mK, 0.04 mK) for each integration time. The right set of figures presents the corresponding MFs for the selected threshold values, represented by triangle, star, and circle symbols. The MFs are plotted as a function of τ using black solid lines. As a reference, the mean values of the MFs obtained from the independent intensity maps are shown as colored solid lines, while the shaded regions represent the rms scatter. From Spina et al. (2021).

computed using the partial derivatives of $\mathbf{M}_{\rm mod}$ with respect to the model parameters and the inverse of the covariance matrix C as

$$F_{ij} = \frac{\partial\, \mathbf{M}_{\rm mod}^T}{\partial \theta_i} \cdot \mathsf{C}^{-1} \cdot \frac{\partial\, \mathbf{M}_{\rm mod}}{\partial \theta_j} \,. \tag{2.2}$$

Constructing a theoretical model for the MFs of the 21 cm signal poses challenges. The expected values of the mock observations, averaged over sample variance and thermal noise, are utilized as $\mathbf{M}_{\rm mod}(\boldsymbol{\theta})$. The sample covariance matrix ($\hat{\mathsf{C}}$) of the MFs is approximated, where the quantities are evaluated with the fiducial set of model parameters. The impact of thermal noise is considered by generating multiple realizations of the noise and combining them with the 21 cm signal for different values of $t_{\rm pix}$.

Scaling with the survey area. The study has focused on a region of the sky with a linear transverse size matching the comoving side of the MDPL simulation box. It has been found that the data covariance matrix scales proportionally to the survey area. To extend the analysis to larger survey areas, it is sufficient to correct the covariance matrix by a multiplicative factor denoted as ζ. This factor is given by the square root of the ratio of the survey area ($\Omega_{\rm surv}$) to the area of the MDPL simulation box ($\Omega_{\rm MDPL}$).

Table 2.1. – Fiducial values and forecast uncertainties for the model parameters that regulate the HIHMR. The top section gives the results from the Fisher-information analysis for one channel at $z = 1$. The bottom sections refer to fits of either second-order polynomials to the measurements in the data cubes for $0.5 < z < 1.5$. The optimistic case (opt) treats all channels as independent data points while the pessimistic one (pes) considers only one channel every five. Adapted from Spina et al. (2021).

	α	M_0 $[10^{10}h^{-1}\,\mathrm{M}_\odot]$	$M_{\rm min}$ $[10^{11}h^{-1}\,\mathrm{M}_\odot]$	γ
Fiducial value θ	0.53	1.50	6.00	0.35
$\Delta\theta$ single channel	0.05	0.15	0.23	0.11
	θ_0	θ_0	θ_0	θ_0
Fiducial value θ	0.53	1.50	6.00	0.35
$\Delta\theta$ data cube (opt)	0.01	0.02	0.02	0.01
$\Delta\theta$ data cube (pes)	0.01	0.03	0.05	0.02

Constraints from $\Omega_{\rm HI}$. To compute the constraints derived from the measurements of $\Omega_{\rm HI}$, a second Fisher matrix G_{ij} is utilized. The combined constraints obtained from both the MFs and $\Omega_{\rm HI}$ are determined using the matrix $(\mathsf{F}+\mathsf{G})^{-1}$, where F represents the Fisher matrix derived from the MF measurements and G is the Fisher matrix computed from the HI cosmic abundance measurements.

2.5. Results

In this section, the main results of the study are presented, which include a Fisher forecast for the HIHMR. The forecast is based on a future wide HI intensity-mapping survey conducted with the SKA-1 MID (Square Kilometre Array Phase 1 Mid-frequency) observatory. This survey is expected to cover an area of approximately $\Omega_{\rm surv} \simeq 20,000$ square degrees (corresponding to a scaling factor $\zeta \approx 5.8$) and a frequency range of $350 < \nu < 1050$ MHz (equivalent to $0.35 < z < 3.06$ in redshift).

Constraints from individual channels. The Fisher forecasts for the HIHMR are presented in Figure 2.3, showing the relative marginal uncertainty (68.3% credibility interval) for each model parameter as a function of $t_{\rm pix}$ (pixel integration time). The orange lines and shaded regions represent the mean and scatter of the results over 30 noise realizations when combining the MFs with $\Omega_{\rm HI}$ data. The blue lines show the mean uncertainty when considering only the MFs. It is observed that combining the MFs and $\Omega_{\rm HI}$ constrains all parameters of the mean HIHMR to better than 10% for large $t_{\rm pix}$ values. The constraints deteriorate as $t_{\rm pix}$ decreases, with significant degradation occurring when $t_{\rm pix}$ is much smaller than 10^4 s.

Figure 2.4 demonstrates that the MFs and $\Omega_{\rm HI}$ provide informative constraints on the halo occupation properties of HI. The posterior distribution of the mean HIHMR shows little scatter around the fiducial model.

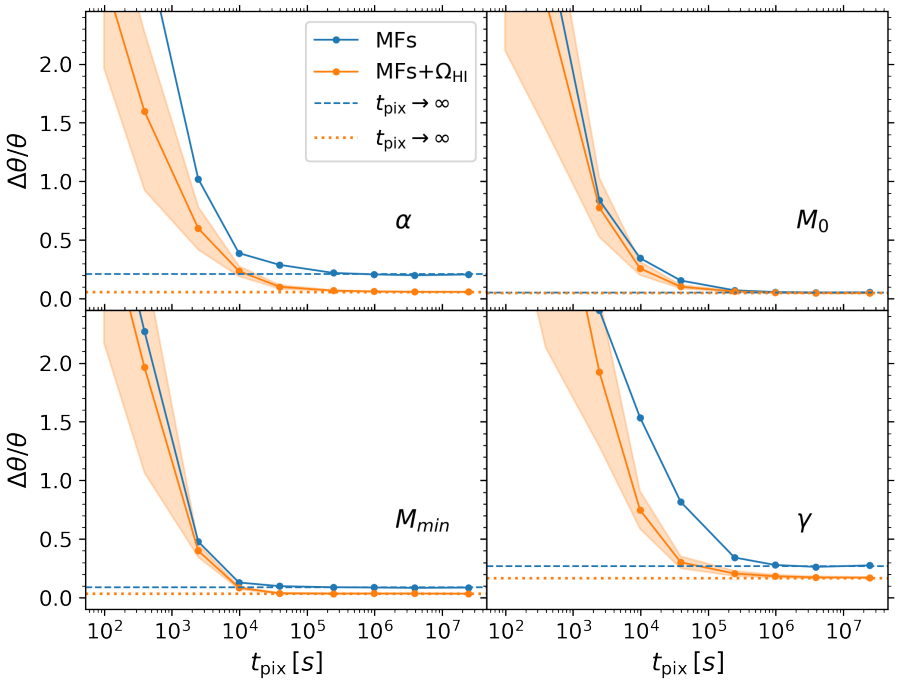

Figure 2.3. – The relative marginalised uncertainty (68.3% credibility interval) for each model parameter is shown as a function of t_{pix}. The continuous lines and shaded regions represent the mean and rms scatter of the results over 30 noise realizations for MFs alone (in blue) and when combining with the Ω_{HI} data (in orange). The dotted and dashed lines indicate the results obtained in the absence of thermal noise. From Spina et al. (2021).

Constraints from the full data cube. The study now shifts focus to utilizing the full data cube provided by radio observations. Instead of considering a single frequency channel, the measurement of the MFs will be repeated in each frequency channel, allowing for the fitting of the redshift evolution of the model parameters. It is expected that the parameters of the HIHMR should vary smoothly with redshift. To demonstrate the power of this technique, the data at $z \simeq 1$ from table 1 of Villaescusa-Navarro et al. (2018) is synthesized in terms of second-order Taylor expansions for the model parameters in the redshift range $0.5 < z < 1.5$. Table 1 in the paper presents the results obtained assuming independent data points for consecutive channels. The analysis is repeated by considering only one channel every five. Although the uncertainty on the model parameters slightly increases in this case, there is still a significant improvement compared to considering a single channel. Comparing the results obtained for $z = 1$ from the multi-channel fits to those presented in a previous section, an improvement in the error bars by up to a factor of ten is observed.

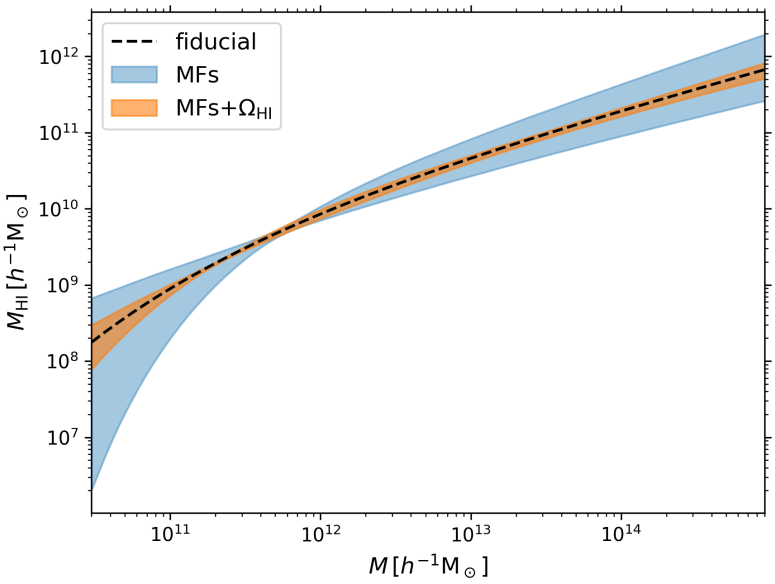

Figure 2.4. – The central 68.3% credibility interval of the mean HIHMR at $z = 1$ is obtained by fitting the MFs (light blue) and combining them with estimates of $\Omega_{\rm HI}$ (orange) in a single frequency channel at $z = 1$, assuming an integration time of $t_{\rm pix} = 3.9 \times 10^4$ s. The dashed curve represents the fiducial relation. From Spina et al. (2021).

2.6. Discussion and conclusion

While power spectra contain the full statistical information for Gaussian random fields, the MFs provide additional information in the case of non-Gaussian fields. It is important to determine whether the fluctuations in the 21-cm brightness temperature distribution deviate significantly from a Gaussian random field, especially considering the large beam size of the SKA-1 MID telescopes. We compute the probability density function (PDF) of the brightness temperature and we show that it is asymmetric and has heavier tails compared to a Gaussian distribution, indicating departures from Gaussianity. The skewness and kurtosis values further support this non-Gaussian behavior. We also compare the measured MFs with the ones obtained for a Gaussian random field. While the surface area covered by the excursion set shows little difference from the Gaussian case, noticeable deviations are observed for the perimeter of the boundary and the Euler characteristic, particularly at extreme threshold parameter values.

The issue of whether the information encoded in the MFs on the large scales probed by SKA-1 MID is sufficient to constrain all the parameters influencing the mean HIHMR is addressed. The assumption of the two-halo regime in the halo model suggests that the data collapses into the linear bias parameter of HI with respect to mass (see Cooray & Sheth, 2002, for a review). However, this assumption only holds to a first approximation, and non-linear corrections are necessary to accurately model the spatial distribution of DM halos. The results indicate that similar corrections are needed to describe the non-

Gaussian features of the HI distribution (see e.g. Desjacques et al., 2018, for a review). Evidence from other studies, such as the bispectrum of 21-cm intensity maps, further supports this need for non-linear corrections (Cunnington et al., 2021). While some level of degeneracy between model parameters is present in the results, combining independent measurements of the HI abundance with the MFs helps break this degeneracy.

The main findings of the study can be summarized as follows:

- Long integration times per pixel, $t_{\rm pix}$, are required to overcome thermal noise and achieve the necessary sensitivity for imaging the HI distribution. Using the MFs of the 21 cm intensity map in a single frequency channel, the parameters of the HIHMR can be measured with a signal-to-noise ratio of one at $t_{\rm pix} \simeq 9.8 \times 10^3$ s. Optimal error bars are obtained with $t_{\rm pix} \simeq$ a few $\times 10^4$ s, corresponding to a total observing time of 4-5 days.

- The loss of information on the mean HI density, $\Omega_{\rm HI}$, during foreground removal can be restored by combining the MFs with independent measurements of $\Omega_{\rm HI}$. This addition slightly improves the constraints on some parameters of the HIHMR, particularly the slope α.

- Combining MF measurements in different frequency channels allows for exquisite constraints on the redshift evolution of the HIHMR, both for the mean and the scatter.

The study uses the SKA-1 MID dark-energy survey as an example of what can be achieved with forthcoming facilities. Higher angular resolution observations with SKA precursors like the Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX, Newburgh et al., 2016b) and the Canadian Hydrogen Intensity Mapping Experiment (CHIME, Bandura et al., 2014), as well as other proposed facilities, have the potential to probe smaller length scales where the HI-to-mass bias relation is more non-linear. With sufficiently long integration times, these experiments may partially remove parameter degeneracies and provide tighter constraints on the HIHMR. Future work will investigate this possibility further.

CHAPTER 3

Detecting signature of Helium reionization through line-intensity mapping

Benedetta Spina[1] & Cristiano Porciani[1]
[1]*Argelander Institut für Astronomie , Auf dem Hügel 71, 53121 Bonn, Germany*
In preparation

3.1. Introduction

During the early stages of cosmic history, the IGM experienced two significant transformations. Initially, there was a transition from neutral hydrogen to ionized hydrogen (HII), which has been extensively reviewed (see Barkana & Loeb, 2005). Subsequently, there was a transition from single-ionized helium to doubly-ionized helium (HeIII), also discussed in various comprehensive reviews (see Furlanetto & Oh, 2008b; Bagla & Loeb, 2009; McQuinn & Switzer, 2009, for reviews). These changes in the ionization states of hydrogen and helium had profound implications for the evolution of the intergalactic medium.

The process of helium reionization is intricately linked to the characteristics and abundance of the early QSO population. While photons emitted by star-forming galaxies during high redshifts play a crucial role in ionizing neutral hydrogen and helium, they do not contribute significantly to the reionization of helium. This is because the ionization of helium requires photons with higher energies, such as those emitted by AGNs. Therefore, in order to gain insights into the activity of QSOs and the formation of galaxies, it is essential to accurately determine the abundance of single-ionized helium throughout cosmic time.

Helium reionization can be investigated through the study of the hyperfine transition at a wavelength of $\lambda_{\rm HeII} = 3.46$ cm in single-ionized helium atoms. Similar to the 21

cm line for neutral hydrogen atoms, this emission line provides valuable insights into the process of helium reionization. The $\lambda_{\rm HeII}$ line is stronger and more spread out compared to other emission lines like [CII] or CO, but it is fainter than the 21 cm line. However, it is less susceptible to contamination from synchrotron radiation and interference from the terrestrial ionosphere, making it a favorable probe for studying helium reionization.

Several upcoming radio surveys aim to observe and study the EoR and its main drivers. These surveys utilize advanced radio telescopes with specific capabilities and differentiate in several aspects with respect to e.g. galaxy surveys. Galaxy surveys primarily involve mapping and characterizing the distribution of galaxies in the Universe. These surveys employ optical or infrared observations to detect and study the light emitted by galaxies. By mapping the positions and properties of galaxies, they provide valuable insights into the large-scale structure of the Universe, the clustering of galaxies, the expansion rate of the cosmos, and the evolutionary processes of galaxies over cosmic time. Galaxy surveys play a crucial role in advancing our understanding of cosmology, galaxy formation, and the nature of DM. On the other hand, radio surveys focus on detecting and studying radio emissions from celestial objects. Radio surveys often target specific frequency ranges and utilize advanced techniques to separate desired signals from background noise and interference. We list some of the upcoming and ongoing radio surveys in Section 3.2

This study focuses on evaluating the feasibility of detecting the fluctuations in the 3.46 cm signal during helium reionization using SKA-1 MID, DSA-2000 and PUMA in single-dish mode (Battye et al., 2013). The objective is to place constraints on the sources responsible for helium reionization within a reasonable observation time. Currently, the 3.46 cm signal has been detected in local HII regions and planetary nebulae using radio telescopes like GBT (Rood et al., 1979, 1984; Balser & Bania, 2018) and VLA (Balser et al., 2006). However, these measurements alone are insufficient to fully constrain the theoretical models describing helium reionization. The upcoming observations with future radio surveys are anticipated to provide groundbreaking results in this field. The consensus among researchers is that the reionization of the Universe took place around a redshift of $z \sim 3.2$. This understanding is supported by various lines of evidence, such as the evolution of the IGM temperature derived from HI Lyα absorption features (Theuns et al., 2002; Becker et al., 2011), observations of quasars (QSOs) (Richards et al., 2006), and the properties of the Lyα forest (Jakobsen et al., 1994). However, recent studies by Giallongo et al. (2015) and subsequent works (Madau & Haardt, 2015; Chardin et al., 2017) suggest that it is important to consider alternative scenarios that involve an increased population of faint QSOs during the early stages of the Universe. These additional faint QSOs may have a significant impact on the photon budget during the EoR and could contribute to the second ionization of helium.

Theoretical studies have explored different approaches to trace helium reionization on small scales, such as investigating ionized regions near accreting black holes (Vasiliev et al., 2019) or utilizing fast radio bursts (Linder, 2020). On larger scales, recent research has focused on the detectability of the HeII hyperfine transition signal (Takeuchi et al., 2014) and its power spectrum (Khullar et al., 2020), including the application of kinetic

Sunyaev-Zeldovich tomography (Hotinli et al., 2022). These investigations aim to enhance our understanding of the process of helium reionization and explore different observational techniques to study it. In contrast to previous estimates by Khullar et al. (2020), which considered a late reionization scenario (Eide et al., 2020) using an interferometric setup, we investigate two scenarios: a standard late-reionization scenario (referred to as Late) and an early-reionization scenario (referred to as Early). In the Early scenario, the reionization of both hydrogen and helium is attributed to a population of high-redshift QSOs, as proposed by Giallongo et al. (2015). By considering these two scenarios, we aim to explore different reionization models and their impact on the detectability of the HeII hyperfine transition signal. In our study, we employ hydrodynamical simulations that have been post-processed to incorporate radiative-transfer equations. By simulating data cubes resembling those expected from the surveys considered, we account for instrumental effects such as angular resolution and thermal noise. To mitigate foreground contamination, we analyze the 3D power spectrum of the simulated signal and assess its signal-to-noise ratio. Our findings present innovative and promising outcomes regarding the detectability of fluctuations in the 3.46 cm signal within a feasible observational timeframe.

The structure of this work is as follows. Starting from Section 3.1 we outline the state-of-the-art theoretical and observational background and our motivation to investigate helium reionization. In Section 3.2 we list the main surveys we consider in this work. In Section 3.3, we provide a detailed description of the hydrodynamical simulations and outline the procedure for generating mock data cubes. Section 3.4 introduces the definition of the power spectrum. We outline the main systematic effects when performing observations in Section 3.5 and we discuss how to account for them. We present and comment our results in Section 3.7. Section 3.8 summarizes our main findings, while Section 3.9 offers a perspective of future challenges and steps.

3.2. Radio surveys

We list here some notable examples of radio surveys.

- Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX): HIRAX is a radio interferometer designed to study the EoR by mapping the redshifted 21 cm line of neutral hydrogen. It consists of an array of compact, wide-field radio antennas located in South Africa (Newburgh et al., 2016b).

- Canadian Hydrogen Intensity Mapping Experiment (CHIME): CHIME is a radio interferometric telescope located in Canada. It is specifically designed for studying cosmology and the EoR through intensity mapping of neutral hydrogen. CHIME covers a large frequency range and provides a wide field of view (Bandura et al., 2014).

- LOw Frequency ARray (LOFAR): LOFAR is a radio telescope array spread across several European countries. It operates at low frequencies and has excellent sensitiv-

ity for observing the EoR. LOFAR employs a combination of phased array stations and dipole antennas to cover a wide range of frequencies (van Haarlem et al., 2013).

- Square Kilometre Array Phase 1 MID (SKA-1 MID): SKA-1 MID is a precursor to the full Square Kilometre Array (SKA) project. It will consist of an array of mid-frequency radio telescopes located in Australia and South Africa. SKA-1 MID will have a large collecting area and high sensitivity, making it well-suited for studying the EoR and other cosmological phenomena (Square Kilometre Array Cosmology Science Working Group et al., 2018).

- The Dark Sector Array 2000 (DSA-2000) is a proposed radio survey telescope with 2000x5m dishes in Nevada. It aims to image the entire sky repeatedly over sixteen epochs, detecting over one billion radio sources and operating as a dedicated survey telescope (Hallinan et al., 2019).

- Packed Ultra-wideband Mapping Array (PUMA) is a proposed radio telescope designed for 21 cm intensity mapping in the post-reionization era, as well as other science goals like Fast Radio Bursts (FRBs), pulsar monitoring, and multi-messenger observations of transients. It operates in the frequency range of 200 MHz to 1100 MHz. In the first phase, it consists of 5000x6m antennas (Slosar et al., 2019) .

For the purposes of this work, we will refer only to SKA-1 MID, DSA-2000 and PUMA. We list in Table 3.1 the main technical features of the telescopes for the detection of the 3.46 cm emission line. In particular, D is the diameter of the antennas, $N_{\rm dish}$ the number of antennas for the survey, $f_{\rm detec}$ the detected frequency range for the HeII signal accessible by the telescope, $T_{\rm sys}(\nu)$ the system temperature[1].

	SKA-1 MID	DSA-2000	PUMA
D [m]	15	5	6
$N_{\rm dish}$	197	2000	5000
$f_{\rm detec}$ [GHz]	0.35-15.3	0.7 - 2.0	0.2 - 1.1
$T_{\rm sys}(\nu)$ [K]	$\sim 25 - 80$	~ 25	~ 25

Table 3.1. – Technical parameters of the radio surveys considered in this work, for the purposes of HeII detection.

3.3. Mock 3.5 cm data cube

In this section, we will provide a comprehensive description of the simulated mock HeII data cubes. We will start by explaining how we obtain the HeII fraction from the

[1]The system temperature will be discussed further in Section 3.5.2.

hydrodynamical simulations. Additionally, we will outline the methods we employ to account for the various sources of noise in the data.

3.3.1. Simulations

To generate the simulated HeII data cubes, we utilize two sets of hydrodynamical simulations of the IGM performed with the Ramses code (Teyssier, 2002). These simulations are then post-processed using the Radamesh code (Cantalupo & Porciani, 2011). The simulated volume corresponds to a cubic box with a comoving side length of $L = 100\,h^{-1}$Mpc, assuming a flat ΛCDM cosmology. During the radiative-transfer calculations, the local baryonic overdensity $\Delta_\mathrm{b}(\mathbf{x})$, defined as the ratio of the baryonic density $\rho_\mathrm{b}(\mathbf{x})$ to the mean baryonic density $\langle\rho_\mathrm{b}\rangle$, is held constant. However, the mean density $\langle\rho_\mathrm{b}\rangle$ scales as $(1+z)^3$ with redshift z. This scaling accounts for the expansion of the Universe and ensures that the mean density evolves consistently with the cosmological model.

In the first suite of simulations (Compostella et al., 2014), helium reionization is primarily driven by AGNs, which become active at redshift $z = 5$. By employing a probabilistic algorithm, the DM halos are assigned AGNs in a manner that reproduces the measurements by Glikman et al. (2011). Two different approximations are used to model the evolution of AGN emissivity, based on the results from Haardt & Madau (2012). However, no significant systematic differences in the relevant observables were found between the two approximations. As a result of these simulations, helium reionization concludes at a relatively late redshift of approximately $z \sim 3$. This corresponds to a scenario of late reionization. The suite comprises a total of six simulation runs, allowing for a comprehensive analysis of the effects and variations in the simulated observables.

The second suite of simulations (Garaldi et al., 2019) adopts a different approach by incorporating the QSO luminosity function derived from Giallongo et al. (2015). This luminosity function implies a significant population of QSOs at high redshift, resulting in an early-reionization model. To calibrate the luminosity of the QSOs, they are placed at the centers of the simulated DM halos. In this scenario, the QSOs contribute sufficient energy to ionize both neutral hydrogen, single-ionized helium, and doubly-ionized helium. Consequently, the reionization of helium completes much earlier compared to standard models, nearly coinciding with the end of hydrogen reionization at a redshift of approximately $z \approx 5$. The suite consists of four simulation runs, providing further insights into the effects and characteristics of early helium reionization.

Figure 3.1 illustrates the evolution of the helium fractions for the Late and Early models. The blue curve represents the HeI fraction, the orange curve represents the HeII fraction, and the green curve represents the HeIII fraction. The continuous curves correspond to the Late model, while the dashed curves correspond to the Early model. The shaded regions around the curves indicate the standard deviation among the simulation runs. In the Late model, the HeI fraction remains at zero throughout the entire redshift range considered. On the other hand, in the Early model, the first and second ionization of helium occur simultaneously. As a result, the majority of HeII is immediately reionized, causing the

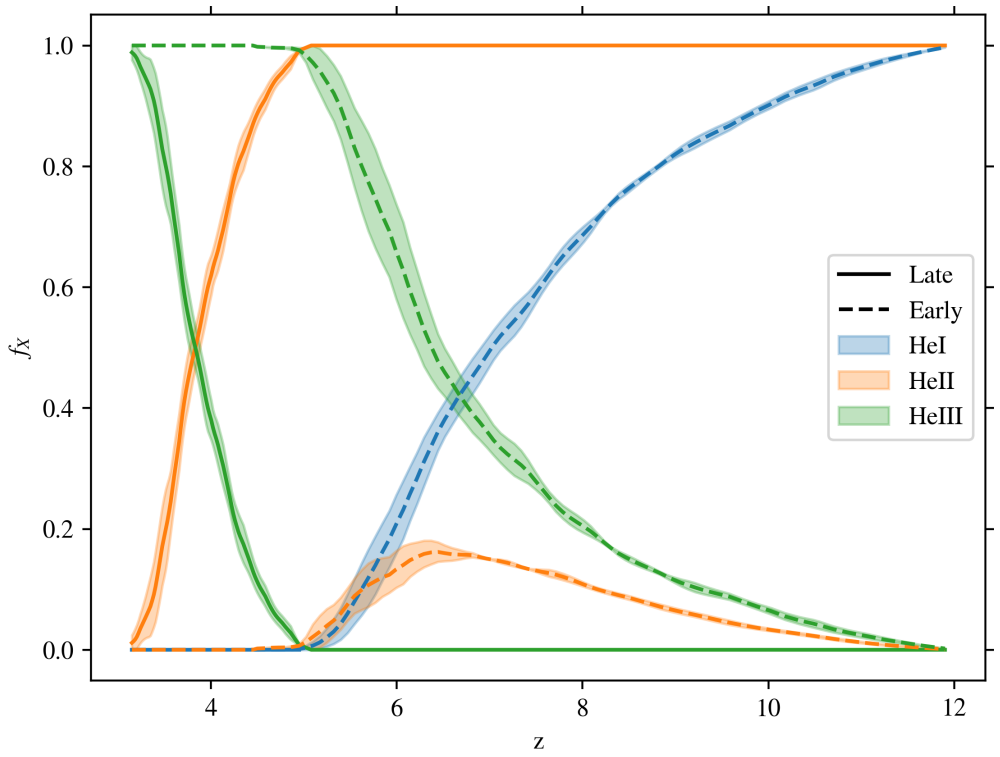

Figure 3.1. – The evolution of the HeI (blue lines), HeII (orange lines), and HeIII (green lines) fractions for the Late reionization (continuous lines) and Early reionization (dashed lines) models as a function of redshift. The figure provides an overview of the evolution of helium ionization states, demonstrating that in the Late reionization model, helium remains predominantly in the single-ionized state (HeII), while in the Early reionization model, both first and second ionization of helium occur simultaneously, resulting in a transient increase in the HeII fraction.

HeII fraction to gradually increase until around $z \sim 6.5$, after which it rapidly decreases. By the end of both simulations, helium is completely ionized in the intergalactic medium.

3.3.2. Brightness temperature

As already shown in Section 1.6.2 the differential brightness temperature δT_b of the 3.46 cm hyperfine transition of HeII relative to the CMB can be expressed as:

$$\delta T_b(z, \mathbf{r}) = \frac{T_S(z, \mathbf{r}) - T_{\rm CMB}(z, \mathbf{r})}{1+z} \left[1 - e^{-\tau(z,\mathbf{r})}\right], \tag{3.1}$$

where $T_S(z, \mathbf{r})$ represents the spin temperature associated with the spin-flip of electrons in HeII atoms at redshift z and position $\mathbf{r}$. Moreover, $T_{\rm CMB}(z, \mathbf{r})$ denotes the CMB temperature, and $\tau(z, \mathbf{r})$ represents the optical depth of the IGM. This equation describes the deviation of the 3.46 cm signal from the CMB due to the interaction between HeII atoms and radiation fields. In the optically thin regime, where the optical depth τ is

much greater than 1, the expression for the differential brightness temperature simplifies to (Furlanetto et al., 2006):

$$\begin{aligned}\delta T_b(z,\mathbf{r}) &\simeq \frac{T_S^X(z,\mathbf{r}) - T_{\rm CMB}(z,\mathbf{r})}{1+z}\tau(z,\mathbf{r}) \simeq \\ &\simeq 0.53\, f_{\rm HeII}(z,\mathbf{r})\Delta_b(1+z)^2\frac{H_0}{H(z)}\times \\ &\quad\times\left[1-\frac{T_{\rm CMB}(z,\mathbf{r})}{T_S(z,\mathbf{r})}\right]\left[\frac{H(z)/(1+z)}{{\rm d}v_\parallel/{\rm d}r_\parallel}\right]\,\mu{\rm K}.\end{aligned} \tag{3.2}$$

In our analysis, we consider $f_{\rm HeII}$ as the local fraction of HeII and $\frac{{\rm d}v_\parallel}{{\rm d}r_\parallel}$ as the velocity gradient along the line-of-sight, accounting for both the Hubble expansion and the peculiar velocity. We make the assumption that the ratio $\frac{H(z)/(1+z)}{{\rm d}v_\parallel/{\rm d}r_\parallel}$ is approximately equal to 1. Additionally, we assume that the spin temperature T_S is much larger than the CMB temperature $T_{\rm CMB}$, resulting in an approximation of $1-\frac{T_{\rm CMB}(z)}{T_S(z)} \sim 1$. These assumptions simplify the expression for the differential brightness temperature and allow us to focus on the dominant factors influencing the signal.

3.4. Power spectrum

The differential brightness temperature in the Fourier space is represented as $\delta T_b(\mathbf{q})$, where $\mathbf{q}$ denotes wavevectors with Cartesian components that are integer multiples of the fundamental frequency $k_{\rm F}$. For this particular analysis, the fundamental frequency is $k_{\rm F} = 2\pi/L = 0.06\,h{\rm Mpc}^{-1}$. The power spectrum is estimated using the following formula:

$$\hat{P}(k) = \frac{1}{N_q}\sum_{\mathbf{q}\in k}|\delta T_b(\mathbf{q})|^2, \tag{3.3}$$

where N_q represents the number of wavevectors $\mathbf{q}$ within the k-bin of width δk. The sum is taken over the wavevectors $\mathbf{q}$ that satisfy the condition $k - \frac{\delta k}{2} \leq |\mathbf{q}| \leq k + \frac{\delta k}{2}$. For a comprehensive description of the power spectrum, we refer to Section 1.5. To simplify the notation, we will indicate the estimator for the power spectrum as $P = \hat{P}$.

The power spectrum of the signal is presented for the Late model (Figure 3.3) and the Early model (Figure 3.2). One notable distinction between the models lies in the redshift evolution of the signal. On small scales, the power spectra exhibit a consistent scaling with the matter power spectrum, showing a decrease with redshift for both models. However, the amplitude at a given redshift is significantly higher in the standard model compared to the alternative model. This discrepancy arises from the different reionization time-scales in the two models.

Additionally, the presence of HeII bubbles created by the AGNs and QSOs has a significant impact on the large-scale behavior of the power spectrum. These bubbles introduce

non-Gaussianity and alter the topology of the HeII field. As the reionization process progresses, the bubbles expand, merging with each other and filling the intergalactic medium. The expansion of the HeII bubbles leads to a distinct pattern in the power spectrum on large scales. As the bubbles grow, they create regions of enhanced HeII fraction surrounded by regions of lower HeII fraction. This spatial distribution introduces fluctuations in the signal, resulting in deviations from the Gaussian distribution expected for the matter component alone. Moreover, the expansion of the bubbles alters the topology of the HeII field. Initially, the bubbles are small and isolated, but as they expand, they start to overlap and form interconnected structures. This evolution modifies the connectivity and geometry of the HeII regions, creating a complex network of interconnected voids and ionized regions. These changes in the topology of the HeII field further contribute to the distinct features observed in the power spectrum on large scales. Overall, the expansion of HeII bubbles driven by the AGNs and QSOs introduces non-Gaussianity and modifies the topology of the HeII field, leading to unique features in the power spectrum on large scales. This highlights the importance of considering the reionization process and its impact on the spatial distribution of the HeII fraction when analyzing the power spectrum measurements.

This behavior is shown in the brightness temperature panels in Figure 3.2 and Figure 3.3. In the Late model, where helium reionization occurs at a later stage, the bubbles that form are predominantly composed of HeIII. This means that within these bubbles, the HeII fraction is close to zero, resulting in a negligible signal contribution. The signal mainly originates from outside of the HeIII bubbles. On the other hand, in the Early model, where helium reionization occurs at an earlier stage, the bubbles that form contain a mixture of HeII and HeIII. Inside the bubbles, the HeIII fraction is dominant, while the HeII fraction is concentrated at the edges of the bubbles. Consequently, the HeII signal is primarily sourced from the boundaries of the HeIII bubbles, where the transition between ionized and neutral regions occurs. The contrasting behavior between the Late and Early models arises from the different sequences of ionization during the reionization process. In the Late model, helium is initially fully ionized, and as reionization progresses, the HeIII bubbles expand and merge, leaving no significant HeII signal within the bubbles. In contrast, in the Early model, the reionization process starts from a predominantly neutral state, resulting in the formation of HeIII bubbles with distinct HeII signal contributions along their perimeters. This difference in the distribution of the HeII signal, with the Late model having minimal signal inside the bubbles and the Early model having concentrated signal at the edges of the bubbles, further contributes to the contrasting features observed in the power spectrum measurements between the two models.

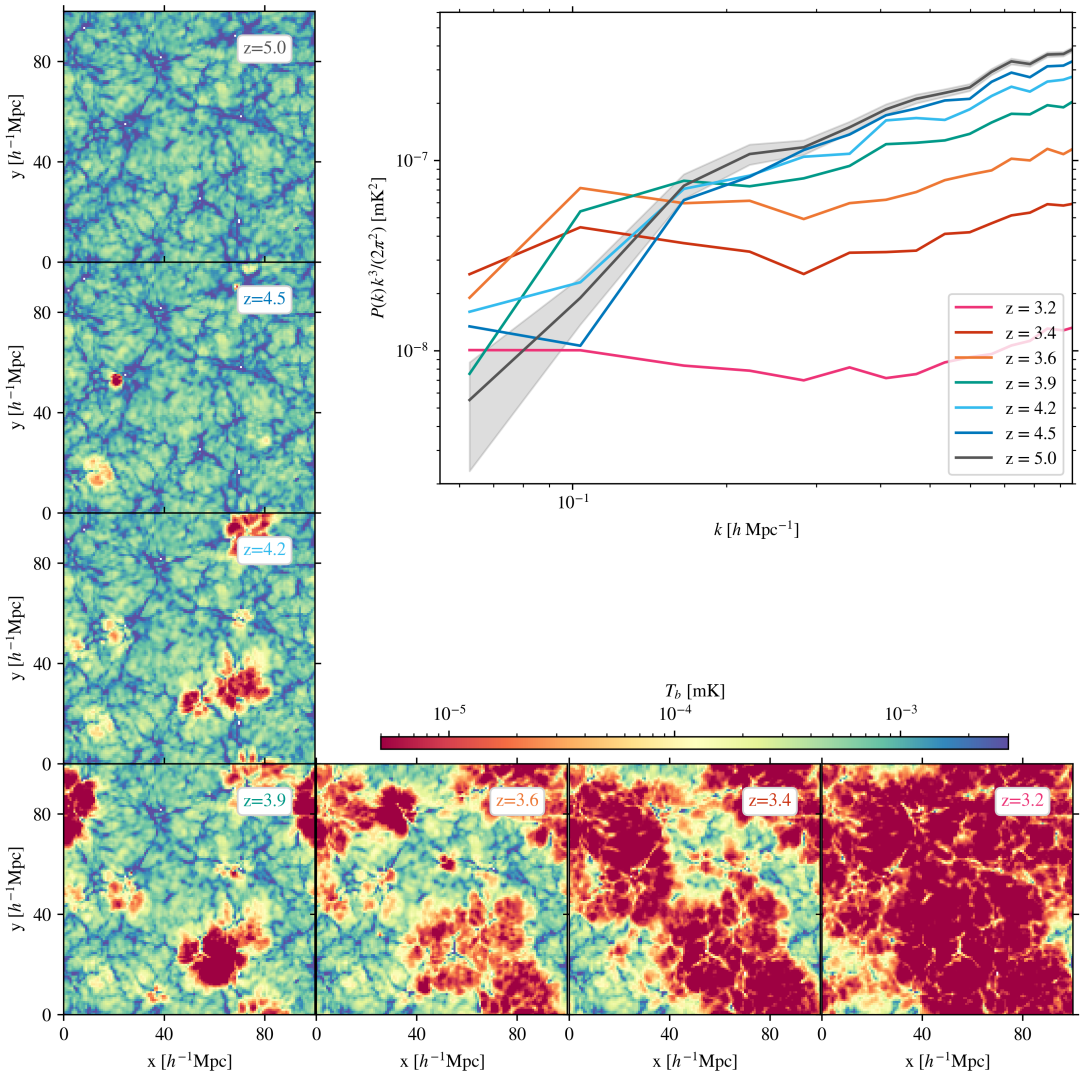

Figure 3.2. – Brightness temperature field snapshots for the Late model at different redshifts, along with the corresponding power spectra. The HeIII bubbles, resulting from late helium reionization, are evident in the temperature field. The power spectrum exhibits increased power on large scales due to the presence of these bubbles, while on small scales it resembles the matter power spectrum.

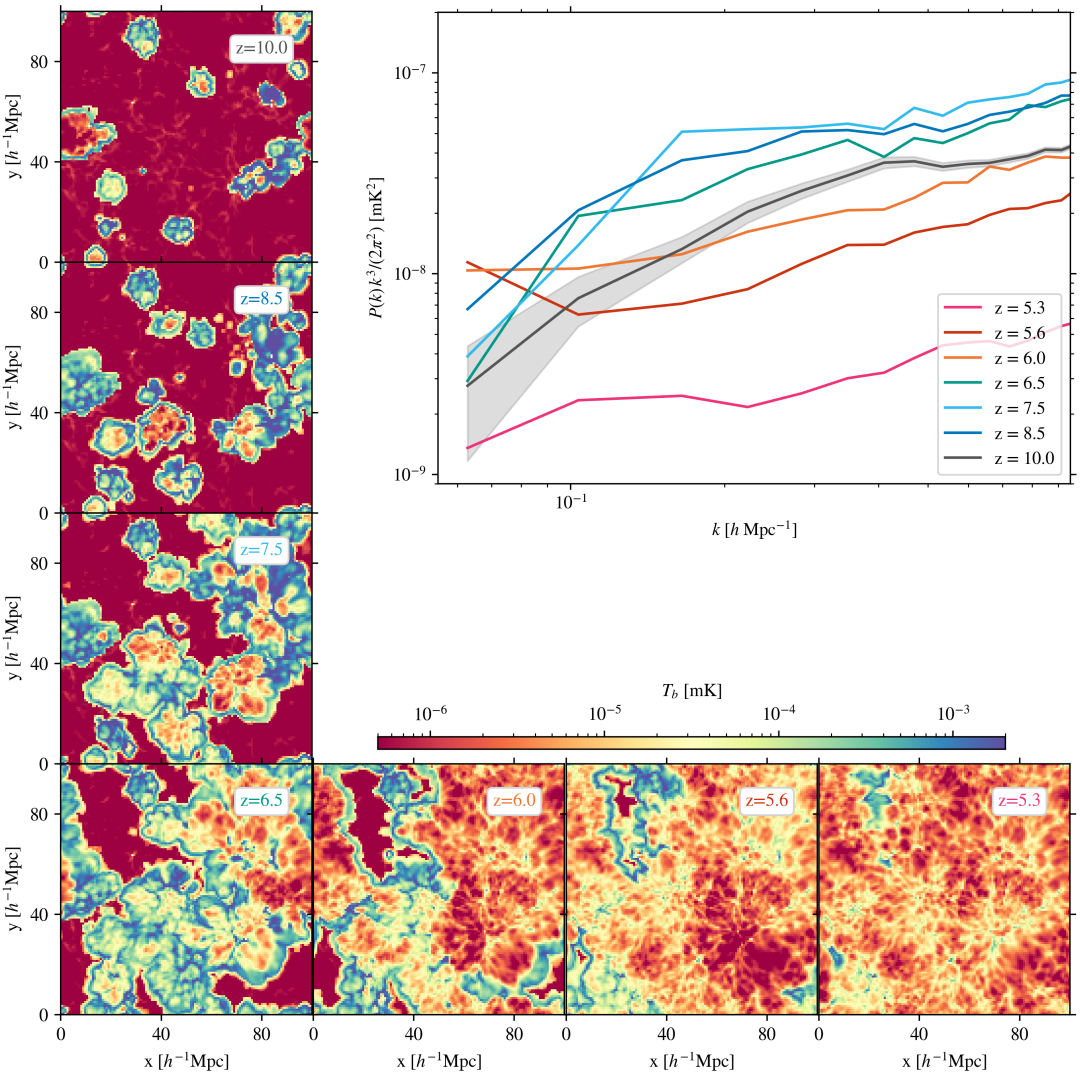

Figure 3.3. – As for Figure 3.2 for the Early model. The HeII signal primarily emanates from the edges of the HeIII bubbles formed during early helium reionization.

3.5. Systematic effects

We carefully consider systematic effects in the observations of the brightness temperature. This includes accounting for the chosen frequency bandwidth, the angular resolution of the telescope, the foreground removal procedure, and the thermal noise of the instrument. By addressing these systematic effects, we ensure the reliability and accuracy of our estimates.

3.5.1. Frequency bandwidth and angular resolution

To simulate the 3.46 cm tomography, we divide the total receiver bandwidth into a number of frequency channels. Each channel corresponds to a specific range of frequencies and provides spectroscopic information about the observed region of the sky. The channel bandwidth plays a crucial role in determining the signal-to-noise ratio of the radio images (see section 3.5.2).

The narrower the channel bandwidth, the better the spectral resolution. However, narrower channels also lead to lower signal-to-noise ratios due to reduced integration time per channel. On the other hand, wider channel bandwidths sacrifice spectral resolution but improve the signal-to-noise ratio by allowing more integrated signal power within each channel. This approach is particularly useful for detecting large-scale structures and improving the overall sensitivity of the observation. The choice of channel bandwidth depends on the specific goals of the observation, the desired level of sensitivity, and the available resources. It is a trade-off between spectral resolution and signal-to-noise ratio, and it is important to optimize these parameters based on the scientific objectives and instrumental constraints of the 3.46 cm tomography. By using the relation between the redshift and comoving distance χ in the background cosmological model, the channel bandwidth $\Delta\nu$ at a specific redshift z corresponds to a radial separation in comoving distance given by

$$\Delta\chi = \int_{z_-}^{z_+} c\frac{\mathrm{d}z}{H(z)} \tag{3.4}$$

where

$$z_\mp = \frac{1}{\dfrac{1}{1+z} \pm \dfrac{\Delta\nu}{2\nu_{\mathrm{rest}}}} - 1 \ . \tag{3.5}$$

By setting $\Delta\chi = L/128 = 0.78h^{-1}$Mpc, i.e. the lower comoving distance accessible on the direction parallel to the line-of-sight with the simulation considered, we obtain the corresponding $\Delta\nu$ at each redshift. In our analysis, we adopt the distant-observer (plane-parallel) approximation, which assumes that the 3.46 cm observations are conducted from a distant location compared to the scale of the observed structures. In this approximation, we choose one of the axes of the simulation box as the line of sight along which the observed volume is projected. By selecting a specific axis as the line of sight, we simplify the analysis by assuming that the variations in the observed 3.46 cm signal are primarily due to the structures along that particular direction. This approximation allows us to focus on the large-scale structures and the statistical properties of the observed data. While the distant-observer approximation neglects the full three-dimensional nature of the observed volume, it provides a useful simplification that enables efficient analysis and interpretation of the data.

To generate synthetic maps with a larger channel bandwidth $\Delta\nu$, we perform spatial averaging of the brightness temperature δT_b at a fixed position on the sky over the corresponding comoving distance scale $\Delta\chi$. This averaging process allows us to create intensity

maps of the 3.46 cm signal at different redshift. Each map represents the integrated signal over a specific frequency range. To incorporate the finite angular resolution of the instrument, we convolve the two-dimensional intensity maps with the telescope beam. For a single dish telescope with a diameter of D, we assume a Gaussian beam profile. The full width at half maximum (FWHM) of the beam, denoted as $\theta_{\rm FWHM}$, is given by the equation:

$$\theta_{\rm FWHM} = 1.2\,\frac{\lambda_{\rm rest}\,(1+z)}{D}, \tag{3.6}$$

where $\lambda_{\rm rest}$ is the rest-frame wavelength of the 3.46 cm signal. The FWHM describes the angular size of the beam, indicating the angular resolution of the telescope. To perform the convolution, we transform the maps and the beam into Fourier space, multiply their Fourier transforms, and then transform back to real space. This convolution process accounts for the blurring effect introduced by the finite angular resolution of the instrument, resulting in the final observed intensity maps that include the combined effects of the instrumental beam and the underlying 3.46 cm signal. The factor of 1.2 in equation (3.6) corresponds to a correction for the non-uniform illumination of the antenna and is not directly associated with the Rayleigh criterion. In an ideal scenario, the numerical factor would be 1.02, as described in previous studies (e.g., Hall, 2005; Chen et al., 2020). This correction accounts for factors such as the antenna efficiency and the shape of the antenna beam pattern. The Gaussian beam with a FWHM given by Equation (3.6) corresponds to an isotropic smoothing in the plane of the sky. The standard deviation of this Gaussian smoothing is denoted by Σ and can be calculated using

$$\Sigma = \frac{\theta_{\rm FWHM}\,(1+z)\,d_{\rm a}}{2\sqrt{2\ln 2}}\,, \tag{3.7}$$

where $d_{\rm a}$ is the angular-diameter distance to redshift z in the background (in a flat Universe, $(1+z)\,d_{\rm a} = \chi$).

3.5.2. Thermal noise

The output of a radio telescope is subject to thermal noise, which can be approximated as Gaussian for single-dish observations. The root mean square (rms) noise fluctuation associated with an integration time $t_{\rm pix}$ per pointing can be calculated using the radiometer equation:

$$\sigma_{\rm N} = \frac{T_{\rm sys}}{\sqrt{t_{\rm pix}\,\Delta\nu}}, \tag{3.8}$$

where $T_{\rm sys}$ represents the total system temperature and $\Delta\nu$ is the channel bandwidth. For an antenna of SKA-1 MID, the system temperature is given by (Square Kilometre Array Cosmology Science Working Group et al., 2018)

$$T_{\rm sys}(\nu) = T_{\rm spl} + T_{\rm CMB} + T_{\rm gal}(\nu) + T_{\rm rx}(\nu), \tag{3.9}$$

where $T_{\rm spl} \approx 3\,{\rm K}$ and $T_{\rm CMB} \approx 2.73\,{\rm K}$ are the spill-over and the cosmic-microwave-background contributions. The contribution of the Galaxy to the total system temperature can be modeled as

$$T_{\rm gal}(\nu) = 25 \left(\frac{408\,{\rm MHz}}{\nu}\right)^{2.75}\ {\rm K} \tag{3.10}$$

and the receiver-noise temperature as

$$T_{\rm rx}(\nu) = 15\,{\rm K} + 30 \left(\frac{\nu}{\rm GHz} - 0.75\right)^2\ {\rm K}. \tag{3.11}$$

For DSA-2000 and PUMA experiments, the system temperature is set at 25 K.

To construct a map of the large-scale structure of the Universe, we can utilize multiple antennas simultaneously. The total observation time required to map the region of the sky with solid angle $\Omega_{\rm surv}$ is given by

$$t_{\rm obs} = \frac{t_{\rm pix}}{N_{\rm dish}, N_{\rm beam}} \frac{\Omega_{\rm surv}}{\Omega_{\rm pix}}, \tag{3.12}$$

where $t_{\rm pix}$ is the integration time per pixel, $N_{\rm dish}$ is the number of available dishes, each equipped with $N_{\rm beam}$ feedhorns (typically 1 or 2), and $\Omega_{\rm pix} \approx \theta^2_{\rm pix}$ is the solid angle covered by a single pixel. This equation accounts for the time required per pixel, the number of dishes, and the solid angle coverage of each pixel, allowing us to estimate the total observation time for mapping the desired region of the sky. To ensure proper sampling of the map and satisfy the Nyquist-Shannon theorem, the angular size of each pixel, $\theta_{\rm pix}$, should be smaller than half of the FWHM of the telescope beam, denoted as $\theta_{\rm FWHM}$. In practical terms, $\theta_{\rm pix}$ is typically chosen to be between $\theta_{\rm FWHM}/7$ and $\theta_{\rm FWHM}/3$ (Marr et al., 2015). This range ensures that the map is adequately sampled, capturing the necessary level of detail without introducing significant aliasing effects. By appropriately selecting the pixel size, we can achieve a balance between spatial resolution and computational efficiency in the map reconstruction process.

It is important to note that this estimate is based on the assumptions and parameters provided, and the actual observation time may vary depending on specific observational constraints, data processing requirements, and other factors. The analytical expression for the power spectrum of the noise, as defined in equation (3.8), is given by $N(k) = \sigma^2_{\rm N} V_{\rm pix}$, where $V_{\rm pix} = A_{\rm pix} \cdot l_{\rm pix}$ represents the volume covered by each pixel, $A_{\rm pix}$ the sky area

covered and $l_{\rm pix}$ the direction along the line-of-sight. By combining Eq. (3.8) and (3.12), we can write the power spectrum of the noise as

$$
\begin{aligned}
N &= \frac{T_{\rm sys}^2}{\left[\sqrt{\left(t_{\rm obs} N_{\rm dish} N_{\rm beam} \frac{\Omega_{\rm surv}}{\Omega_{\rm pix}}\right) \delta\nu}\right]^2} A_{\rm pix}^2 l_{\rm pix} = \\
&= \frac{T_{\rm sys}^2}{t_{\rm obs} N_{\rm dish} N_{\rm beam} \frac{\Omega_{\rm pix}^2}{\Omega_{\rm surv}} \delta\nu} A_{\rm pix}^2 l_{\rm pix} = \\
&= \frac{T_{\rm sys}^2}{t_{\rm obs} N_{\rm dish} N_{\rm beam} \delta\nu} \frac{A_{\rm pix}^2}{\Omega_{\rm pix}^2} \Omega_{\rm surv} l_{\rm pix}.
\end{aligned}
\tag{3.13}
$$

3.5.3. Foreground removal

The LIM technique used to map the large-scale structure of the Universe encounters several obstacles, particularly due to foreground contamination. Both Galactic and extragalactic foreground sources, including synchrotron emission from the Milky Way and extragalactic point sources, can have intensities several orders of magnitude higher than the desired HI signal. Initially, the task of disentangling these foreground components may appear daunting. However, the HeII transition, with its relatively high rest frequency, is less susceptible to foreground contamination, providing some advantage in this regard.

A crucial distinction exists between foregrounds and the 3.46 cm signal, which can be effectively leveraged. Foregrounds typically exhibit smooth frequency-dependent variations at a specific sky position, whereas the 3.46 cm signal is expected to display poor frequency correlation. This fundamental difference serves as the foundation for foreground separation techniques (e.g., Shaver et al., 1999; Di Matteo et al., 2002; Oh & Mack, 2003; Zaldarriaga et al., 2004). The concept involves exploiting the distinct spectral characteristics of foregrounds and the 3.46 cm signal to statistically differentiate between them.

While the concept of distinguishing foregrounds from the 3.46 cm signal based on their spectral behaviors is promising, its practical implementation necessitates careful attention to several complex factors. One crucial consideration is the presence of frequency-dependent beam shapes, which can introduce additional challenges when separating foregrounds from the 3.46 cm signal. Addressing and accounting for these complexities is essential to successfully remove foreground contamination and extract the desired cosmological information from observations at 3.46 cm.

Despite the challenges involved, researchers have put forward and evaluated various foreground-cleaning techniques using simulated data (see recent reviews by Wolz et al., 2015). These techniques strive to distinguish the foregrounds from the 3.46 cm signal, but they may inadvertently lead to unintended consequences, such as the removal of significant power on large scales from the 3.46 cm signal.

In our approach, inspired by Alonso et al. (2017) and Cunnington et al. (2019), we assume that foreground subtraction eliminates Fourier modes of δT_b with comoving radial wavenumbers satisfying

$$k_{\parallel} < k_{\parallel}^{\mathrm{FG}} \approx \frac{\pi H(z)}{c\,(1+z)\,\xi}, \tag{3.14}$$

where ξ is a dimensionless parameter of order unity. The parameter $k_{\parallel}^{\mathrm{FG}}$ denotes the lower limit of the wavenumber range where foregrounds can be successfully distinguished from the 3.46 cm signal. The precise value of ξ is difficult to ascertain and relies on the specific foreground removal technique employed. Nonetheless, a reasonable approximation for 21 cm experiments, as proposed by Cunnington et al. (2019), is $\xi \approx 0.1$. We conservatively decide to adopt the same estimate for ξ, even though the 3.46 cm signal is less prone to foregrounds contamination. We set the cut-off scale to $k_{\parallel}^{\mathrm{FG}} \approx 0.01\,h\,\mathrm{Mpc}^{-1}$, which is slightly larger than the fundamental frequency $k_{\mathrm{f}} = 2\pi/L$ corresponding to the size of our simulation box. By excluding the Fourier modes with $k_{\parallel} = 0$, we eliminate the monopole and the largest-scale fluctuations from the brightness temperature distribution δT_b. Consequently, the average value of δT_b across the entire simulation box becomes zero. This poses an additional challenge for detecting the 3.46 cm signal, as the mean signal no longer contributes to the overall measurement, thereby increasing the required sensitivity.

The 3.46 cm signal is not only influenced by the ionization state of helium, but also by the underlying DM density. On large scales (small k), the power spectrum of the 3.46 cm signal can be extended by rescaling the matter power spectrum. This is illustrated in Figure 3.4 where the various effect of the systematic effects on the power spectra are presented. The measured power spectrum from the simulations is shown (blue continuous line), accounting for the pixel size (light blue dash line) and the Gaussian smoothing (green dash-dotted line) caused by the angular resolution of the instrument. Additionally, the theoretical power spectrum of matter[2] is scaled (orange continuous line) by a factor C_1 to match the measured power spectrum at $k = 1h\mathrm{Mpc}^{-1}$ and smoothed (orange dash line) for comparison. The smoothing is made via the convolution with the window function (Li et al., 2016)

$$W(k) = e^{-k^2\Sigma^2} \int_0^1 e^{-k^2\left(\sigma_{\parallel}^2 - \Sigma^2\right)\mu^2} \mathrm{d}\mu\,, \tag{3.15}$$

where $\sigma_{\parallel}$ is the one given by the minimum frequency bandwidth accessible, as described in Section 3.5.1, and $\mu = \cos\theta$ with θ the spherical polar angle in k-space. The analysis focuses on k values ranging from $0.01, h, \mathrm{Mpc}^{-1}$ to $2, h, \mathrm{Mpc}^{-1}$ to examine the behavior of the power spectrum across different scales. On small k values, the dominant influence of the HeII bubbles created by AGNs and QSOs becomes apparent. This means that the extension of the power spectrum using the rescaled matter power spectrum is conservative since it does not account for the effect of the bubbles. The impact of the bubbles on

[2]The matter power spectrum is computed with `CAMB` (Lewis & Challinor, 2011).

the power spectrum at small scales needs to be taken into consideration to obtain a comprehensive understanding of the observed signal.

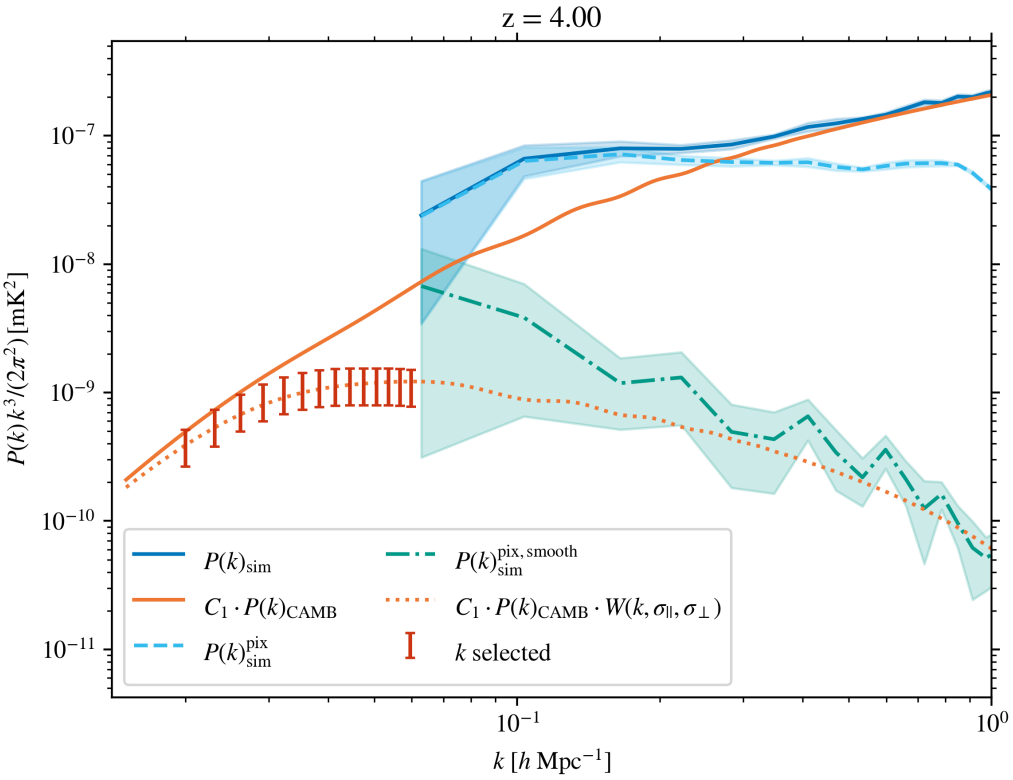

Figure 3.4. – Power spectrum of the 3.46 cm line signal. The measured power spectrum from simulations is shown (blue continuous line), along with the power spectrum considering the pixel size $\theta_{\rm pix}$ (light blue continuous line), Gaussian smoothing due to angular resolution (green continuous line). The rescaled matter power spectrum (orange continuous line) and smoothed (orange dotted line) are also shown. The dominant effect of HeII bubbles on small k values indicates the conservative nature of the extension. In red, we display the values considered in the following analysis.

3.6. Signal-to-noise ratio

To assess the detectability of the signal by current radio surveys, it is useful to define the signal-to-noise ratio (SNR) for the power spectrum as

$$\mathrm{SNR}^2(k) = \frac{P^2(k)}{\sigma_P^2(k)} = \tag{3.16}$$

$$= N_q \frac{P^2(k)}{[P(k)+N]^2}. \tag{3.17}$$

Assuming a cubic geometry, the number of modes within the k-bin of width δk is

$$N_q = \frac{1}{2} 4\pi k^2 \delta k \frac{V_{\rm surv}}{(2\pi)^3} = \frac{k^2 \delta k}{(2\pi)^2} V_{\rm surv}\,, \tag{3.18}$$

where $V_{\rm surv} = A_{\rm surv} \cdot l_{\rm s}$ is the volume of the survey, given by the observed area perpendicular to the line-of-sight $A_{\rm surv}$ and the comoving distance in the direction parallel to the line-of-sight $l_{\rm s}$.

3.7. Results

In this section we present the findings and analyses of our study on the detectability of the HeII signal using radio astronomy surveys. We examine the integration time, signal-to-noise ratio, and power spectrum behavior for both late and early reionization models. Additionally, we explore the influence of various factors, such as observed volume, redshift range, and survey parameters, on the detectability of the HeII signal. The results shed light on the prospects and challenges associated with detecting the HeII signal and provide insights into the optimal strategies for future radio astronomy surveys.

While the minimum value accessible is $\Delta\chi$ as definied in Eq. (3.4), it is more convenient to consider a larger volume, i.e. a broader redshift range, to increase the SNR. We are interested in quantifying how much integration time is needed to access the HeII signal, therefore it is convenient to express the SNR as a function of the integration time. Combining Eqs. (3.13), (3.17) and (3.18), we can write the SNR as

$$\begin{aligned} {\rm SNR}^2(k) &= \frac{k^2 \delta k}{(2\pi)^2} A_{\rm surv} l_{\rm s} \left[1 + \frac{N}{P(k)}\right]^{-1} = \\ &= \frac{k^2 \delta k}{(2\pi)^2} A_{\rm surv} l_{\rm s} \left[1 + \frac{1}{P(k)} \frac{T_{\rm sys}^2}{t_{\rm obs} N_{\rm dish} N_{\rm beam} \delta\nu} \frac{A_{\rm pix}^2}{\Omega_{\rm pix}^2} \Omega_{\rm surv} l_{\rm pix}\right]^{-1}. \end{aligned} \tag{3.19}$$

The SNR depends on several quantities, some related to the frequency and frequency bandwidth we consider for the observations $(T_{\rm sys}(\nu), l_{\rm pix}(\nu, \delta\nu), \delta\nu)$, to the radio-telescope we consider $(T_{\rm sys}(\nu), N_{\rm dish}, N_{\rm beam}, \Omega_{\rm surv})$, to the signal itself and the scale considered $(P(k), k)$ and to the integration time $t_{\rm obs}$. In order to maximize the SNR, it is useful to consider the integrated signal-to-noise ratio SNRt, i.e. summing the contribution on all scales, as

$${\rm SNRt} = \frac{1}{2\pi} \sqrt{A_{\rm surv} l_{\rm s}} \sum_k \left\{ k\sqrt{\delta k} \left[1 + \frac{1}{P(k)} \frac{T_{\rm sys}^2}{t_{\rm obs} N_{\rm dish} N_{\rm beam} \delta\nu} \frac{A_{\rm pix}^2}{\Omega_{\rm pix}^2} \Omega_{\rm surv} l_{\rm pix}\right]^{-1/2} \right\}. \tag{3.20}$$

A large observed volume results in an increased SNRt, making a broad redshift range optimal for maximizing SNRt. Since the power spectrum is dependent on redshift, we utilize the mean power spectrum within the considered redshift range.

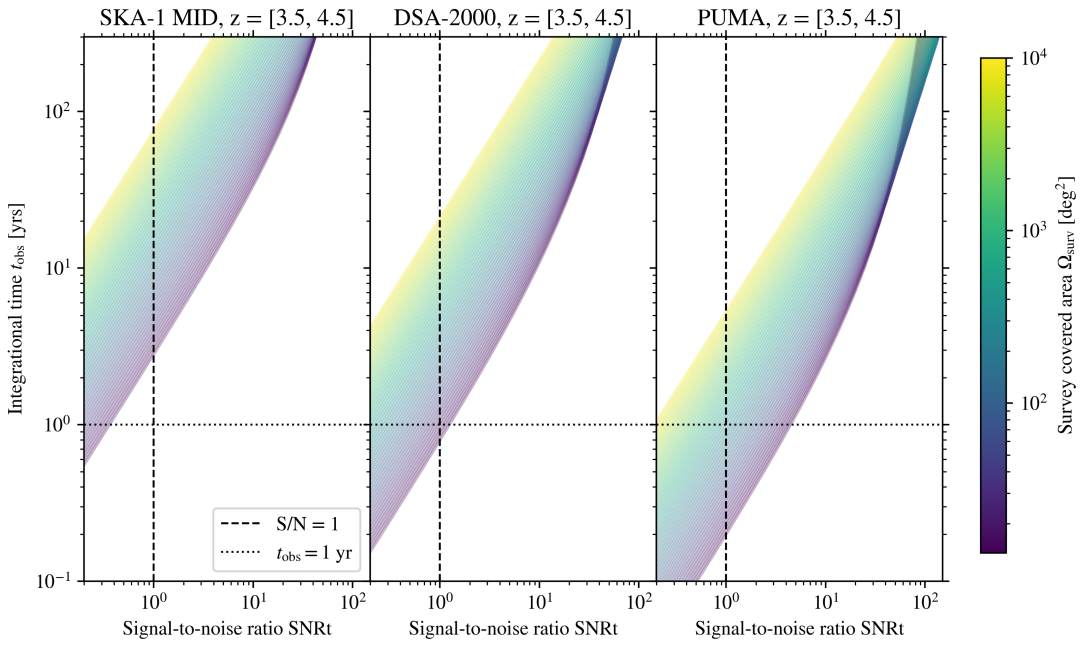

Figure 3.5. – Integration Time vs. Signal-to-Noise Integrated Ratio (SNRt) in the redshift range $3.5 \leq z \leq 4.5$. The figure illustrates the relationship between the integration time t_{obs} and the signal-to-noise integrated ratio SNRt. The integration time is influenced by the specific redshift range considered and the survey area Ω_{surv}, represented using different colors. A reference integration time of 1 year is depicted as a dotted black line, while the reference SNRt of 3 is shown as a dashed black line. Each panel corresponds to a distinct survey.

We illustrate our findings in Figures 3.5 and 3.6, which displays the integration time as a function of SNRt for a fixed observed volume determined by the selected redshift range (respectively $3.5 \leq z \leq 4.5$ and $5.5 \leq z \leq 6.5$) and covered area. The graph showcases the results for the three chosen experiments. The k values used for computing the SNRt are displayed in Figure 3.4 for $z = 4$, which includes both the k values obtained from the simulation boxes and those from the extended power spectrum. Our results indicate that the best outcomes are achieved when observing a relatively small region of the sky. This counterintuitive finding arises from the fact that, for a fixed observation time, the detected signal is more significant in a small area that can be quickly surveyed compared to a larger area that requires more time to cover (and overall provide a weaker signal for this reason). Similarly, for a fixed SNRt, a larger area to be surveyed requires a longer integration time. This relationship arises because a larger area necessitates more observations and data acquisition to achieve the desired SNRt, resulting in increased integration time. The integration time as a function of SNRt exhibits an upper limit on the signal quality. Regardless of the area covered, the SNRt will not continue to improve indefinitely even with extended observation durations. This behavior is evident in the curves displayed on

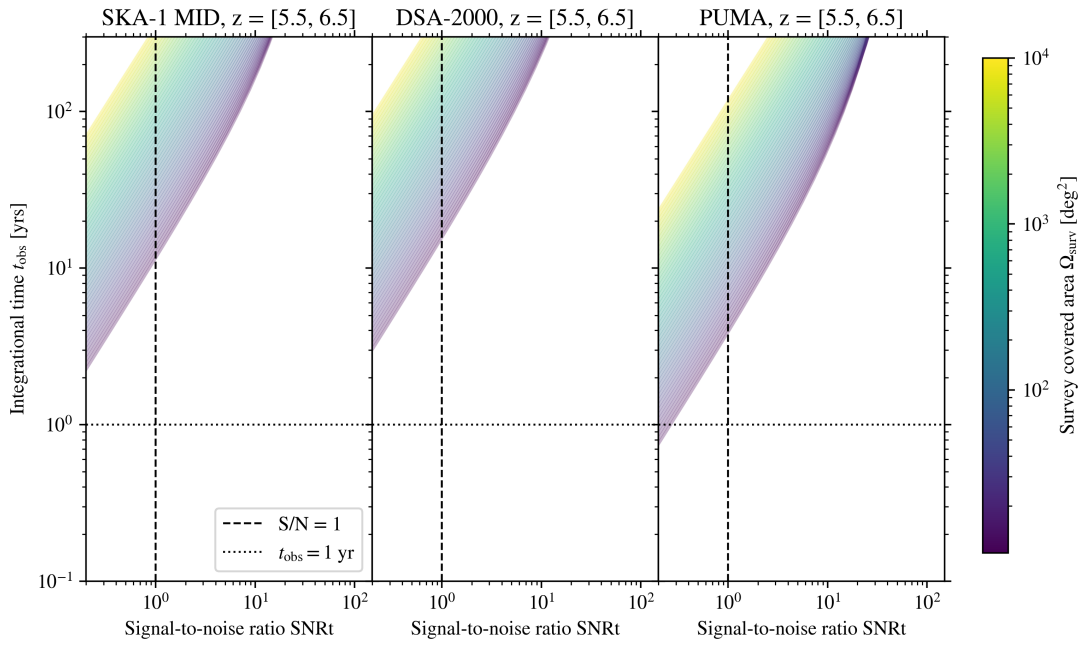

Figure 3.6. – As in Figure 3.5, in the redshift range $5.5 \leq z \leq 6.5$.

the right side of each panel, illustrating that beyond a certain point, further increase in integration time does not yield significant improvements in the SNRt.

Among the three surveys considered, PUMA shows the most promising prospects for detecting the HeII signal in a late reionization model, essentially because of its large number of antennas. In this case, the integration time needed to obtain a good SNRt is of a few 10^{-1} yrs, more precisely an integration time of $t_{\rm obs} = 5 \times 10^{-1}$ yrs is needed to achieve a SNRt of 3 considering the smallest accessible area. In the case of an early reionization model, the integration time required for detection increases significantly due to several factors. Firstly, the signal itself is weaker since the HeII fraction never exceeds 0.2, as the first and second ionization of Helium occur simultaneously. Additionally, the impact of accounting for pixel size and beam smoothing becomes more significant at higher redshifts, resulting in a damping effect on the power spectrum. These combined factors contribute to the substantial increase in integration time by several orders of magnitude in the case of an early reionization model.

Detecting the HeII signal at high redshifts, particularly for the early reionization model, poses significant challenges and demands advanced surveys that are not currently available. However, the results obtained in our study indicate that the detection of the HeII signal at lower redshifts is more feasible. Such a detection could provide valuable insights and constraints on the nature of the reionization process, potentially allowing us to exclude certain reionization models. The figures presented in this study illustrate the improved

prospects for detecting the HeII signal at lower redshifts and highlight the importance of future observations in advancing our understanding of the reionization history.

3.8. Summary

Helium reionization is a pivotal milestone in the cosmic history, marking the phase when the Universe transitioned from being predominantly neutral to ionized. Understanding this epoch allows us to reconstruct the sequence of events that shaped the Universe and laid the foundation for the formation of cosmic structures. This process is driven by the emission of ionizing photons from the first generations of stars, galaxies, and black holes. Investigating this process enables us to probe the nature and evolution of these early sources and the interplay between them and the intergalactic medium. Moreover it is tightly connected to hydrogen reionization since the two processes happened roughly in the same time. By studying helium reionization, we can gain insights into the later stages of hydrogen reionization.

This study represents a innovative effort in utilizing line-intensity mapping to establish crucial constraints on the initiation and progression of helium reionization. By forecasting advanced radio surveys, we have explored the feasibility of detecting and tracking the HeII signal, leading to promising findings that can significantly enhance our understanding of the sources behind helium reionization. The application of line-intensity mapping techniques offers a powerful and innovative approach to shed new light on this important cosmic epoch, potentially unlocking valuable insights into the early Universe's evolution.

One of the main challenges in this endeavor is separating the HeII signal from foreground contamination (depite being less prone to foregrounds contamination than the HI signal because of its higher rest frequency). Galactic and extragalactic foregrounds, such as synchrotron emission and extragalactic point sources, can significantly outshine the faint HeII signal. Foreground separation techniques that exploit the different spectral behaviors of foregrounds and the HeII signal have been developed, paving the way for accurate measurements. Radio surveys as SKA-1 MID, DSA-2000 and PUMA - in a single-dish approach aiming for intensity mapping measurements - are well-suited for detecting the HeII signal. These surveys enable us to explore the power spectrum of the HeII signal and observe distinct features caused by the presence of HeII bubbles created by astrophysical sources.

Examining the power spectra of the matter and HeII signal, we found that the behavior of the HeII signal varies between models with late and early reionization. This highlights the significance of accounting for the effects of these bubbles in analyzing the HeII signal. When considering the sensitivity of HeII signal detection, we find that the integration time, the beam size and the observed volume play crucial roles. Observing smaller regions of the sky for a fixed time can yield larger signals but it does not give us access to small k modes, while larger areas necessitate longer integration times but capture the complexity of the

large-scale structures. The redshift range and survey area impact the required integration time for detecting the HeII signal as well.

Although detecting the HeII signal in an early reionization model presents challenges due to a weaker signal, redshift-dependent brightness temperature, and damping effects, our analysis revealed that a detection at lower redshifts can still provide valuable insights. By constraining the type of reionization model, such a detection could potentially exclude certain models and enhance our understanding of helium reionization. In particular, PUMA - followed by DSA-2000 - stands out as one of the most promising endeavors for detecting the HeII signal in a late reionization model. One of the key advantages of the PUMA experiment is its large number of antennas, which enables it to achieve higher sensitivity and better angular resolution compared to other experiments.

Our exploration of helium reionization and the detection of the HeII signal through radio surveys and intensity mapping has offered valuable insights, allowing us to forecast for upcoming and ongoing experiments.

3.9. Challenges and next steps

In the ongoing exploration of helium reionization and the detection of the HeII signal, there are several potential avenues for future research and development.

Radio surveys. Continued advancements in radio astronomy instrumentation and techniques will play a crucial role. This includes the development of next-generation radio telescopes with even higher sensitivity, wider frequency coverage, and improved angular resolution. These advancements will enable precise measurements of the HeII signal, better separation of foregrounds, and enhanced mapping of the LSS of helium reionization.

Theoretical modelling. Foreground contamination remains a major challenge in detecting the faint HeII signal. Further research and development are needed to refine foreground removal methods and minimize their impact on the HeII power spectrum. This involves exploring innovative statistical techniques, machine learning algorithms, and sophisticated data analysis approaches to effectively separate foregrounds from the desired signal. Theoretical modeling and simulations play a crucial role in understanding the underlying physics of helium reionization and its impact on the HeII signal. Future research should focus on refining and expanding existing models, incorporating more astrophysical processes. Simulations should strive to capture the intricate connection between HeII ionization, density fluctuations, and the formation of HeIII bubbles.

Cross-correlation with different probes. Cross-correlation studies between the HeII signal and other cosmological probes can provide valuable insights. Exploring the synergies between HeII observations and other datasets, such as galaxy surveys, CMB measurements, and Lyman-alpha forest observations, can offer a more comprehensive picture of the reionization process and the underlying astrophysics.

CHAPTER 4

How to identify galaxies quenched by reionizaton

Benedetta Spina[1], Enrico Garaldi[2] & Cristiano Porciani[1]
[1]*Argelander Institut für Astronomie , Auf dem Hügel 71, 53121 Bonn, Germany,*
[2]*Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany*
In preparation

4.1. Introduction

The reionization of the Universe has a profound impact on the early evolution of galaxies. As they form stars, their cold gas reservoir is usually replenished through the cooling of the hot gas surrounding the halo as well as cold accretion streams (Birnboim & Dekel, 2003; Bauermeister et al., 2010). During the EoR, high-energy photons emitted by ionizing sources can reach and heat up the gas within the IGM, preventing gas accretion. This process has the potential to suppress the formation of new stars, along with other mechanisms. For example, the gas can be directly removed from the galaxy through ram-pressure stripping when it rapidly falls into the hot intra-cluster medium of a cluster environment (Gunn & Gott, 1972; Bekki, 2009). Alternatively, the environment may prevent the accretion of fresh gas onto the galaxy, leading to quenching as the star formation exhausts the remaining gas reservoir over time. Feedback and outflows are also considered to play a significant role in gas removal from galaxy halos (Moore et al., 1996; Balogh et al., 2016). In the overconsumption scenario, the depletion of gas is enhanced by outflows generated by star formation and these outflows can result from radiation pressure or subsequent supernovae (McGee et al., 2014).

Halos with low circular velocities are highly susceptible to a range of energetic feedback processes, including those triggered by supernovae and ultraviolet radiation (Mac Low & Ferrara, 1999; Bullock et al., 2000). In the initial phases of cosmic evolution, the first generations of stars and galaxies emit substantial amounts of ionizing photons, which contribute to the formation of the meta-galactic UV background and play a crucial role in the reionization of the Universe and of low mass galaxies. This contribution can manifest in various ways, such as inducing photoevaporation, extending cooling times, and impeding gas inflows, resulting in the suppression of dwarf galaxy formation (Efstathiou, 1992; Gnedin, 2000; Iliev et al., 2005).

The quenching of dwarf galaxies during reionization is believed to have a significant impact on their subsequent evolution (McGaugh & Wolf, 2010; Wheeler et al., 2014; Faisst et al., 2017; Katz et al., 2020). These galaxies have originated within atomic cooling halos during the early Universe, i.e. within halos massive enough to cool their gas down to low enough temperatures for atomic hydrogen to efficiently condense, leading to the formation of small and low-mass galaxies, which are the progenitors of modern-day dwarf galaxies. For this reason, dwarf galaxies, with their lower masses and shallower gravitational potential wells, are expected to be more vulnerable to the effects of reionization compared to larger galaxies. Consequently, star formation in dwarf galaxies can be suppressed, influencing observable properties such as stellar masses, sizes, and gas content. Investigating the quenching of dwarf galaxies during reionization offers valuable insights into the interplay between astrophysical processes in galaxy formation and the cosmic reionization epoch. It contributes to our understanding of galaxy formation, evolution, and the role played by the ionization state of the Universe in shaping their properties. The inefficiency of star formation in the lowest mass DM halos and the presence of ancient stellar populations in ultrafaint dwarf galaxies suggest the existence of a process that quenches star formation in these galaxies during the early Universe (Smith et al., 2008; Wheeler et al., 2014; Katz et al., 2020).

Numerical simulations have established the effects of reionization on the gas content of low-mass dwarf galaxies, but disagreements persist regarding the characteristic mass of suppression, the timing and environmental dependence of suppression, the subsequent star formation rates (SFRs), and the relative importance of different modes of suppression (Gnedin, 2000; Wise et al., 2014; Ocvirk et al., 2016; Wu et al., 2019).

Detecting the signatures of reionization in the Local Group has proven to be complex, as most dwarf galaxies in the group show extended star formation histories that are inconsistent with early truncation by reionization (Grebel & Gallagher, 2004; Weisz et al., 2014; Skillman et al., 2017). The discovery of faint dwarf galaxies with the Sloan Digital Sky Survey has increased the prospects of identifying true fossils of reionization, but the influence of other environmental factors complicates the identification process (Martin et al., 2004; Belokurov et al., 2010; Rhode et al., 2013). The Local Cosmology from Isolated Dwarfs program focuses on isolated dwarf galaxies in the Local Group and finds that their extended star formation histories are not compatible with reionization models (Cole et al., 2007; Monelli et al., 2010; Hidalgo et al., 2011). A comprehensive comparison

between reionization models and the star formation histories of Local Group dwarf galaxies is lacking. There is no general consensus, with some observations indicating suppression of star formation due to reionization (Weisz et al., 2014; Brown et al., 2014; Bettinelli et al., 2018) and others showing no definite signature. Additionally, recent theoretical works suggest that star formation may be reignited long after reionization, further complicating the interpretation of the role of reionization in dwarf galaxy suppression (Benítez-Llambay et al., 2015; Wright et al., 2019; Ledinauskas & Zubovas, 2018). The limited time resolution of inferred star formation histories and the possibility of star formation reigniting long after reionization complicate the interpretation of the role of reionization in dwarf galaxy suppression.

The primary objective of this study how the characteristics of the galaxies contribute to the quenching process. A particular emphasis is placed on examining the role of the environment in which galaxies reside. The redshift at which reionization occurs locally - namely the local reionization redshift $z_{\rm reion}$ - is a crucial parameter that can influence the formation and evolution of galaxies. We aim to explore how the timing and progression of reionization, whether it is self-induced or induced from a nearby galaxy, can shape the quenching of galaxies. By exploring these aspects, we seek to gain insights into the mechanisms that drive quenching and how they are influenced by both internal galaxy properties and the larger-scale cosmic environment. We aim to provide a simple but powerful tool to complement observations, able to determine the quenching probability for galaxies with given observed features. Determining observationally the cause of a galaxy's quenching is challenging; however, we present here a simple prescription to predict whether and how a galaxy has quenched because of reionization by interpreting other observable properties.

The chapter is organised as follows. After introducing the suite of simulations used in this work in Section 4.2, we define and study the redshift of reionization for the galaxies in the simulations in Section 4.3. We also explore different reionization mechanisms and the role on the environment. In Section 4.4 we discuss several definitions of quenching by reionization and we classify the galaxies in the simulation accordingly. In Section 4.5, four key observables characterizing quenched galaxies are described and analyzed. The results of this project are shown and discussed in Section 4.6, and then summarized in Section 4.7. We briefly outline next steps and challenges in Section 4.8.

4.2. Simulations

We base our analysis on the THESAN suite of simulations, large volume ($L_{\rm box} = 95.5$ cMpc) radiation-magneto-hydrodynamic simulations that simultaneously model the large-scale statistical properties of the intergalactic medium during reionization and the resolved characteristics of the galaxies responsible for it. The flagship simulation has DM and baryonic mass resolutions of $m_{\rm P} = 3.1 \times 10^6 {\rm M}_\odot$ and $5.8 \times 10^5 {\rm M}_\odot$, respectively, enabling predictions down to the atomic cooling limit. The gravitational forces are softened on

scales of 2.2 ckpc with the smallest cell sizes reaching 10 pc at $z = 5.5$. The simulations use an efficient radiation hydrodynamics solver (Arepo-RT, Kannan et al., 2019) that precisely captures the interaction between ionizing photons and gas, coupled to well-tested galaxy formation (IllustrisTNG) and dust models (McKinnon et al., 2016, 2017) to accurately predict the properties of galaxies.

By incorporating detailed models of galaxy formation and radiation hydrodynamics, the Thesan simulations allow for a exploration of the interaction between reionization and the suppression of star formation in dwarf galaxies. The high-resolution of the simulations enables the modelling of individual dwarf galaxies and the accurate modeling of their properties. This allows us to examine the effects of reionization on the gas content, star formation rates, and the overall evolution of dwarf galaxies. The simulations can explore how the heating and photoionization of the gas on the outskirts of dwarf galaxies affect their ability to form stars. Additionally, the simulations can investigate how the ionizing radiation and resulting feedback processes, such as photoevaporation and gas outflows, impact the gas content and star formation rates of dwarf galaxies.

Utilizing the LHaloTree algorithm, we can track the accretion history of each subhalo, providing crucial insights into the processes affecting dwarf galaxies, particularly quenching. Analyzing the merger trees allows us to identify environmental factors, such as proximity to massive galaxies or ionized regions, influencing star formation quenching in dwarfs. Additionally, we make use of the 81 snapshots of the simulation, encompassing a comprehensive description of the system. These snapshots offer an extensive array of properties for each resolution element, providing a detailed picture of the simulation at specific cosmic times. Halos and subhalos, and their associated properties are identified at run time through the FOF and SUBFIND algorithms (Springel et al., 2001) and collected in the Group and Subgroup catalogs. The simulation suite also includes high time-cadence Cartesian outputs of the full particle data, which are mapped onto a regular Cartesian grid. From these outputs, we extract the HII fraction and the underlying DM density fields. These lower-resolution, high time-cadence outputs are well-suited for investigating the large-scale properties of the reionization process.

4.3. Reionization redshift

The local reionization redshift identifies the redshift at which the process of reionization takes place in a given region. We define it as the point in time when the fraction of ionized hydrogen becomes greater than 0.5 and remains so thereafter, i.e.

$$z_{\rm reion} = \min_z \left(f_{\rm HII} > 0.5 \right) . \tag{4.1}$$

We tested that this definition gives the most stable and consistent results among the galaxies we selected.

In our approach, we calculate the local reionization redshift by tracking the evolution of individual galaxies throughout the simulation, rather than for each spatial location in

the simulation. This method enables us to consistently account for the galaxy motion and understand how its reionization properties change over time, without being limited to a fixed position in the simulation. By tracing the galaxy's evolution, we can capture the dynamic nature of reionization and its effects on the galaxy's surroundings.

It is important to note that the choice of the region over which the HII fraction is calculated can significantly impact the determination of the local reionization redshift. To account for the environmental effects on different scales on reionization, we apply different levels of smoothing on the HII fraction field. We convolve the field with a Gaussian kernel with different standard deviations σ = 0.1, 0.2, 0.5, 1.0 Mpc, helping us analyzing the impact of the surrounding environment on the reionization process. The local reionization redshifts computed at these scales is indicated as $z^{\sigma}_{\rm reion}$. If the smoothing scale is very small, it includes a significant contribution from the HI inside the galaxies due to their self-shielding properties. In such case, $z^{\sigma}_{\rm reion}$ is governed by the internal properties of galaxies, and it becomes insensitive to the influence of the external environment. On the other hand, if the smoothing scale is too large, it diminishes the contribution of individual galaxies, making it harder to assess their specific role in the reionization process. By carefully selecting an appropriate smoothing scale, we can strike a balance and gain insights into both the internal properties of galaxies and the external environment during reionization.

To specifically investigate the impact of the onset of (local) reionization on galaxies, our analysis focuses on galaxies that form in neutral regions and subsequently undergo reionization. We exclude galaxies that form in regions already ionized from our study. By excluding galaxies formed in ionized regions, we can isolate the effects of the reionization front and better understand its influence on the properties and evolution of galaxies. This approach allows us to observe how the transition from a neutral to an ionized environment affects the gas dynamics, star formation, and quenching processes within these galaxies. When observing a galaxy is then necessary to account for the environment as well, in order to understand if the galaxy has formed before or later its surrounding were ionized, and consequently apply the results obtained in this project.

In addition to tracking the reionization redshift, we also define the reionization mass $M_{\rm reion} = M_{\rm DM}\,(z = z_{\rm reion})$. This parameter represents the DM mass associated with each galaxy at the specific redshift of reionization, i.e. when we expect to see the effect of reionization. Sampling galaxies according to their reionization mass and redshift allows us to investigate the relationship between the properties of galaxies and their reionization history. This approach provides valuable insights into how the mass of a galaxy at the time of reionization influences its subsequent evolution and quenching process. By analyzing the distribution of galaxies across different reionization masses and redshifts, we can identify patterns and trends that shed light on the impact of reionization on galaxy formation and evolution. It enables us to explore questions such as whether there is a mass threshold for efficient reionization, how the timing of reionization affects the subsequent star formation and quenching of galaxies, and whether there are distinct evolutionary paths for galaxies based on their reionization properties.

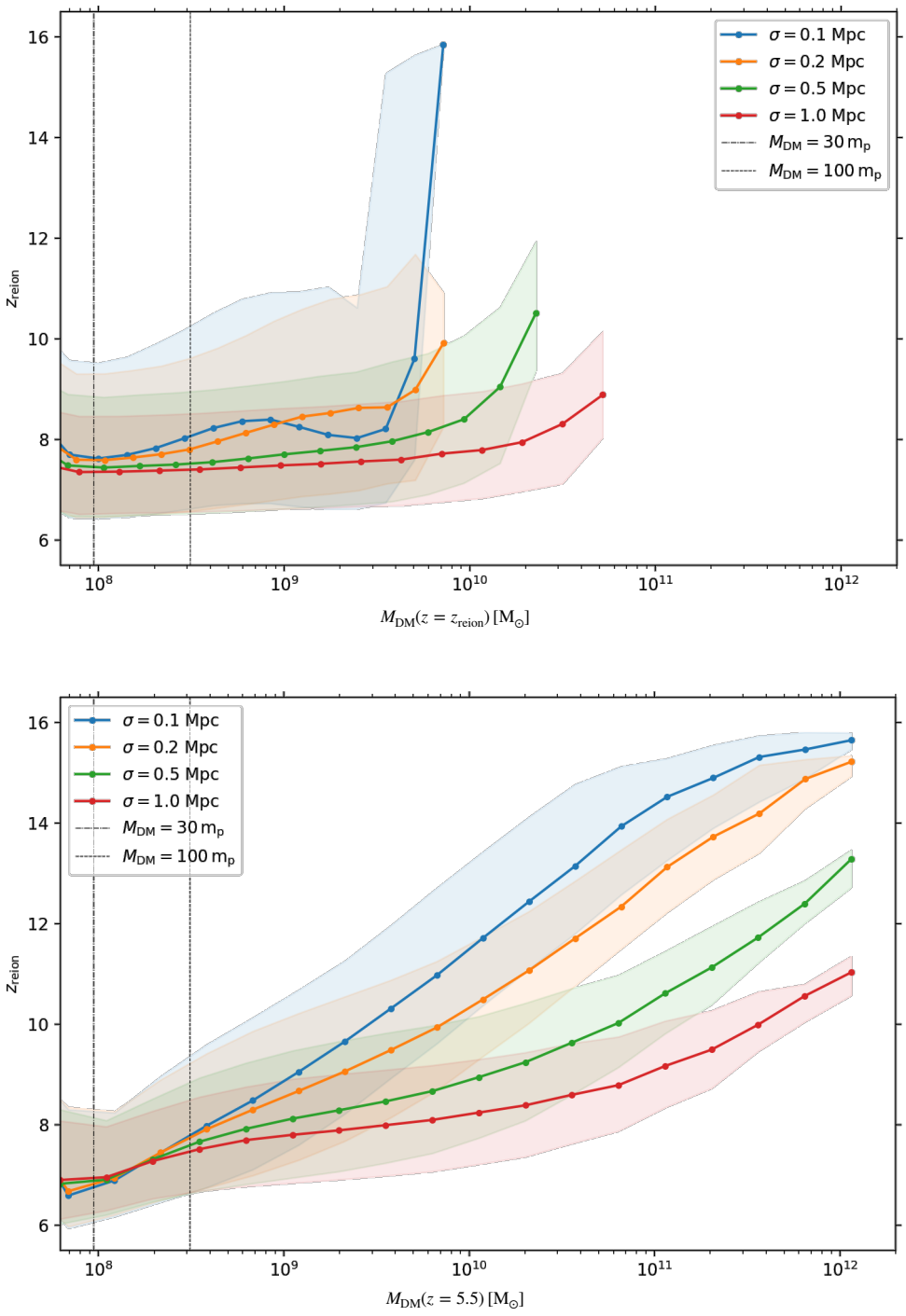

Figure 4.1. – Relation between the reionization redshit $z_{\rm reion}$ and the DM mass at reionization time $M_{\rm DM}(z = z_{\rm reion})$ (upper panel) and at the end of the simulation (lower panel) for different smoothing scales (σ = 0.1, 0.2, 0.5, 1.0 Mpc respectively as blue, orange, green and red continuous lines). We also show the threshold of 30 $m_{\rm P}$ and 100 $m_{\rm P}$ DM particles, above which we will perform the analysis. We observe a clear trend where higher values of mass accretion rates correspond to higher values of reionization redshifts, but as the smoothing scale increases, the correlation between mass accretion rate and reionization redshift becomes less pronounced.

In the top panel of Figure 4.1 we explore the relationship between $z_{\rm reion}$ and $M_{\rm reion}$ while considering different levels of smoothing. Our analysis reveals a general trend where higher values of $M_{\rm reion}$ correspond to higher values of $z_{\rm reion}$. This trend suggests that galaxies with larger masses tend to experience reionization at earlier cosmic times. However, we observe that as the smoothing scale increases, the correlation between $M_{\rm reion}$ and $z_{\rm reion}$ becomes less steep. In other words, the relationship between galaxy mass and reionization redshift becomes less pronounced with larger smoothing volumes. This effect arises because larger smoothing volumes encompass a larger portion of the surrounding environment, diluting the specific impact of the galaxy's internal properties on the reionization process.

As the smoothing scale continues to increase significantly we observe a plateau in the correlation between $M_{\rm reion}$ and $z_{\rm reion}$. This plateau indicates that for scales larger than 0.5 Mpc the specific characteristics of the galaxy have less influence on the timing of reionization. Instead, the dominant factor becomes the overall state of the surrounding environment, which becomes increasingly averaged over larger smoothing volumes. This flattening of the correlation curve and the subsequent plateau at larger smoothing scales suggest that the mid reionization point, where the galaxy's reionization is roughly halfway complete, is less affected by the galaxy's internal properties and more determined by the global properties of the reionization process.

Upon investigating the relationship between $z_{\rm reion}$ and the DM mass at the end of the simulation (bottom panel of Figure 4.1), we consistently observe that higher mass galaxies tend to undergo reionization at earlier cosmic times. This finding reaffirms the notion that the first reionized regions are the ones around larger galaxies, so reionization starts from the most biased regions in the Universe. Similar to our previous analysis, we find that the impact of smoothing on the correlation between $z_{\rm reion}$ and galaxy mass persists. However, when comparing galaxies at the same point in time (end of the simulation), the plateau in the correlation is less pronounced. The diminished visibility of the plateau can be attributed to the fact that at the end of the simulation, galaxies have already undergone substantial evolution and reionization processes, i.e. galaxies coming from large overdensities grow more, being the first to collapse, and hence ionize earlier their surroundings.

4.3.1. Reionization front and SFR

We can investigate the influence of the reionization front on the star-formation histories of galaxies by categorizing them with respect to $z_{\rm reion}$ and the mass at the end of the simulation. Galaxy quenching can occur due to various reasons and reionization represents one potential factor. For instance, the star formation activity within a galaxy might deplete its gas reservoirs, hindering the formation of new stars and leading to a decline in its star-formation rate. Similarly, if a galaxy undergoes a merger, the star formation process may continue independently of reionization thanks to the new gas reservoirs acquired. Nonetheless, the results from our analysis indicate a discernible impact of reionization on the selected galaxies, particularly evident in the reduced star-formation rates for low-mass

galaxies after reionization. The subsequent behavior of the star-formation rates depends on a multitude of factors, and while reionization plays a role, its significance may become less central over time. We show in Figure 4.2 the median SFR of galaxies with similar $z_{\rm reion}$ (rows, also the vertical shadow area) and similar DM mass at $z = 5.5$ (column). It is evident how reionization has an impact on the SFR of these galaxies, in particular on dwarf galaxies (first 2 columns).

4.3.2. Inside-out vs. Outside-in reionization scenario

The process of reionization can occur in different ways, known as inside-out and outside-in reionization. In inside-out reionization, ionization starts from a galaxy and gradually propagates outward. In outside-in reionization, ionization begins from the surrounding regions and moves inward towards the galaxy.

Inside-out reionization occurs when a galaxy forms stars, and the ionizing radiation from these stars begins to ionize the surrounding IGM. This initial ionization creates small, localized bubbles around the galaxy. As more stars form and emit ionizing radiation, these bubbles expand and merge, gradually encompassing larger volumes of the IGM. This process continues as ionized regions propagate outward from the galaxy, leading to the progressive ionization of the surrounding environment. By studying the reionization redshift at different smoothing scales, which provide information about the volume considered around the galaxy, we can observe the sequential ionization around the galaxy, reflecting the inside-out progression. On the other hand, outside-in reionization occurs when a galaxy is primarily ionized by a nearby source of ionizing radiation. The ionization front originates in regions that are relatively far away from the galaxy and gradually moves inward toward the galaxy's location.

By examining the reionization redshift at different smoothing scales, we can investigate the characteristics of inside-out and outside-in reionization. The smoothing scales we consider in this work are $\sigma = 0.2$, 0.5 Mpc, large enough to neglect the contribution of HI inside the galaxy and small enough to study the environment around the galaxy.

In Figure 4.3, we present an illustration of the inside-out scenario (left) and the outside-in scenario (right). For each case, we depict the DM and stellar mass accretion histories of the galaxy on the left. We also show the redshift at which the first star forms in the galaxy ($z_{\rm star}$), the reionization redshifts ($z_{\rm reion}^{0.2\,\rm Mpc}$, $z_{\rm reion}^{0.5\,\rm Mpc}$) and the threshold of $100\,m_{\rm P}$ that defines the formation of a DM halo in the simulation. On the right side, we display the galaxy's position in the HII fraction field of a slice through the simulation box at selected redshifts, with the dimension along the line-of-sight fixed. Zooming closely to the galaxy's position, we indicate the spherical volumes[1] corresponding to the smoothing scales at $\sigma = 0.2, 0.5$ Mpc. These scales are significantly larger than the typical size of a galaxy, which corresponds to roughly one pixel in the HII fraction field.

[1]The effective volume covered by a 3-dimensional Gaussian filter with standard deviation σ is $V_G = (\sigma\sqrt{2\pi})^3$. This corresponds to a spherical volume with radius $r \simeq 1.55\sigma$.

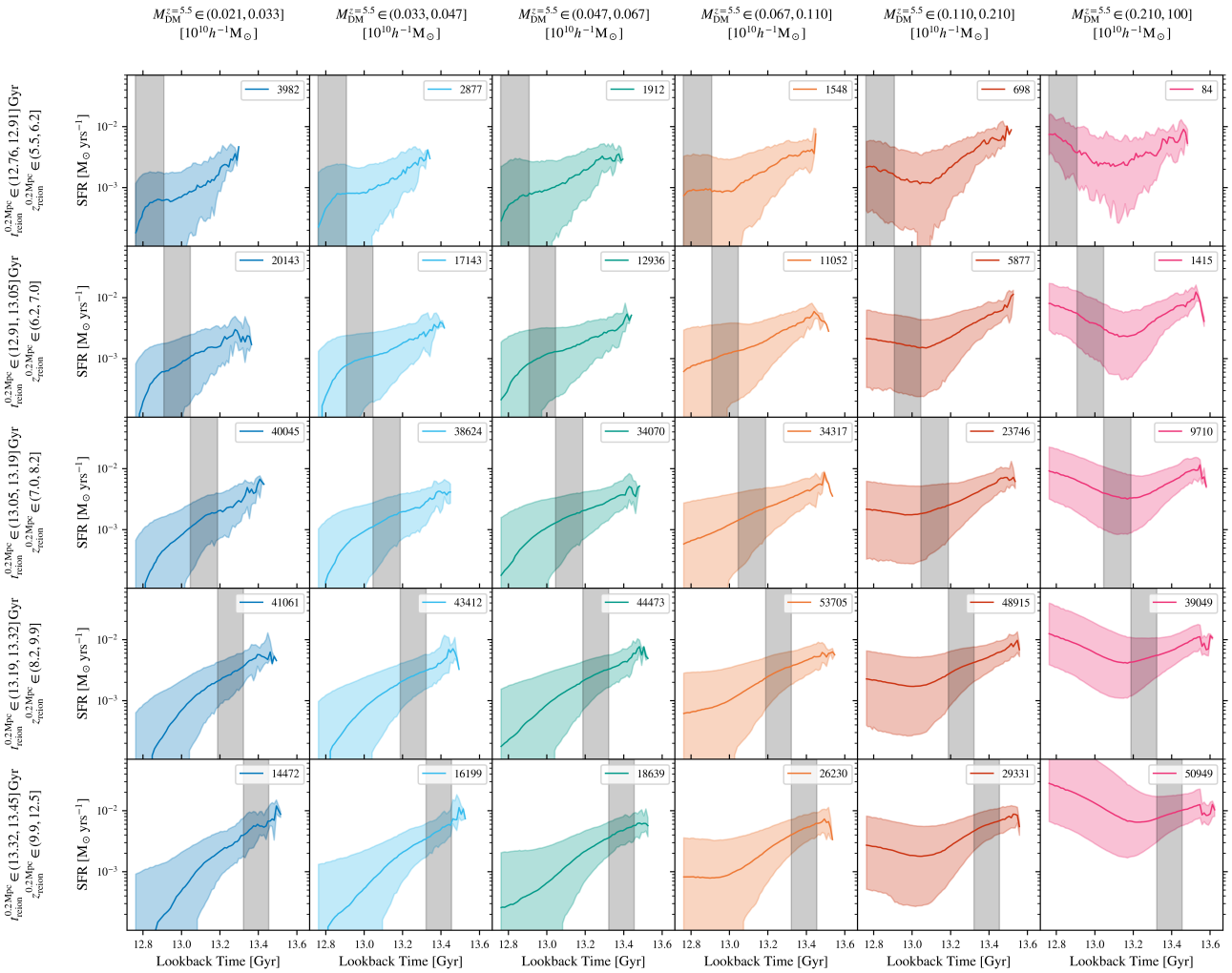

Figure 4.2. – Median SFR of galaxies with similar $z_{\rm reion}$ and DM mass at $z = 5.5$ as a function of the lookback time. We present the median SFR of galaxies binned into rows based on similar $z_{\rm reion}$ (also indicated by the vertical shadow area) and columns based on similar mass. The colored shadow area represents the 15th and 85th percentiles of the data distribution and for each panel the number of galaxies satisfying that combination are reported. In addition to the $z_{\rm reion}$ binning, the corresponding values for the lookback time are reported. The figure clearly illustrates the impact of reionization on the SFR of these galaxies, particularly on dwarf galaxies (first 2 columns).

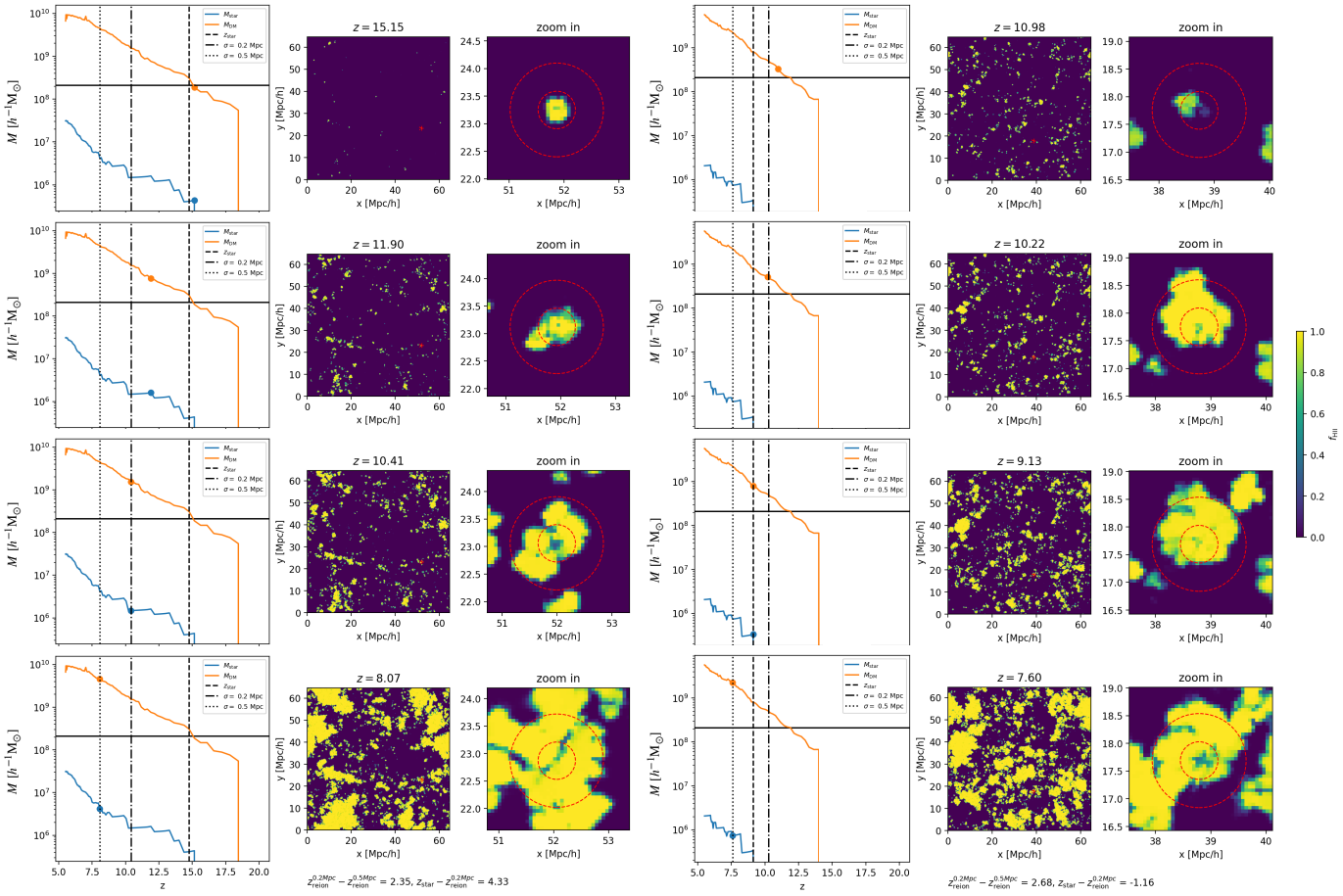

Figure 4.3. – Inside-Out and Outside-In Reionization Scenarios. The figure presents an illustration of the inside-out scenario (left) and the outside-in scenario (right) for selected galaxies. On the left, we display the DM (blue continuous line) and stellar (orange continuous line) mass accretion histories of the galaxy, with indications of the redshift at which the first star forms in the galaxy ($z_{\rm star}$, dashed black line) and the reionization redshifts ($z_{\rm reion}^{0.2,\rm Mpc}$, dash-dotted black line; $z_{\rm reion}^{0.5,\rm Mpc}$ dotted black line). Additionally, we include the threshold of $100\,m_{\rm P}$ as a black continuous line that defines the formation of a DM halo in the simulation. On the right side, we show the galaxy's position in the HII fraction field at specific redshifts (the dimension along the line-of-sight is kept fixed). To provide further insight, we zoom closely to the galaxy's position and indicate the spherical volumes corresponding to the smoothing scales at $\sigma = 0.2, 0.5$ Mpc (red dashed line). These scales are significantly larger than the typical size of a galaxy, which corresponds to roughly one pixel in the HII fraction field.

We define inside-out reionized galaxies the ones with $z_{\rm star} > z_{\rm reion}^{0.2\,\rm Mpc} > z_{\rm reion}^{0.5\,\rm Mpc}$, corresponding to the first quadrant in Figure 4.4. Here we display the distribution of the galaxies selected by combinations of $z_{\rm star}, z_{\rm reion}^{0.2\,\rm Mpc}, z_{\rm reion}^{0.5\,\rm Mpc}$. The upper part of the figure shows galaxies for which a small surrounding volume is ionized before a larger one; the right part galaxies who form stars before the surroundings are ionized. The contour lines indicate the number distribution of galaxies, showing that the majority undergoes an outside-in reionization scenario. The color-code informs about the mean DM mass of galaxies with the same characteristics, showing that high-mass galaxies tend to reionize from the inside-out.

All the other cases display an outside-in reionization scenario, in particular:

- $z_{\rm reion}^{0.2\,\rm Mpc} > z_{\rm reion}^{0.5\,\rm Mpc} > z_{\rm star}$, the galaxy is ionized by a nearby source (second quadrant in Figure 4.4),
- $z_{\rm reion}^{0.5\,\rm Mpc} > z_{\rm reion}^{0.2\,\rm Mpc} > z_{\rm star}$, the galaxy is ionized by a distant source (third quadrant in Figure 4.4),
- $z_{\rm star} > z_{\rm reion}^{0.5\,\rm Mpc} > z_{\rm reion}^{0.2\,\rm Mpc}$, the galaxy form a star but the photons emitted are not energetic enough to ionize the surroundings (fourth quadrant in Figure 4.4).

4.3.3. Environment

There are clear differences between the galaxy populations we outlined, particularly concerning their cosmic environment. Galaxies reionized inside-out, for instance, display a preference for occupying low-density regions in the Universe. In contrast, galaxies reionized outside-in are more commonly found in higher density environments. This environmental distinction plays a crucial role in shaping the star-formation histories of these galaxies.

Low-density regions provide favorable conditions for the development of galaxies reionized inside-out, since the expansion of the reionization front originates from the central regions of these galaxies, propagating outward. Conversely, in higher-density environments, the ionizing photons responsible for galaxy ionization can originate from neighboring galaxies. The clustering of galaxies in these denser regions leads to more complex interactions and a higher probability of sharing ionizing photons. Consequently, the reionization process in higher-density regions may involve contributions from multiple galaxies.

We show in Figure 4.5 the distribution of the underlying density for galaxies reionized inside-out and outside-in (according to the classification made in Section 4.3.2). We compute the mean density in spherical volumes centered in the galaxy with radii of R = 0.5, 1.0 Mpc, at the time of reionization $z_{\rm reion}^{0.2\,\rm Mpc}$. In general, inside-out galaxies exhibit a lower density in their environment, while outside-in galaxies a larger one. However, it is also interesting to notice the differences within the same population. For example,

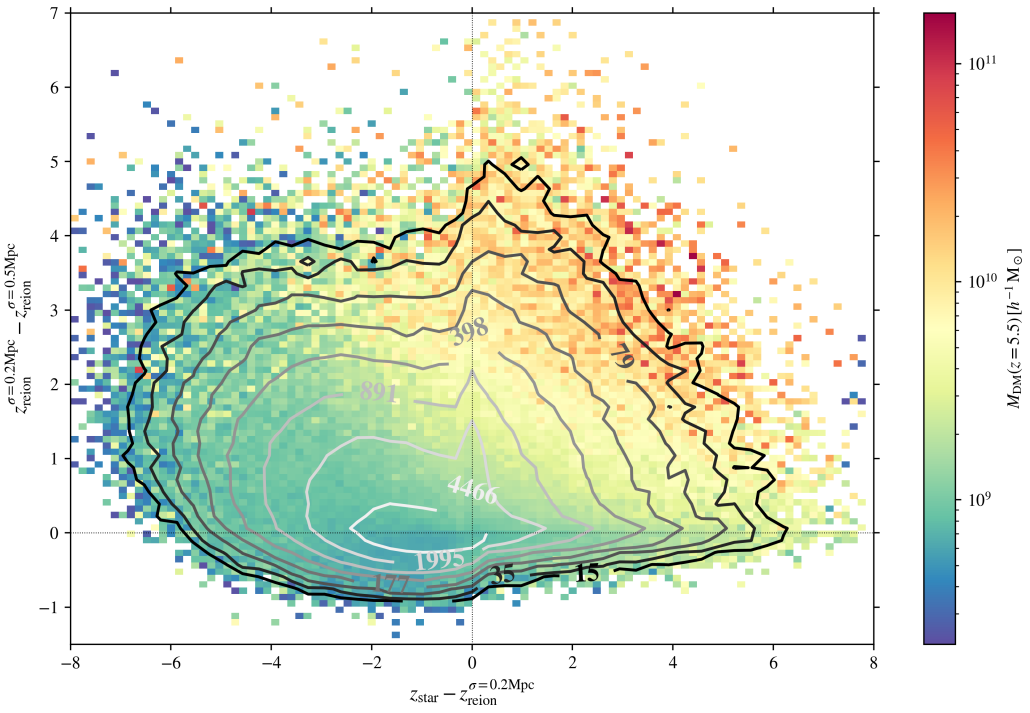

Figure 4.4. – Distribution of galaxies selected based on combinations of $z_{\rm star}$, $z_{\rm reion}^{0.2,{\rm Mpc}}$, and $z_{\rm reion}^{0.5,{\rm Mpc}}$. The color-code refers to the mean DM mass at $z = 5.5$ for galaxies with similar features. The upper part of the figure illustrates galaxies in which a smaller surrounding volume is ionized before a larger one, while the right part shows galaxies that form stars before their surroundings are ionized. The contour lines represent the number distribution of galaxies.

galaxies ionized by a nearby galaxy (light blue) tend to be on a lower-density region than galaxies ionized by a galaxy more far away (petrol).

Understanding the environmental influences on galaxy formation and reionization processes provides critical insights into the intricate interplay between galaxies and their surroundings. By examining the distribution and behavior of inside-out and outside-in recognized galaxies across different cosmic environments, we can unravel the impact of local and global factors on their star-formation histories during the epoch of reionization.

4.4. To be quenched or not to be quenched

Before proceeding to classify galaxies as quenched, it is important to define consistently what quenched means. Operatively defining quenched and non-quenched galaxies involves setting specific criteria or thresholds based on observable properties of galaxies. The criteria to distinguish between actively star-forming galaxies and those with suppressed star formation are based both on the specific observational data available and on the scientific questions being addressed. Different criteria can lead to slightly different samples of

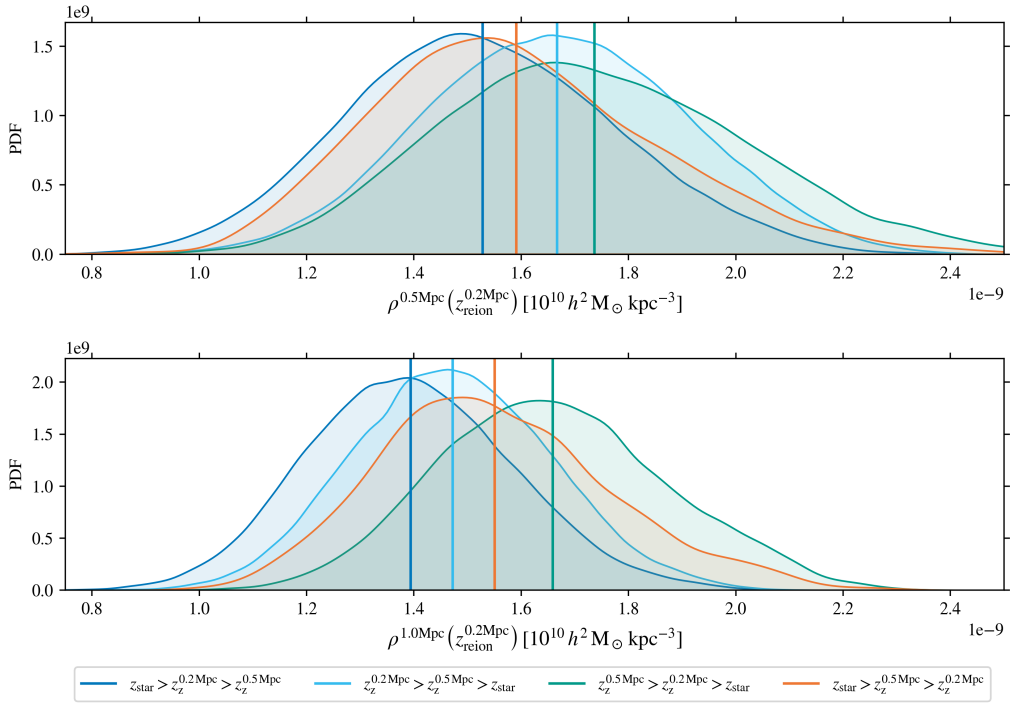

Figure 4.5. – The distribution and its mean of the underlying density for inside-out and outside-in galaxies. The average density is computed within spherical volumes centered on the galaxy with radii of $R = 0.5,\ 1.0$ Mpc, at the time of reionization $z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}$. Inside-out galaxies (in blue) generally exhibit a lower density in their environment, while outside-in galaxies tend to have a higher density. Notably, differences are also observed within the same population. Galaxies ionized by a nearby galaxy (light blue) typically reside in lower-density regions compared to galaxies ionized by galaxies located farther away (petrol).

quenched galaxies, and the results should be interpreted in the context of the adopted definition. Also, quenching may not be an abrupt process, and some galaxies may exhibit intermediate properties between quenched and non-quenched states. Here we list some common approaches to operatively define quenched and non-quenched galaxies:

- SFR Threshold. One of the simplest methods is to set a threshold on the galaxy's SFR. Galaxies with SFRs below a certain value are classified as quenched, while those above it are considered actively star-forming (e.g. Looser et al., 2023).

- Specific Star Formation Rate (sSFR). The sSFR, defined as the SFR normalized by the stellar mass, is another parameter used for quenching classification. A low sSFR indicates a quenched galaxy, while a high sSFR corresponds to a non-quenched one (e.g. Ahad et al., 2023).

- Balmer Absorption Lines. The strength of Balmer absorption lines in galaxy spectra can be used as an indicator of recent star formation. Weak Balmer lines suggest

a quenched galaxy, while strong Balmer lines imply active star formation (e.g. Domínguez et al., 2015).

Moreover, various mechanisms can induce quenching in galaxies, and these mechanisms can lead to distinct timescales for the suppression of star formation. For instance, processes involving the removal of gas can rapidly quench star formation within a relatively short span of 10 to 100 million years; on the other hand, mechanisms that hinder the accretion of fresh gas necessitate longer quenching timescales, spanning from 100 million years to 1 billion years (Croton et al., 2006; Fabian, 2012; King & Pounds, 2015; Whitaker et al., 2021; Williams et al., 2021). Consequently, these different quenching mechanisms leave distinct imprints on the chemical enrichment levels of galaxies. Such complexities necessitate a comprehensive understanding of how these mechanisms operate at different cosmic epochs.

To examine the specific effects of reionization on galaxies shortly after it occurs, we focus on the immediate aftermath rather than studying events much later. In order to achieve a robust classification of galaxies, we explored various approaches and found that dividing them into four distinct groups provide the most stable results. This classification is based on computing the SFR at the time of reionization (more specifically, at $z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}$) and after specific intervals of 50 and 150 Myr later (based on the recent findings in Looser et al., 2023). By averaging the SFR over a few $\times 10$ Myr intervals around these timescales, we minimize fluctuations and obtain more reliable groupings. The four types are defined as follows:

- Type 0: Galaxies with extremely suppressed star formation after both 50 Myr and 150 Myr from reionization, indicated by

$$\frac{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}} + 50\,\mathrm{Myr}\right)}{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}\right)} < 0.1 \wedge \frac{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}} + 150\,\mathrm{Myr}\right)}{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}\right)} < 0.1\,. \tag{4.2}$$

- Type 1: Galaxies that experience strong suppression of star formation 50 Myr after $z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}$ but show some recovery at 150 Myr, i.e.

$$\frac{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}} + 50\,\mathrm{Myr}\right)}{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}\right)} < 0.1 \wedge \frac{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}} + 150\,\mathrm{Myr}\right)}{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}\right)} > 0.1\,. \tag{4.3}$$

- Type 2: Galaxies that maintain a relatively high SFR both 50 Myr and 150 Myr after $z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}$, as indicated by

$$\frac{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}} + 50\,\mathrm{Myr}\right)}{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}\right)} > 0.1 \wedge \frac{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}} + 150\,\mathrm{Myr}\right)}{\mathrm{SFR}\left(z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}\right)} > 0.1\,. \tag{4.4}$$

- Type 3: This category comprises the remaining galaxies that do not fall into the previous groups. Specifically, these are galaxies where star formation initiates much later than the time of reionization, even though the underlying DM halo existed before the region was ionized[2].

This classification allows us to investigate the immediate impact of reionization on galaxies' star formation, enabling us to draw meaningful insights into the interconnection between reionization and the subsequent evolution of galaxies in the early Universe. We show the median SFRs of galaxies classified in the four groups in Figure 4.6. Here, we observe a column-wise arrangement of galaxies with similar masses, while the rows represent different types of SFR behavior. We notice distinct patterns in the distribution of galaxies with varying SFR responses to reionization.

Starting with galaxies of type 2, which are unaffected by reionization, we observe that their percentage increases with mass. This aligns with our expectations, as larger galaxies possess a higher gas accretion rate, making them less susceptible to the disruptive effects of reionization. On the other hand, galaxies of type 3, which begin their star formation significantly after reionization, show a diminishing trend with increasing mass. The reason behind this trend lies in the fact that more massive galaxies tend to initiate star formation earlier due to their ability to accumulate gas and form stars more rapidly, resulting in an inside-out type of reionization. It is noteworthy to mention that the percentage of galaxies in type 0, experiencing complete quenching due to reionization, is found to be very low. This observation may be attributed to the mass range covered in the simulation, which might not be sufficient to effectively investigate the smaller galaxies, as they tend to feel the effects of reionization more significantly. Despite the low occurrence of type 0 galaxies, we find a consistent number of galaxies in type 1. These galaxies are initially impacted by reionization, leading to a suppression of star formation. However, they manage to recover and form stars after a delay of at around 100 Myr. This category of galaxies is of particular interest for our study, as it highlights the correlation between reionization and subsequent star formation, revealing the diverse and intricate processes shaping the evolution of galaxies in the early Universe.

4.5. How to observe a quenched galaxy

In this section, we focus on four key observables that play a crucial role in characterizing quenched galaxies: stellar mass ($M_{\rm star}$), star metallicity ($Z_{\rm star}$), gas metallicity ($Z_{\rm gas}$), and the specific star-formation rate. Each of these quantities provides unique information about the past and present star-forming history, the chemical composition, and the ongoing star formation activity of a galaxy. We briefly explore how these observables are obtained and their significance in unraveling the intricate processes behind quenching.

[2]Let us remark that in Section 4.3 we excluded from the analysis galaxies whose DM halo formed after the region was ionized, while in this case the DM halo formed already but the galaxy did not produce any star until much later after reionization.

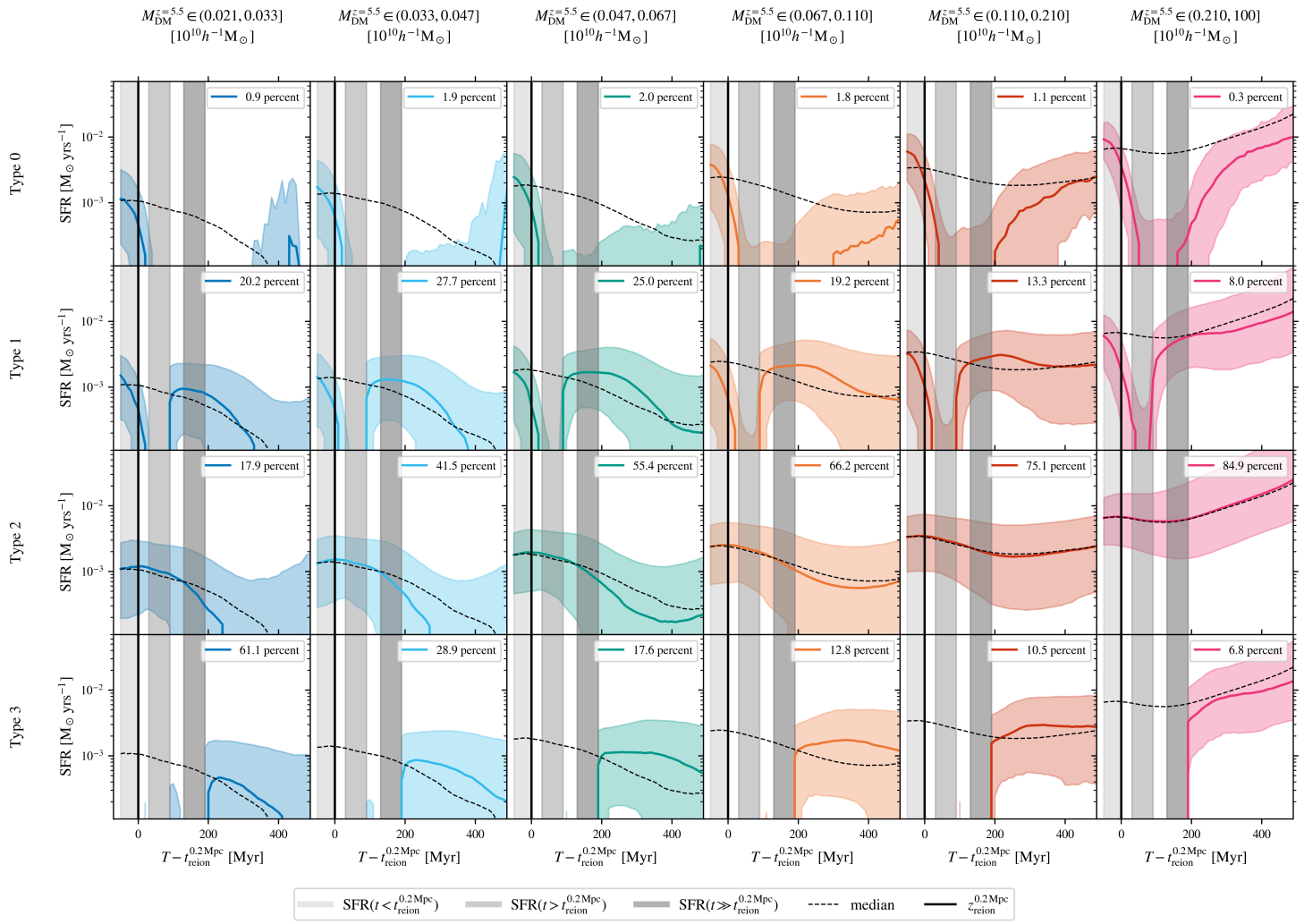

Figure 4.6. – Galaxy response to reionization: diverse SFHs in the early Universe. The figure presents galaxies grouped by similar masses in columns, with each row representing different types of SFR behavior. Each panel shows the median SFR (continuous colored line) and the 15th and 85th percentiles (colored shadow area) as a function of time after reionization for galaxies with the same features. The median SFR of all galaxies with similar mass is displayed in each panel as black dashed line. The vertical bands indicate the time intervals used to compute the SFR, at reionization (light grey), after 50 Myr (medium grey), and after 150 Myr (dark grey). The figure also includes the percentage of galaxies belonging to each typology within a given mass range.

- M_{star}. Stellar mass can be determined through various observational methods. One common approach is to use multi-band photometry to fit the spectral energy distribution (SED) of the galaxy, which provides an estimate of the stellar mass-to-light ratio (Cid Fernandes et al., 2005; Mitchell et al., 2013). Moreover, by studying the dynamics of stars within the galaxy it is possible to infer its total mass (Puchwein & Springel, 2013). Stellar mass is a crucial quantity because it gives us insights into the total amount of stellar material formed and helps classify galaxies into different mass regimes, such as low-mass dwarf galaxies or massive ellipticals.

- Z_{star}. The star metallicity of galaxies is primarily derived through spectroscopic observations, by studying the spectral absorption lines of stars and measuring the abundance of heavy elements relative to hydrogen (Sánchez-Blázquez et al., 2006; Peeples & Somerville, 2013). Star metallicity is essential because it reflects the galaxy's chemical enrichment history, revealing information about its past star formation episodes and the contribution of different stellar populations.

- Z_{gas}. Observing gas metallicity is also achieved through spectroscopy. Emission lines from ionized gas, such as HII regions or planetary nebulae, provide valuable information about the gas composition. These emission lines arise from the recombination of ionized atoms, and their relative strengths enable astronomers to estimate the gas metallicity. Gas metallicity is crucial because it indicates the level of chemical enrichment and the capacity of the galaxy to form new stars. Low gas metallicity may imply that the galaxy has not undergone many generations of star formation or has experienced strong outflows.

- sSFR. To measure the specific star formation rate, information on both star formation rates and stellar masses are needed. Star formation rates are often determined using observations of ultraviolet, infrared, or radio emission associated with ongoing star formation (Domínguez et al., 2015; Barnes et al., 2017). Combining this with the stellar mass measurement provides the sSFR. sSFR is important because it characterizes a galaxy's current level of star formation relative to its mass.

Overall, these observed quantities play a critical role in understanding the life cycle of galaxies. They provide insights into their past and present star formation activities, their chemical composition, and their interactions with their surrounding environments. By studying these properties across different galaxy populations and cosmic epochs, we gain a deeper understanding of galaxy formation, evolution, and the underlying physical processes driving their diverse behaviors in the Universe.

4.6. Results

We present in this Section the main results of our study. We have categorized galaxies into four groups according to their SFR, as explained in Section 4.4, and selected four

observables for these galaxies, as described in Section 4.5. By combining these quantities in Figure 4.7, we gain insight into the interplay between galaxy properties and SFR. The figure displays the fraction of galaxies for each category across their stellar mass, star and gas metallicity, and specific star formation rate.

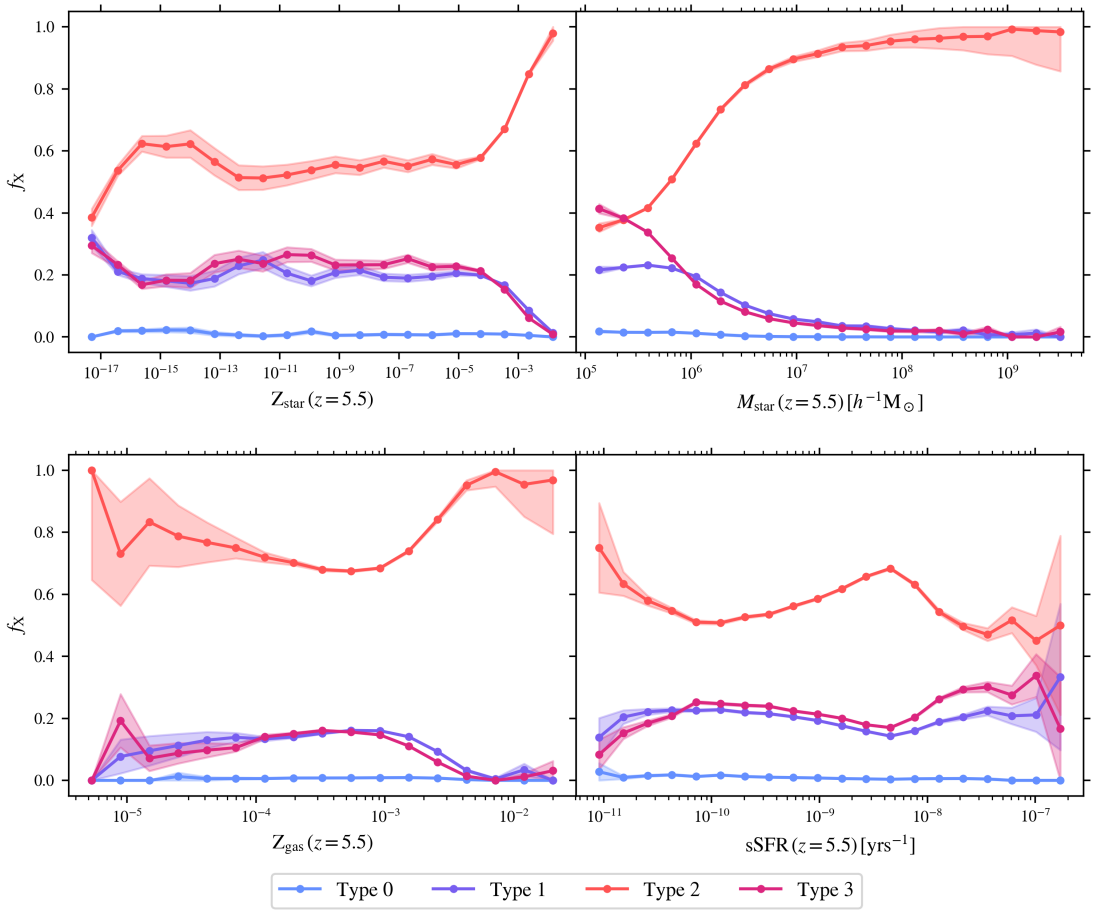

Figure 4.7. – The fraction of galaxies for each category is shown as a function of different observables. Each panel corresponds to a specific observable (stellar mass, star metallicity, gas metallicity, and specific star formation rate) computed at redshift 5.5, and the colored lines (type 0 in light blue, type 1 in purple, type 2 in red, and type 3 in magenta) represent the fraction of galaxies belonging to each category across the range of that observable.

Type 0 galaxies, experiencing complete quenching due to reionization, are found to be a minority in our study. Surprisingly, their distribution does not appear significantly influenced by the observables we have selected. To delve further into galaxies entirely quenched by reionization, we must explore galaxies with lower masses. Type 1 galaxies are particularly intriguing, and their dependence on galaxy features offers valuable insights. We observe that their fraction increases for galaxies with low stellar mass, as well as low stellar and gas metallicity. In these cases, the galaxies lack sufficient material to

sustain star formation after the gas is heated by the reionization front. Essentially, these galaxies form all the stars they can during the initial phase and then exhaust their available resources, leading to quenching at the redshfit we observe. Interestingly, the fraction of type 1 galaxies seem relatively independent of their sSFR. The fraction remains roughly constant, which could be attributed to their low SFR and low stellar mass, resulting in a consistent ratio between these quantities. Type 2 galaxies remain unaffected by reionization entirely, and their fraction increases notably for galaxies with higher stellar mass and stellar and gas metallicity. These galaxies possess sufficient material to sustain star formation, undisturbed by the reionization process. On the other hand, type 3 galaxies initiate star formation significantly later after reionization. We observe that they follow a trend similar to type 1 galaxies, where the fraction of type 3 galaxies increases for galaxies with low stellar mass and stellar and gas metallicity. This behavior is consistent with the idea that galaxies with fewer resources, both in terms of mass and metallicity, are more prone to delay star formation after reionization.

Moreover, we explore the connection between the selected observables by constructing a 2-dimensional histogram in a corner plot in Figure 4.8. In particular, we investigate how the fraction of type 1 galaxies varies for different combinations of observables. We observe a higher fraction of type 1 galaxies for galaxies with low stellar mass and low metallicity, which aligns with our expectations. However, we also identify some intriguing combinations in the observables. For instance, the fraction of type 1 galaxies is rather high (approximately 0.3) for galaxies with low stellar mass but high stellar metallicity, appearing in the upper left corner of the central panel. Similarly, we notice a distinct group of galaxies with very low sSFR but high gas metallicity in the lower right corner of the lower right panel. Additionally, there is another interesting subgroup of galaxies with low stellar mass but high gas metallicity located in the lower left corner of the upper left panel. The large scatter in the data, observed in particular in the Z_{star} − sSFR panel, could be reduced by filtering bins with low galaxy number count. However, we decide to show here the data unfiltered to emphasize the complexity of the problem and provide valuable insights into the intricate interaction between galaxy features and the quenching process.

4.7. Summary

Throughout our study, we delved into the fascinating phenomenon of quenching SF in galaxies and its connection to the epoch of reionization. Quenching refers to the suppression or halt in star formation within galaxies, and it can occur due to various mechanisms, including gas removal and starvation processes. In principle, when reionization happens, the gas within these galaxies is heated and dispersed, halting the star formation process. Our research focused on understanding the impact of reionization on galaxy formation and evolution, and to achieve this, we utilized state-of-the-art hydrodynamical simulations, namely the THESAN suite of simulations.

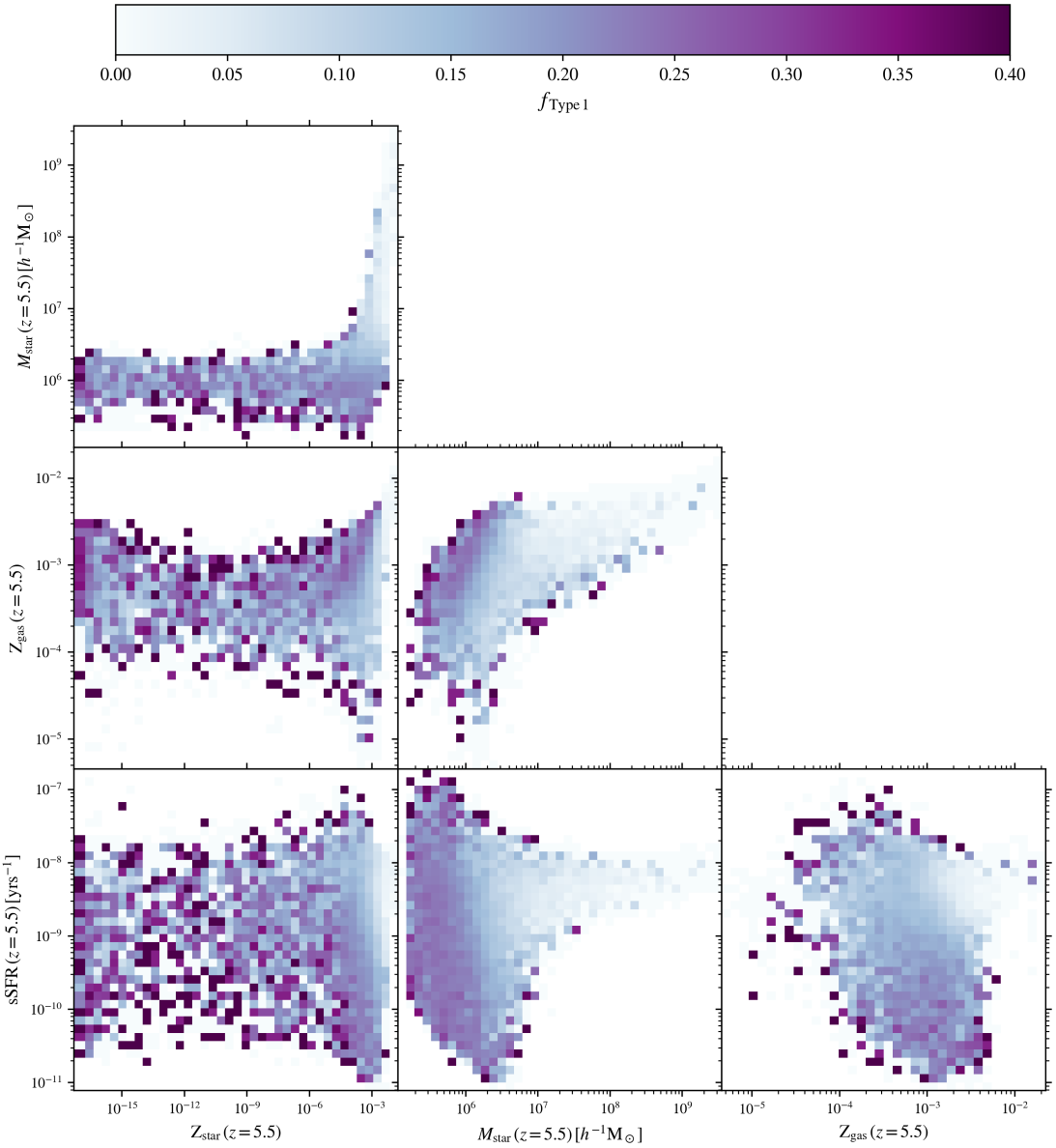

Figure 4.8. – Exploring the relationship between selected observables for type 1 galaxies. The figure presents 2-dimensional histograms, illustrating the fraction of type 1 galaxies, across different combinations of the four selected observables.

We explored the effects of different reionization mechanism on galaxies, particularly focusing on the redshifts of reionization ($z_{\mathrm{reion}}^{0.2\,\mathrm{Mpc}}$ and $z_{\mathrm{reion}}^{0.5\,\mathrm{Mpc}}$) and their relation to galaxy properties. We found that the reionization redshifts had a discernible impact on the star formation histories of galaxies, particularly for low-mass galaxies. Inside-out galaxies tend to be located in low-density regions of the Universe, where the reionization front expands from the central region outward. The first stars formed in these galaxies play a

crucial role in ionizing the surrounding medium. On the other hand, outside-in galaxies are more prevalent in higher-density environments, where nearby galaxies may contribute to the ionization process. Our results also indicated that galaxies ionized by nearby galaxies tend to be in lower-density regions than galaxies ionized by more distant galaxies. This suggests that the environment plays a significant role in the reionization process and its impact on galaxies.

We then developed classification schemes to categorize galaxies based on their SFR at different epochs relative to reionization. We identified four main types of galaxies based on their SFR behavior after reionization. Type 0 galaxies experienced complete quenching due to reionization, but their fraction was found to be relatively low. Type 1 galaxies exhibited a suppression of star formation shortly after reionization but managed to recover and form stars after a delay of around 100 Myr. These galaxies were of particular interest as they revealed the correlation between reionization and subsequent star formation. Type 2 galaxies showed no significant impact from reionization, indicating a robustness to this cosmic event. Their percentage increased with galaxy mass, suggesting that larger galaxies with higher gas accretion rates were less affected by reionization. Type 3 galaxies initiated star formation significantly later than the time of reionization, and this behavior diminished with increasing galaxy mass, indicating an inside-out type of reionization in more massive galaxies. We also explored the connections between the fractions of type 1 galaxies and various observables, such as stellar mass, star and gas metallicity, and specific star formation rate. We observed interesting trends in these connections, such as a higher fraction of type 1 galaxies for galaxies with low stellar mass and low metallicity. Additionally, we noticed intriguing combinations, like the fraction of type 1 galaxies being around 0.3 for galaxies with low stellar mass but high stellar metallicity, or with very low specific star formation rate but high gas metallicity.

Determining the exact cause of a galaxy's quenching through observations is very difficult. We present a robust and straightforward tool that complements observations by predicting the quenching probability for galaxies based on their observed features. We also gain insights on the environmental role in the quenching mechanism, connecting it to thee underlying DM density.

Our findings shed light on the intricate correlation between reionization, galaxy properties, and the subsequent star formation history. This study contributes to our understanding of how reionization impacts the formation and evolution of galaxies in the early Universe, offering valuable insights into the intricate processes that have shaped cosmic structures throughout cosmic history.

4.8. Challenges and next steps

We discuss here next steps and challenges in the study of reionization-quenched galaxies.

Refine and Expand the Analysis. Continuously refine and improve the classification and analysis of galaxies based on their star formation histories and the impact of reionization.

This may involve exploring different time intervals for averaging SFR, considering alternative quenching criteria, and incorporating more sophisticated statistical techniques to minimize uncertainties and fluctuations in the results. Additionally, expand the analysis to include a broader range of galaxy properties and parameters to better understand the diverse quenching mechanisms at play. Furthermore, it is crucial to conduct a multi-dimensional fit that predicts the galaxy's typology based on its observational features. By employing advanced statistical techniques and machine learning algorithms, we can develop a robust model that accurately classifies galaxies into their respective categories.

Incorporate Machine Learning Techniques. Utilize machine learning techniques to analyze the dataset and identify complex correlations and patterns that might be challenging to detect using conventional analysis methods. Machine learning can help uncover subtle relationships between different observables and reveal hidden trends that can provide deeper insights into the quenching mechanisms and their connections to reionization. We have already made use of the `K-Means`[3] and `Linear SVC` algorithms to classify quenched galaxies and the `RandomForestRegressor` and `PCA` algorithms to analyze which galaxy's features might influence quenching. These preliminary results are not showed here, going beyond the scope of this project.

Explore Additional Observables. Broaden the investigation to include a wider range of observables that could provide deeper insights into the quenching and reionization effects on galaxies. Consider properties such as galaxy morphology, gas content, dust extinction, and kinematics, as these factors play critical roles in shaping galaxy evolution. Integrating these additional observables into the analysis can lead to a more comprehensive understanding of the correlation between reionization and galaxy quenching.

Include Simulations with Different Parameters. Conduct simulations with varying cosmological and astrophysical parameters to assess the robustness of the findings. Investigate how changes in parameters related to cosmology, feedback processes, and galaxy formation affect the observed trends. By exploring a broader parameter space, it is possible to identify the key drivers and sensitivities of the quenching process.

Comparisons with Observations. Validate the simulation results by performing detailed comparisons with observational data from large galaxy surveys and other astrophysical observatories. Ensure that the simulated galaxies reproduce the observed properties, redshift distributions, and quenching trends seen in the real Universe. Address any discrepancies between the simulations and observations, and refine the model accordingly.

The refined analysis, expanded parameter exploration, and incorporation of machine learning methods can collectively contribute to a more comprehensive and accurate model of the intricate processes that shaped the evolution of galaxies and the Universe as a whole.

[3] Refer to https://scikit-learn.org/stable/ for all the algorithms cited.

CHAPTER 5

Conclusions

This thesis has presented a comprehensive exploration of the EoR and its implications for our understanding of the early Universe. The EoR represents a crucial phase in the cosmic history, marking the transformation of neutral elements into ionized species. This intricate process is primarily driven by photons emitted from the earliest luminous sources, making the understanding of the EoR tightly intertwined with the formation and evolution of these sources. To comprehend the dynamics of the EoR, it becomes imperative to gain insights into the growth of cosmic structures and the intricate interplay between galaxies and the intergalactic medium. In this work, we have delved into various aspects of this phenomenon, aiming to unravel the broader picture.

The motivation behind this thesis also arises from the multitude of ongoing and upcoming experiments dedicated to reveal the heart of the EoR. We employ the LIM technique, a powerful tool that allows us to survey large volume of the sky without the need of resolving each object. LIM enables us to study the HI and HeII atoms through the detection of their hyperfine transition. The utilization of LIM promises to advance our understanding of the early Universe, enriching our knowledge of the cosmic dawn and the critical processes shaping the cosmos during this pivotal period.

Through three distinct projects, namely the investigation of 21 cm intensity mapping to constrain the HIHMR, the study of helium reionization using intensity mapping techniques and the investigation of the quenching of dwarf galaxies after reionization, we have made significant progress in addressing key questions regarding the nature and evolution of cosmic structures during this pivotal epoch.

The HI-halo mass relation at $z \sim 1$

The first project of this thesis focused on simulating 21 cm intensity mapping to study and constrain the HIHMR, a crucial aspect in comprehending the formation and evolution of galaxies.

We employed synthetic data derived from N-body simulations to construct realistic 21 cm mock intensity maps at $z \simeq 1$. The HI distribution was represented using a parameters-dependent analytical expression overlaid on the DM halos. Instrumental effects, such as telescope beam smoothing, instrumental thermal noise, and foregrounds removal, were carefully considered to ensure accurate representation. The resulting mock intensity maps offer a valuable tool for studying the large-scale structure of the Universe and probing the HIHMR with precision. Our analysis, based on the Minkowski functionals, has allowed us to characterize the spatial distribution of the HI signal and probe the underlying physical mechanisms driving the HIHMR. This statistical tool quantify the geometry and topology of brightness-temperature isocontours in a single frequency channel.

Utilizing the Fisher information formalism, we forecast constraints on the parametric model for the HIHMR based on a 20,000 deg^2 survey conducted with the Square Kilometre Array Phase 1 (SKA-1) MID observatory in single-dish mode. With a channel bandwidth of 2 MHz, integration times of a few $\times 10^4$ s per pointing allow us to image the smoothed HI distribution at redshift $z \simeq 1$ and efficiently constrain the HIHMR parameters from the Minkowski functionals. Additionally, we find that incorporating an independent measurement of the mean HI density provides tighter constraints on some parameters. By combining results from different frequency channels, we obtain exquisite constraints on the evolution of the HIHMR, particularly in the central frequency range of the data cube. This approach holds significant promise for furthering our understanding of galaxy formation and evolution during the epoch of reionization and provides valuable insights into the connection between HI content and DM halos. Ultimately, our work contributes to the broader field of cosmology and deepens our understanding of the intricate processes that shaped the early Universe.

When HeII goes to HeIII

The second project delved into the study of helium reionization, a key process during the epoch of reionization that is intricately linked to the evolution of the intergalactic medium and the formation of the first luminous sources, for which robust constraints on its underlying drivers and endpoint are still lacking.

In our investigation, we explore the implications of different reionization models on the helium reionization process. In the early model, we consider a scenario where QSOs play a prominent role in driving reionization, due to their high-energy photons capable of ionizing both hydrogen and helium atoms. To accurately account for the contribution of quasars, the simulations we analyze implement the quasar luminosity function from

Giallongo et al. (2015), providing a robust framework for understanding the ionizing effects of these energetic sources at high redshifts ($z > 5$). We also examine the standard and more conservative late reionization model, where AGNs become significant contributors from a redshift of $z = 5$. This approach ensures that we thoroughly consider the impact of AGNs in the later stages of the EoR, providing a valuable contrast to the early quasar-driven model.

We obtain 3.46 cm mock intensity maps implementing the same techniques we used in our first project, therefore accounting for the instrumental effects of three radio surveys (SKA-1 MID, DSA-2000, PUMA). We investigate the diverse reionization models and their respective impacts on the power spectrum of HeII intensity maps. Remarkably, our findings demonstrate that such instruments hold the potential to offer valuable constraints on the timeline of helium reionization and prove the feasibility of using intensity mapping to probe the signatures (sources and end-point) of helium reionization.

Observing reionization-quenched galaxies

In our third project, we focus on exploring the quenching of galaxies driven by reionization. As the reionization front heats the gas, the star formation process in low-mass galaxies can be stifled. Intriguingly, state-of-the-art observations reveal varying behaviors for galaxies with similar characteristics and in the same environment. To delve deeper into this phenomenon, we employ sophisticated hydrodynamical simulations to investigate how reionization impacted distinct galaxy populations.

We start by defining the redshift of reionization for each galaxy and we investigate different reionization mechanisms, i.e. inside-out and outside-in reionization. We find that the environment in which these galaxies have formed plays a significant role, with inside-out reionized galaxies located in low-density regions. Galaxies are then categorized into four types based on their star formation histories and their responses to reionization. A small percentage of galaxies (type 0) experienced complete quenching due to reionization, while a larger fraction (type 1) initially suppressed star formation but eventually recovered. The majority of galaxies (type 2) remained unaffected, with a consistent amount (type 3) initiating star formation much later after reionization. We explored the correlation between these galaxy types and various observables, such as stellar mass, metallicity and specific star formation rate. This analysis shed light on the intricate interplay between reionization and subsequent star formation, offering crucial insights into galaxy formation in the early Universe. Our work significantly advances the understanding of the complex processes that shaped galaxies during this transformative period.

Collectively, the findings from both projects contribute to our broader understanding of the epoch of reionization and its implications for the formation and evolution of cosmic structures. The constraints on the HIHMR obtained through 21 cm intensity mapping offer valuable insights into the relationship between HI content and halo mass, shedding light on the processes governing galaxy formation. Meanwhile, the exploration of helium reionization through intensity mapping techniques provides a new avenue to study the correlation

between helium distribution and large-scale structure, advancing our understanding of the complex astrophysical processes in the early Universe.

Outlook and Open Questions

We discuss here some insights on the next steps and open questions in the advancing our understanding of the epoch of reionization (EoR).

Future Observational Campaigns. Dedicated observational campaigns using upcoming instruments such as SKA-1 MID, DSA-2000, PUMA and their precursor telescopes are necessary to constrain the EoR. These instruments will provide unprecedented sensitivity and resolution, and they will observe the core of the EoR for the first time ever. The observations will focus on mapping the signal coming from the hyperfine transitions in HI and HeII atoms with higher precision and over a wider range of redshifts. Moreover, only dwarf galaxies in the local Universe have been observed so far; observing of galaxies in different environment will shed light on the quenching processes affecting their stellar accretion histories.

Refinement of Theoretical Models. The results obtained from the projects provide valuable constraints on the HIHMR, on the distribution of helium and on the processes governing the quenching of dwarf galaxies. These constraints will be crucial for refining theoretical models and simulations of galaxy formation, structure growth, and the reionization process. Theoretical efforts will focus on incorporating the observed constraints into simulations, advancing our understanding of the physical processes that drove the EoR.

Cross-correlations. A promising avenue for future research is the exploration of multi-tracer techniques. Combining observations of different cosmological tracers, such as the 3.46 cm and 21 cm signals, and other emission lines, will provide a more complete and comprehensive picture of the EoR. Multi-tracer studies will enable us to disentangle the various astrophysical and cosmological effects that contribute to the observed signals. By analyzing multiple tracers simultaneously, we can better constrain the properties of ionized regions, the distribution of HI, and the nature of ionizing sources. Moreover, different lines are affected by diverse foregrounds contamination, therefore their combination will improve the goodness of the signal.

These steps are essential for advancing our understanding of the EoR, shedding light on the astrophysical processes, and unraveling the mysteries of the early Universe. By combining observational and theoretical efforts, we can continue to make significant strides in deciphering the complex nature of the EoR and its role in shaping the cosmic structures we observe today.

APPENDIX A

The HI-halo mass relation at $z \sim 1$ paper

The paper Spina et al. MNRAS (2021), 505, 3492 is reproduced below in its original form with permission by Oxford University Press.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **505,** 3492–3504 (2021) https://doi.org/10.1093/mnras/stab1555
Advance Access publication 2021 June 2

The H I–halo mass relation at redshift $z \sim 1$ from the Minkowski functionals of 21-cm intensity maps

Benedetta Spina,[1]★ Cristiano Porciani[1] and Carlo Schimd[2]
[1]*Argelander Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany*
[2]*Aix Marseille Universite, CNRS, CNES, LAM, F-13007 Marseille, France*

Accepted 2021 May 25. Received 2021 May 25; in original form 2021 January 20

ABSTRACT

The mean and the scatter of the H I content of a dark matter halo as a function of the halo mass are useful statistics that can be used to test models of structure and galaxy formation. We investigate the possibility of constraining this H I–halo mass relation (HIHMR) from intensity maps of the redshifted 21-cm line. In particular, we use the geometry and topology of the brightness-temperature isocontours in a single frequency channel as quantified by the Minkowski functionals. First, we generate mock maps from a large *N*-body simulation considering the impact of thermal noise and foreground removal. We then use the Fisher information formalism to forecast constraints on a parametric model for the HIHMR. We consider a 20 000 deg^2 survey (originally proposed for dark energy science) conducted with the Square Kilometre Array Phase 1 (SKA-1) MID observatory operating in single-dish mode. For a channel bandwidth of 2 MHz, we show that an integration time of a few $\times\, 10^4$ s per pointing is sufficient to image the smoothed H I distribution at redshift $z \simeq 1$ and to measure the HIHMR in a nearly optimal way from the Minkowski functionals. Tighter constraints on some of the parameters can be obtained by using also an independent measurement of the mean H I density. Combining the results from different frequency channels provides exquisite constraints on the evolution of the HIHMR, especially in the central frequency range of the data cube.

Key words: galaxies: evolution – intergalactic medium – large-scale structure of Universe – radio lines: general.

1 INTRODUCTION

Cold-gas reservoirs in galaxies provide the raw fuel for star formation. Assessing how they vary across different galaxy populations and environments is of paramount importance to constrain models of galaxy assembly and evolution.

Neutral hydrogen (H I) makes up the bulk of the cold gas. In the post-reionization Universe, H I can be found almost exclusively within self-shielded clouds inside galaxies and galaxy clusters. The total H I content of dark matter (DM) haloes is thus a simple descriptive statistic that can be used to compare theoretical models with observations. It gives the H I mass in a halo by summing up the contributions from central and satellite galaxies (as well as from diffuse gas). A strong theoretical prejudice is that this quantity should mainly depend on the halo mass. In fact, gas-accretion rates (both from the intergalactic medium and via galaxy mergers) are expected to be regulated by the halo masses and so are also several processes that deplete the H I content (like the efficiency of galactic winds or photoionization). Extra dependencies on top of the halo mass will appear as scatter around the mean relation. Measuring the amplitude of this scatter provides an empirical test of the assumption that halo mass is the main driver of the H I content.

The H I–halo mass relation (HIHMR) has been the subject of many studies based on (post-processed) hydrodynamical simulations (Davé et al. 2013; Villaescusa-Navarro et al. 2014; Crain et al. 2017; Villaescusa-Navarro et al. 2018; Ando et al. 2019) and semi-analytical models of galaxy formation (Kim et al. 2017; Zoldan et al. 2017; Baugh et al. 2019; Spinelli et al. 2020). In parallel, observational constraints at redshift $2.2 \lesssim z \lesssim 5$ have been obtained for damped Lyman- α absorbers (Barnes & Haehnelt 2014). On the other hand, in the local Universe, extensive investigations have been carried out to elucidate the HIHMR of H I-rich galaxies from the Arecibo Fast Legacy ALFA Survey (ALFALFA, Giovanelli et al. 2005). For instance, Paul, Choudhury & Paranjape (2018) have predicted the HIHMR by assuming that there exists a scaling relation between the H I content and the optical properties of a galaxy. By cross-matching the ALFALFA H I sources with the optical group catalogue from the Sloan Digital Sky Survey, Ai & Zhu (2018) have estimated the total H I mass for rich galaxy groups containing eight members or more. Combining a similar technique with information about the H I-weighed clustering of the ALFALFA sources, Obuljen et al. (2019) have constrained the parameters of an analytical model for the HIHMR. A direct measurement of the relation has been recently performed by Guo et al. (2020) who stacked the ALFALFA spectra of the members of the SDSS groups.

At intermediate redshifts ($0.1 \lesssim z \lesssim 2.2$), we have very little information on the H I content of DM haloes as the 21-cm line is too faint to observe individual galaxies with reasonable integration times. A way out is to detect the collective emission from galaxies that occupy large 'voxels' in our past-light cone defined by the angular resolution and the bandwidth of the observations (21-cm tomography). The principal challenge for these intensity-mapping experiments is the subtraction of Galactic and extra-galactic foregrounds that are orders

★ E-mail: bspina@astro.uni-bonn.de

of magnitude brighter than the 21-cm signal. In order to filter them out, their properties and the instrumental systematics need to be characterized to extremely precise levels.

There is increasing interest in developing large-scale surveys that map the intensity of the 21-cm emission from neutral hydrogen to probe unprecedented volumes. Such experiments are expected to provide pivotal contributions to cosmology (e.g. Barkana & Loeb 2005; Loeb & Wyithe 2008; Monsalve et al. 2019) and, at higher redshifts, to our understanding of the epoch of reionization (e.g. Yue & Ferrara 2019).

Information about the HIHMR is key for predicting the signal of 21-cm intensity-mapping experiments. In full analogy with the halo model for DM (e.g. Cooray & Sheth 2002) and galaxy clustering (e.g. Scoccimarro et al. 2001), the H I distribution on large scales can be described in terms of its halo-occupation properties. By using the halo model to fit a compilation of data at low and high redshift, Padmanabhan, Choudhury & Refregier (2015) have estimated that the amplitude of the intensity-mapping signal at intermediate redshifts is between 50 and 100 per cent uncertain. Switzer et al. (2013) have obtained the first observational constraint on the amplitude of 21-cm fluctuations at $z \simeq 0.8$. There are some difficulties, however, in fitting this result together with the low-redshift studies and the Damped Lyman- α systems (DLAs) data into a consistent picture (Padmanabhan & Refregier 2017; Padmanabhan, Refregier & Amara 2017).

In this work, we take the opposite step and investigate the possibility of inferring the HIHMR from the 21-cm intensity maps. However, in order not to weaken the constraining power of these experiments for cosmology, we do not use the power spectrum and focus on the geometrical and topological properties of the highly non-Gaussian maps as quantified by the Minkowski functionals (MFs) of their isocontours. Several morphological indicators have been already discussed in the 21-cm literature to characterize the growth of H II bubbles during the epoch of reionization, namely the genus curve (Hong et al. 2010), the MFs (Gleser et al. 2006; Chen et al. 2019), and the Minkowski tensor (Kapahtia et al. 2018). Here, we apply one of these methods to the post-reionization Universe in order to constrain a parametric model for the HIHMR (mean and scatter). We use mock data based on a large *N*-body simulation and the Fisher information matrix to derive the constraints on the model parameters. The rationale behind our method is rather simple: since DM haloes of different masses trace the underlying mass–density distribution differently, the detailed morphology of the brightness-temperature maps should reflect the HIHMR.

We analyse here two-dimensional maps corresponding to individual frequency channels in a tomographic data cube. The main reason for this choice is that we can robustly estimate the covariance matrix of the data without having to run hundreds of high-resolution *N*-body simulations. Our method, however, can be straightforwardly generalized to three dimensions by directly measuring the MFs in the full data cube.

The proposed approach requires rather long radio integration times as it is necessary to image the redshifted 21-cm signal with a sufficient signal-to-noise ratio to measure the MFs reliably. In order to reduce the observing time, we smooth the maps with an isotropic Gaussian filter before measuring the MFs.

We compute how the constraints on the HIHMR vary with the observing time for an array of radio telescopes used in single-dish mode. As a reference case, we consider the Square Kilometre Array Phase 1 (SKA-1) MID observatory. We provide a proof of concept of the method by only considering mock data at $z \simeq 1$ but, of course, there is no particular difficulty to repeat the investigation for other redshifts.

This paper is organized as follows. In Section 2, we introduce the HIHMR and explain how we use it to generate mock intensity maps starting from a large *N*-body simulation. The definitions of the MFs and their application to the 21-cm intensity maps are presented in Section 3. Our implementation of the Fisher matrix formalism is described in Section 4. Our results are presented in Section 5 and discussed in Section 6. Finally, in Section 7, we briefly summarize the philosophy behind our approach and our main findings.

2 MOCK 21-CM INTENSITY MAPS

In the post-reionization Universe, H I is mostly found within the DM haloes that host galaxies and galaxy clusters. However, it is a matter of fact that H I intensity-mapping experiments will survey large fractions of the sky. Therefore, in order to forecast the constraining power of the forthcoming 21-cm experiments for cosmology and astrophysics, we need to simulate the large-scale structure of the Universe over sizeable volumes while resolving low-mass structures and accounting for the physics that regulates hydrodynamics, feedback, and radiative transfer in the interstellar medium. Fulfilling all these requirements together is prohibitive for state-of-the-art software and computing facilities. A possible way forward is to operate at a simpler level of understanding by combining high-resolution *N*-body simulations with a statistical description of the H I content of DM haloes. In this section, we describe how we construct mock H I intensity maps following this approach. Finally, we illustrate how we account for the different sources of noise in the intensity maps.

2.1 *N*-body simulation

We use the MULTIDARK-PLANCK (MDPL) simulation (Klypin et al. 2016) that assumes a ΛCDM background and considers a set of cosmological parameters, which is compatible with the fit in Planck Collaboration XVI (2014). The simulation evolves $N = 3840^3$ DM particles of mass $m_{\rm p} = 1.51 \times 10^9 h^{-1} {\rm M}_\odot$ (where the Hubble constant is written as $H_0 = 100\,h$ km s^{-1} Mpc^{-1}) in a (periodic) cubic box of comoving side $L = 1\,h^{-1}$ Gpc. DM haloes are identified using a standard 'friends-of-friends' algorithm with a linking length of $\ell = 0.2\,L/N^{1/3}$. Only haloes containing at least 20 particles are considered in this work.

2.2 H I–halo mass relation

A common working hypothesis is that mass is the main driver for the H I content of a DM halo. In this case, it makes sense to introduce the HIHMR $M_{\rm H\,I}(M)$ that gives the mean H I mass found in a DM halo of mass M (Pontzen et al. 2008; Bagla, Khandai & Datta 2010; Barnes & Haehnelt 2014; Seehars et al. 2016; Padmanabhan et al. 2017; Padmanabhan & Refregier 2017; Villaescusa-Navarro et al. 2018; Baugh et al. 2019). We adopt the following functional form:

$$M_{\rm H\,I}(M) = M_0 \left(\frac{M}{M_{\rm min}}\right)^{\alpha} \exp\left[-\left(\frac{M_{\rm min}}{M}\right)^{\gamma}\right] . \quad (1)$$

The HIHMR is a scale-invariant power law of slope $\alpha > 0$ with an exponential cutoff at small halo masses (i.e. for $M \lesssim M_{\rm min}$). This suppression reflects the fact that low-mass haloes cannot self-shield from the UV background and gas cooling is inhibited in them (e.g.

Rees 1986; Efstathiou 1992). The parameter $\gamma > 0$ determines how sharp the cut-off is while M_0 fixes the overall normalization of the HIHMR – note that $M_{\rm HI}(M_{\rm min}) = M_0/e$. For our forecasts at $z = 1$, we assume the fiducial values $M_0 = 1.5 \times 10^{10} h^{-1}{\rm M}_\odot$, $M_{\rm min} = 6.0 \times 10^{11} h^{-1}{\rm M}_\odot$, $\alpha = 0.53$, and $\gamma = 0.35$, which provide an accurate description of the Illustris TNG simulation (Villaescusa-Navarro et al. 2018, where the local H I density is obtained by subtracting the molecular fraction estimated with a chemical-equilibrium model from the total hydrogen abundance) and are also in agreement with the current observational estimates for the H I density (see e.g. Crighton et al. 2015; Hu et al. 2019).

In reality, many other factors influence the H I content of a DM halo, for instance, hydrodynamic processes as well as radiative and mechanical feedback from star formation and accretion onto compact objects. The clustering properties of H I-rich galaxies in the present-day Universe suggest that halo spin also plays a role in determining the cold-gas content (Papastergis et al. 2013). We treat these secondary dependencies beyond halo mass in a statistical way as scatter around the HIHMR. Therefore, we assume that, at fixed halo mass, $M_{\rm HI}$ is a random variable that follows a lognormal distribution whose mean is given by equation (1). This is equivalent to say that $\ln[M_{\rm HI}/(1\,h^{-1}\,{\rm M}_\odot)]$ (at fixed M) is a Gaussian random variable with mean $\ln[M_{\rm HI}(M)/(1\,h^{-1}\,{\rm M}_\odot)] - \sigma^2/2$ and standard deviation σ. We use the fiducial value of $\sigma = 1$ as it approximately matches the scatter measured in the semi-analytical models analysed in Spinelli et al. (2020).

2.3 From haloes to brightness temperature

We associate an H I mass to each DM halo by randomly sampling the corresponding lognormal distribution. We then use the Cloud-In-Cell mass-assignment scheme to build a map of the H I density on a regular Cartesian grid with 210^3 cells. In order to account for redshift–space distortions, we deposit the H I at the location,

$$\boldsymbol{s} = \boldsymbol{x} + \frac{1+z}{H(z)}\boldsymbol{v}_\parallel, \tag{2}$$

where $\boldsymbol{x}$ is the actual comoving position of the halo, $\boldsymbol{v}_\parallel$ is its peculiar velocity along the line of sight, and $H(z)$ denotes the Hubble parameter in the background. Note that this neglects both the relative position and the relative motion of the neutral hydrogen with respect to the halo centre of mass.

Eventually, we compute the brightness temperature using

$$\delta T_{\rm b}(\boldsymbol{s}) = 189\, h\, \frac{H_0}{H(z)}\,(1+z)^2\,\frac{\rho_{\rm H\,I}(\boldsymbol{s})}{\rho_{\rm c,0}}\,{\rm mK}\,, \tag{3}$$

where $\rho_{\rm c,\,0}$ is the critical density of the Universe at redshift $z = 0$ and H_0 denotes the Hubble constant.

2.4 Frequency bandwidth and angular resolution

We now perform a mock 21-cm tomography of the simulated data. For each pointing of the radio telescope, spectroscopic information is collected by partitioning the total receiver bandwidth into a number of frequency channels. The signal-to-noise ratio of radio images depends critically on the channel bandwidth (see Section 2.5). By indicating with χ the radial comoving distance in the background cosmological model, a channel bandwidth $\Delta\nu$ at redshift z corresponds to the radial separation,

$$\Delta\chi = \int_{z_-}^{z_+} c\,\frac{{\rm d}z}{H(z)}, \tag{4}$$

where

$$z_\mp = \frac{1}{\dfrac{1}{1+z} \pm \dfrac{\Delta\nu}{2\nu_{\rm rest}}} - 1\,. \tag{5}$$

For a rest-frame frequency of $\nu_{\rm rest} = 1420.406$ MHz, the cell size of the Cartesian grid we use to sample the H I density corresponds[1] to $\Delta\nu \simeq 1$ MHz at $z = 1$. In order to produce synthetic maps with a larger $\Delta\nu$, we average $\delta T_{\rm b}$ at a fixed position on the sky over the corresponding length-scale $\Delta\chi$. This way, we can produce $210\,(1\,{\rm MHz}/\Delta\nu)$ non-overlapping intensity maps of the 21-cm signal at $z = 1$. One example with $\Delta\nu = 2$ MHz is shown in the top left-hand panel of Fig. 1. Wherever some H I is found along the line of sight, a brightness temperature fluctuation of a few mK is recorded.

In order to account for the finite angular resolution of the instrument, we convolve the two-dimensional maps with the telescope beam. For a single dish of diameter D, we assume a Gaussian beam with full width at half-maximum[2],

$$\theta_{\rm FWHM} = 1.2\,\frac{\lambda_{\rm rest}\,(1+z)}{D}, \tag{6}$$

and perform the convolution in Fourier space. After setting $D = 15$ m (for SKA-1 MID) and $\lambda_{\rm rest} = 21.16$ cm, this corresponds to an isotropic Gaussian smoothing in the plane of the sky with standard deviation

$$\Sigma = \frac{\theta_{\rm FWHM}\,(1+z)\,d_{\rm a}}{2\sqrt{2\ln 2}}\,, \tag{7}$$

where $d_{\rm a}$ is the angular-diameter distance to redshift z in the background (in a flat universe, $(1+z)\,d_{\rm a} = \chi$). It follows that at $z = 1$, $\theta_{\rm FWHM} \approx 1.94$ deg and $\Sigma \approx 33.1\,h^{-1}$ Mpc. Once convolved with the beam, the typical 21-cm signal from $z = 1$ assumes values of the order of 0.1 mK (see the top right-hand panel of Fig. 1 for an example).

2.5 Thermal noise

The output of a radio telescope is contaminated by thermal noise. For single-dish observations, the noise is Gaussian to good approximation. Based on the radiometer equation, the rms noise fluctuation associated with an integration time (per pointing) $t_{\rm pix}$ is

$$\sigma_{\rm N} = \frac{T_{\rm sys}}{\sqrt{t_{\rm pix}\,\Delta\nu}}, \tag{8}$$

where $T_{\rm sys}$ is the total system temperature. For an antenna of SKA-1 MID, the system temperature can be obtained by summing up several components (Square Kilometre Array Cosmology Science Working Group et al. 2018):

$$T_{\rm sys}(\nu) = T_{\rm spl} + T_{\rm CMB} + T_{\rm gal}(\nu) + T_{\rm rx}(\nu), \tag{9}$$

where $T_{\rm spl} \approx 3$ K and $T_{\rm CMB} \approx 2.73$ K denote the spill-over and the cosmic microwave background contributions, respectively. The Galaxy contribution can be modelled as

$$T_{\rm gal}(\nu) = 25\left(\frac{408\,{\rm MHz}}{\nu}\right)^{2.75}{\rm K}, \tag{10}$$

[1] We adopt here the distant-observer (plane-parallel) approximation and use one of the axes of the simulation box as the line of sight.

[2] The coefficient of 1.2 represents a correction for the non-uniform illumination of the antenna and is not related to the Rayleigh criterion. In the ideal case, the numerical factor would be 1.02 (e.g. Hall 2005; Chen et al. 2020).

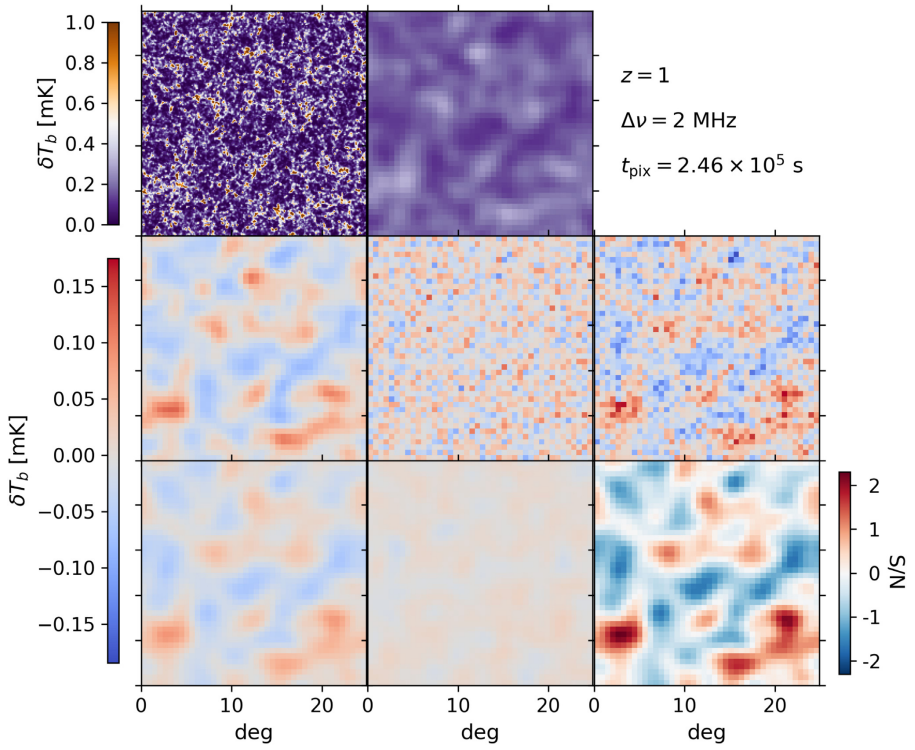

Figure 1. Steps towards building mock 21-cm intensity maps. The original brightness-temperature distribution associated with the DM haloes in the simulation (top left-hand panel; see Section 2.3) is smoothed to simulate the finite angular resolution of radio telescopes (top right-hand panel, see Section 2.4). Long-wavelength Fourier modes in the data cube are lost during the foreground removal (Section 2.6), which shifts the mean signal towards zero (centre left-hand panel; note the different colour scale). Thermal noise is generated at the pixel level (central panel; see Section 2.5) and added to the signal (centre right-hand panel). Finally, some smoothing is applied as a data processing technique to enhance the signal-to-noise ratio (bottom left-hand panel). The smoothed noise map and the signal-to-noise map are shown in the bottom central and bottom right-hand panels, respectively.

and the receiver-noise temperature as

$$T_{\rm rx}(\nu) = 15\,{\rm K} + 30\left(\frac{\nu}{\rm GHz} - 0.75\right)^2\ {\rm K}. \tag{11}$$

For the 21-cm line emitted at $z = 1$, we obtain $T_{\rm sys} = 26.22$ K. The rms noise is thus much larger than $\delta T_{\rm b}$ and integration times of a few days per pointing are needed to see the H I signal emerge above the noise at the pixel level (for a narrow bandwidth of $\Delta\nu = 2$ MHz).

We now imagine to build a pointed map. Assuming that multiple antennas can be used simultaneously, the total time required to map the large-scale structure of the Universe over the solid angle $\Omega_{\rm surv}$ is then

$$t_{\rm obs} = \frac{t_{\rm pix}}{N_{\rm dish}\,N_{\rm beam}}\,\frac{\Omega_{\rm surv}}{\Omega_{\rm pix}}\ , \tag{12}$$

where $N_{\rm dish}$ is the number of available dishes (each with $N_{\rm beam}$ feedhorns, typically 1 or 2) and $\Omega_{\rm pix} \simeq \theta_{\rm pix}^2$ is the solid angle covered by a single pixel. In order to satisfy the Nyquist-Shannon theorem and produce a properly sampled map, $\theta_{\rm pix}$ must be smaller than $\theta_{\rm FWHM}/2$. In practice, it is usually chosen to be between $\theta_{\rm FWHM}/7$ and and $\theta_{\rm FWHM}/3$ (Marr, Snell & Kurtz 2015).

As an example, we estimate the time that would be needed to map the H I distribution in the MDPL box at $z = 1$. The simulation box subtends a solid angle $\Omega_{\rm surv} \simeq 591$ deg^2 and we use 42^2 pixels to cover the whole area. Assuming to observe with all the 197 antennas of the SKA-I MID telescope, we find $t_{\rm obs} \simeq 9\,t_{\rm pix}$ (for $N_{\rm beam} = 1$), i.e. of the order of a month for $\Delta\nu = 2$ MHz.

2.6 Foreground removal

Reaching the necessary sensitivity is only one of the difficulties that we need to face in order to map the large-scale structure of the Universe in 21 cm. Actually, the major challenge is the presence of Galactic and extra-galactic foregrounds that are several orders of magnitude brighter than the H I signal. In particular, extragalactic point sources and synchrotron emission from the Milky Way give major contributions. At first, it might seem impossible to separate these components. However, foregrounds should vary smoothly as a function of frequency (at a given position on the sky) while the fluctuating 21-cm signal is expected to be poorly correlated in frequency space. It has been suggested that this difference could be exploited to separate the 21-cm signal from the foregrounds

(e.g. Shaver et al. 1999; Di Matteo et al. 2002; Oh & Mack 2003; Zaldarriaga, Furlanetto & Hernquist 2004). Although the concept is very promising, complex details need to be taken into account in practical implementations like, for instance, dealing with frequency-dependent beam shapes.

In spite of the difficulties, several foreground-cleaning techniques have been proposed and tested against simulated data (see e.g. Wolz et al. 2015, for a recent review). In addition to separating the foregrounds, these methods usually introduce unintended consequences such as removing large-scale power from the 21-cm signal. Following Alonso et al. (2017) and Cunnington et al. (2019), we assume that the foreground subtraction erases all the Fourier modes of $\delta T_{\rm b}$ with comoving radial wavenumber,

$$k_{\parallel} < k_{\parallel}^{\rm FG} \approx \frac{\pi\, H(z)}{c\,(1+z)\,\xi}\,, \tag{13}$$

where ξ is a dimensionless parameter of order unity. Basically, $k_{\parallel}^{\rm FG}$ denotes the minimum wavenumber for which foregrounds are separable from the signal. It is difficult to assign a precise value to ξ (which is method-dependent), however, assuming $\xi \simeq 0.1$ appears to be a reasonable estimate (Cunnington et al. 2019). For the cut-off scale at $z = 1$, we therefore adopt $k_{\parallel}^{\rm FG} \approx 0.01\, h\,{\rm Mpc}^{-1}$. This value is only slightly larger than the fundamental frequency $k_{\rm f} = 2\pi/L$ associated with our simulation box. We thus only erase the Fourier modes with $k_{\parallel} = 0$ and $k_{\parallel} = k_{\rm f}$. Note that subtracting the $\boldsymbol{k} = 0$ mode has the important consequence of setting the mean value of $\delta T_{\rm b}$ (over the whole simulation box) to zero (from the original value of approximately 0.2 mK in the beam-convolved maps, cf. the top right- and centre left-hand panels in Fig. 1). This, of course, makes the sensitivity requirement for detecting the signal even more demanding.

2.7 Smoothed maps

The centre left-hand panel in Fig. 1 shows the 21-cm signal that one would obtain after the foreground removal but in the absence of thermal noise. In this particular map, $\delta T_{\rm b}$ has a mean value of -0.0065 mK and an rms scatter of 0.035 mK. The adjacent panel on the right-hand side shows a realization of the thermal noise (assuming $t_{\rm pix} = 2.46 \times 10^5$ s) with an rms scatter of 0.038 mK. Comparing at face value, these two maps give the wrong impression that 68 h of integration time might not be enough to image the line emission coming from $z = 1$. This is because the 21-cm signal has been smoothed by the telescope beam while we have generated the thermal noise at the pixel level (as we are assuming to have a different pointing per pixel). The observed map (i.e. the sum of the signal and the noise) is shown in the centre right-hand panel of Fig. 1. In order to reduce the impact of the fine-grained noise on the measurement of the MFs, we further smooth the observed map with a two-dimensional Gaussian filter that has the same width as the telescope beam (we adopt Gaussian smoothing for simplicity although Wiener filtering would be a nearly optimal choice given that the signal is only weakly non-Gaussian, see Section 6). In the bottom panels, we show, from the left- to right-hand side, the smoothed total map, the smoothed noise map (in which the rms scatter drops to 0.007 mK), and the corresponding signal-to-noise ratio (which assumes the peak value of 2.3). It is evident that the smoothed observed map (bottom left-hand panel) presents the same characteristic features of the original one (centre left-hand panel) and can be used to study the morphological properties of the 21-cm signal. This is the subject of the next section.

3 MINKOWSKI FUNCTIONALS

3.1 Basics

A digital image can be thought of as a polyconvex set formed by the finite union of compact and convex subsets in $\mathbb{R}^d$ (its pixels). For such a system, the field of integral geometry provides a family of morphological descriptors known as MFs (or quermassintegrals or intrinsic volumes). Basically, these functionals (Minkowski 1903) measure the size and the connectivity of subsets of $\mathbb{R}^d$ in terms of different quantities. In d dimensions, there exist $d + 1$ MFs that we denote with the symbols $V_0, \ldots, V_{\rm d}$ (as frequently done in the mathematical literature for the intrinsic volumes). Hadwiger's completeness (or characterization) theorem (Hadwiger 1957) states that the linear combinations of the MFs, $\mathcal{F} = \sum_i f_i\, V_i$ (where $f_i \in \mathbb{R}$ and the index i runs from 0 to d) are the only functionals that satisfy the following properties: (i) additivity, $\mathcal{F}(A \cup B) = \mathcal{F}(A) + \mathcal{F}(B) - \mathcal{F}(A \cap B)$; (ii) invariance under rigid motion, $\mathcal{F}(gA) = \mathcal{F}(A)$, where g is an element of the group of translations and rotations in $\mathbb{R}^d$; and (iii) conditional continuity under the Hausdorff measure. In this sense, the MFs completely characterize the morphology of a compact subset of $\mathbb{R}^d$. All the notions above can be generalized to non-Euclidean spaces of constant curvature, like the celestial sphere (see e.g. Schmalzing & Gorski 1998, and references therein).

The MFs have been first introduced in cosmology by Mecke, Buchert & Wagner (1994) and Schmalzing & Buchert (1997, see also Gott et al. (1990)) in order to characterize the large-scale distribution of galaxies. Subsequently, they have been used to measure the morphology of temperature anisotropies in the cosmic microwave background (e.g. Schmalzing & Gorski 1998; Novikov, Feldman & Shandarin 1999). More recently, MFs have been employed to study the morphology and characterize the different stages of cosmic reionization (Gleser et al. 2006; Chen et al. 2019).

3.2 Application to 21-cm maps

Although it would be possible to compute the MFs of $\delta T_{\rm b}$ in the three-dimensional data cube, we perform here a simpler experiment and compute the MFs of two-dimensional maps corresponding to one frequency channel. The main motivation for this choice is that we can more easily estimate the covariance matrix for the measurements without having to generate hundreds of high-resolution simulation boxes (see also Section 4). To this end, we do not build a mock light cone by stitching together different snapshots of the MDPL simulation but only analyse the simulation output at $z = 1$. As already mentioned in Section 2.4, this provides us with 210 $(1\,{\rm MHz}/\Delta\nu)$ non-overlapping intensity maps of the 21-cm signal at $z = 1$. Although these maps are not fully independent (due to correlations along the line of sight), we use them to derive the expected signal in one channel (i.e. the average over the different maps) and the corresponding covariance matrix (see Section 4 for further details).

Let us consider the brightness-temperature map corresponding to a frequency channel and chose a threshold value τ. The excursion set $E_\tau = \{\boldsymbol{\varphi} \in \Omega : \delta T_{\rm b}(\boldsymbol{\varphi}) > \tau\}$ collects all the points where $\delta T_{\rm b}$ exceeds the threshold. We define the MFs of the excursion set as

$$V_0(\tau) = \int_{E_\tau} {\rm d}\Omega\,, \tag{14}$$

$$V_1(\tau) = \int_{\partial E_\tau} {\rm d}\ell\,, \tag{15}$$

$$V_2(\tau) = \frac{1}{2\pi} \int_{\partial E_\tau} \kappa\, {\rm d}\ell\,, \tag{16}$$

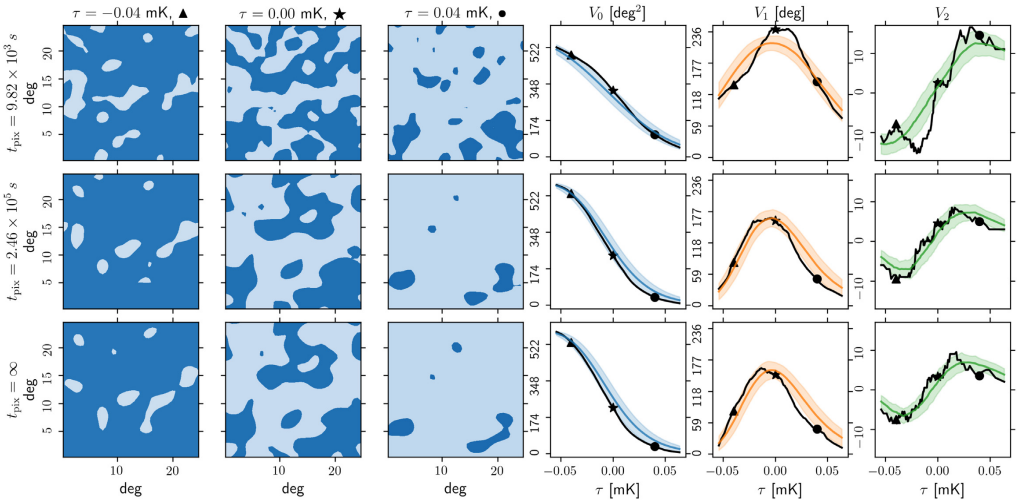

Figure 2. In order to investigate the impact of thermal noise on the MFs, we consider here three intensity maps obtained at $z = 1$ (with $\Delta\nu = 2$ MHz) from the same region of the sky presented in Fig. 1 but obtained varying the integration time per pixel (from the top to bottom, $t_{\rm pix} = 9.82 \times 10^3$ s, 2.46×10^5 s, ∞). The left set of figures shows the regions above (dark blue) and below (light blue) three brightness-temperature thresholds (from the left to right, $\tau = -0.04$ mK, 0 mK, and 0.04 mK). The right set of figures shows the corresponding MFs for the selected threshold values (triangle, star, and circle symbols) and as a function of τ (black solid lines). As a reference, we also show the mean (coloured solid lines) and the rms scatter (shaded regions) of the MFs extracted from the 105 volume slices (with transverse size L^2 and line-of-sight thickness corresponding to $\Delta\nu = 2$ MHz) that fill the MDPL simulation box at $z = 1$.

where $d\Omega$ is the surface element of E_τ, $d\ell$ is the length element of its (smooth) boundary $\partial E_\tau = \{\boldsymbol{\varphi} \in \Omega : \delta T_{\rm b}(\boldsymbol{\varphi}) = \tau\}$, and κ is the local geodesic curvature of ∂E_τ. In simple words, V_0 gives the surface area covered by the excursion set while V_1 is the perimeter of its boundary. Finally, in the flat-sky approximation used here, V_2 coincides with the Euler characteristic χ of E_τ, i.e. the number of connected regions in the excursion set, N_+, minus the number of holes, N_-.

There exist efficient algorithms to compute the MFs for digitized images (e.g. Schmalzing & Buchert 1997; Schmalzing & Gorski 1998). We use the public code MINKFNCTS2D[3] that only exploits information within the image and does not require boundary conditions[4] as described in Mantz, Jacobs & Mecke (2008). Isolines are built with the 'marching square' algorithm that uses contouring cells obtained by combining 2×2 blocks of pixels. In consequence, given an input image with N^2 pixels, the code analyses the central $(N - 1)^2$ pixels and neglects a narrow 'frame' with half-pixel thickness lying along the boundary (corresponding to an area of $2N - 1$ pixels).

Some illustrative examples are provided in the left half of Fig. 2. We use the slice with $\Delta\nu = 2$ MHz presented in Fig. 1, observe it for three different observation times $t_{\rm pix}$ (rows, increasing from the top to bottom), and consider three different temperature thresholds (columns, increasing from the left- to right-hand side). The dark regions in each subpanel represent the excursion set (above threshold). The right half of the figure shows the MFs as a function of the threshold temperature. The triangle, star and circle highlight the values for the images displayed in the panels on the left. The black solid line, instead, shows how the MFs change when many more thresholds are considered. This trend is compared with the mean and the standard error obtained by averaging over the 105 slices contained in our simulation box (coloured lines and shaded regions). The usefulness of the MFs can be more clearly understood by connecting the different plots. When τ is low (left-hand column), the whole observed domain is included in the excursion set with the exception of a few isolated holes. Therefore, V_0 is large while V_1 is small (as only the boundaries of the holes contribute to it) and χ is negative as in 'swiss-cheese' sets dominated by the holes. On the other hand, when τ approaches the median value of the brightness temperature (centre panels), V_0 assumes intermediate values, V_1 reaches a local maximum, and χ is close to zero as in 'sponge-like' structures where the regions above and below threshold are interconnected. Finally, when τ is large (right-hand panels), only a few isolated regions form the excursion set. In this case, both V_0 and V_1 are low while χ is large and positive as in meatball topologies dominated by isolated connected regions. Although this trend holds true for every smooth temperature distribution, the morphological properties of $\delta T_{\rm b}$ are encoded in the precise shape of the $V_i(\tau)$ curves.

It is interesting to investigate how the general pattern is altered by thermal noise. As $t_{\rm pix}$ is reduced (moving from the bottom up), more and more small-scale structures appear in the excursion sets that substantially modify the values of the MFs. In particular, their presence reduces the range of variability of V_0, shifts V_1 towards larger values, and increases the extreme positive and negative values of V_2.

The sensitivity of the MFs to the HIHMR is investigated in Fig. 3, in which we individually vary the parameters of the relation given in equation (1) by ±50 per cent and compare the results to the fiducial case and its standard error. Notice that each parameter modifies differently, and to a varying extent, the shape of the MFs and the location of their extremal points. In particular, increasing M_0 has a similar effect as decreasing $M_{\rm min}$ and vice versa. The figure clearly

[3] https://github.com/moutazhaq/minkfncts2d.

[4] Although our synthetic maps are extracted from simulations with periodic boundary conditions, we analyse them as one would do for observational data.

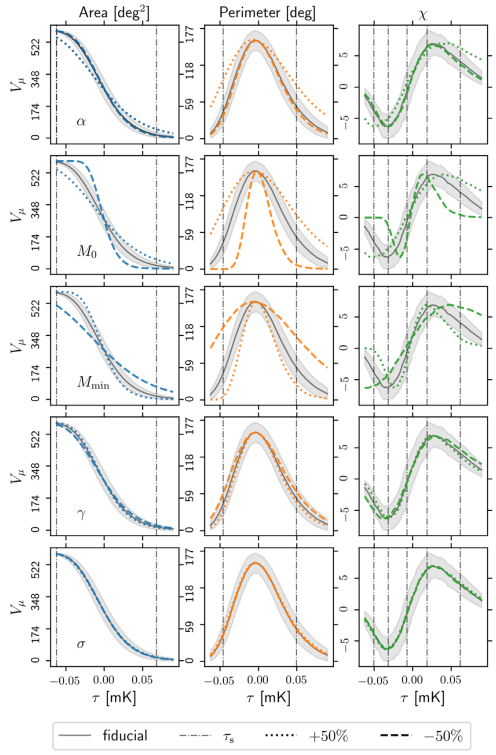

Figure 3. Dependence of the MFs on the parameters of the HIHMR in the absence of thermal noise (i.e. assuming $t_{\rm pix} \to \infty$). We vary each parameter by ±50 per cent with respect to the fiducial value at $z = 1$. Shown are the corresponding MFs averaged over the 105 volume slices that cover the MDPL simulation box (dotted and dashed lines). As a reference, we also show the results for the fiducial case (black solid lines) and their rms scatter (shaded region). The vertical dash–dotted lines indicate the set of thresholds we use in the final analysis presented in Section 4. We make use of slightly different values when we account for thermal noise.

shows that information on the HIHMR is encoded in the MFs. Modifying the parameters of the HIHMR also changes the mean H I density of the Universe and, through equation (3), the amplitude of $\delta T_{\rm b}$. However, this systematic shift is lost due to the procedure of foreground subtraction.

The bottom panel of Fig. 3 shows that the MFs are basically unaffected by relatively large variations of the parameter σ. This reflects the fact that the comoving volume associated with a cylinder of radius Σ and height $\Delta\nu$ is large enough to wash out any stochasticity in the HIHMR. For this reason, we do not consider σ any further in this work and only provide forecasts for the mean H I mass as a function of halo mass.

3.3 Selected observables

In practice, we need to select a finite number of threshold values with which to perform the measurements. Although, in principle, one might want to consider a large number of thresholds, there is an actual limitation due to the fact that we need to estimate the covariance matrix of the measurement errors. Since we can only use $S = 105$ slices for this purpose, we need to limit the size of the data set. The smoothness of the black solid curves in Fig. 2 reveals that the MFs measured from the same map using different values of τ are strongly correlated. This effect is particularly evident for V_0 and V_1. After long experimenting in the attempt to minimize these correlations (and get nearly optimal constraints on the HIHMR using a small number of thresholds), we end up picking two values of τ for V_0, two for V_1 and five for V_2 (as indicated in Fig. 3).

In order to simplify the notation, we combine all the measurements into a single seven-dimensional data vector $\boldsymbol{M}$.

The parameters in equation (1) also determine the mean H I density, $\bar{\rho}_{\rm HI}(z)$, and thus the overall level of the fluctuations, $\delta T_{\rm b}$. However, due to the foreground subtraction, we cannot measure the mean brightness temperature of the 21-cm fluctuations. This fact might weaken the constraints that the MFs can impose on the HIHMR. In fact, the differences between the curves in Fig. 3 would be larger if one could measure the shift in brightness temperature associated with the change in the mean H I density. In order to account for this missing piece of information, we combine the measurements of the MFs with observational constraints on the cosmic abundance of H I, conventionally parameterized as $\Omega_{\rm H\,I}(z) = \bar{\rho}_{\rm H\,I}(z)/\rho_{\rm c,0}$ (for a recent review see e.g. Péroux & Howk 2020). Current estimates at $z \sim 1$ are not fully consistent. In fact, while some studies of DLAs at $z \sim 1$ find $\Omega_{\rm H\,I} \simeq (6 \pm 2) \times 10^{-4}$ (Rao, Turnshek & Nestor 2006; Rao et al. 2017), other estimates are approximately a factor of 3 lower (Neeleman et al. 2016). Similarly, stacking the 21-cm emission from star-forming galaxies gives $\Omega_{\rm H\,I} < 3.7 \times 10^{-4}$ (Kanekar, Sethi & Dwarakanath 2016). Attempts to combine data from low to high redshifts and fit the evolution of the H I abundance with a smooth curve give $\Omega_{\rm H\,I}(z = 1) = (6.1 \pm 0.4) \times 10^{-4}$ (Crighton et al. 2015; Hu et al. 2019). We use this result to set constraints on the parameters of the HIHMR.

4 FISHER MATRIX

4.1 Minkowski functionals

Let us now imagine to fit the measurements of the MFs with a theoretical model $\boldsymbol{M}_{\rm mod}(\boldsymbol{\theta})$ that depends on a set of tunable parameters, $\boldsymbol{\theta}$. Assuming that the measurement errors follow a multivariate Gaussian distribution, we can write the likelihood function $\mathcal{L}$ of the model parameters as

$$-2\ln\mathcal{L}(\boldsymbol{\theta}) \propto (\boldsymbol{M} - \boldsymbol{M}_{\rm mod})^T \cdot \mathsf{C}^{-1} \cdot (\boldsymbol{M} - \boldsymbol{M}_{\rm mod})\,, \tag{17}$$

where C denotes the covariance matrix of the data. In order to forecast the model constraints that can be set by future measurements, we use the Fisher-information formalism. We thus compute the Fisher matrix with elements

$$F_{ij} = \frac{\partial\,\boldsymbol{M}_{\rm mod}^T}{\partial\theta_i} \cdot \mathsf{C}^{-1} \cdot \frac{\partial\,\boldsymbol{M}_{\rm mod}}{\partial\theta_j}\,. \tag{18}$$

This expression assumes that C is computed for the fiducial model and not varied with the model parameters (as routinely done in cosmological studies). We finally use F^{-1} as a proxy for the asymptotic covariance matrix of the estimates for $\boldsymbol{\theta}$. This matrix represents the main result of our study.

It is not easy to build a theoretical model for the MFs of the 21-cm signal. For a zero-mean Gaussian random field y over a two-dimensional space, the MFs can be expressed in terms of the power spectrum (Adler 1981; Tomita 1990; Novikov, Schmalzing &

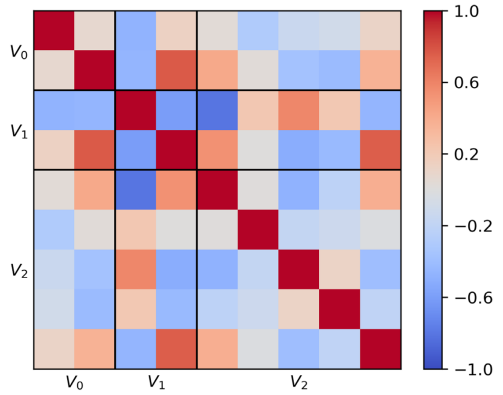

Figure 4. The sample correlation matrix of the MFs for $t_{\rm pix} \to \infty$.

Mukhanov 2000),

$$V_0(\tau) = \frac{1}{2}\mathrm{erfc}\left(\frac{\tau}{\sqrt{2}\sigma_0}\right), \tag{19}$$

$$V_1(\tau) = \frac{1}{2^{7/2}}\frac{\sigma_1}{\sigma_0}\exp\left[-\frac{1}{2}\left(\frac{\tau}{\sigma_0}\right)^2\right], \tag{20}$$

$$V_2(\tau) = \frac{1}{2^{5/2}\pi^{3/2}}\left(\frac{\sigma_1}{\sigma_0}\right)^2\frac{\tau}{\sigma_0}\exp\left[-\frac{1}{2}\left(\frac{\tau}{\sigma_0}\right)^2\right], \tag{21}$$

where τ is the threshold defining the excursion set (see Section 3.2), $\sigma_0^2 = \langle y^2\rangle$ is the variance of the Gaussian random field, and $\sigma_1^2 = \langle y_{;i}\, y^{;i}\rangle$ (where the index i runs from 1 to 2 and we adopted Einstein summation convention) is the variance of its covariant derivative. For a weakly non-Gaussian field, they can be perturbatively expanded in terms of 'skewness parameters' (Matsubara 2003). However, in the general case, one must rely on numerical simulations. A convenient method for 21-cm tomography is to paint H I on top of DM haloes extracted from an N-body simulation as we have done to generate our mock data cubes. In this case, at fixed cosmology, the model parameters coincide with those of the HIHMR. Therefore, we replace $\boldsymbol{M}_{\rm mod}(\boldsymbol{\theta})$ with the expectation value of our mock observations $\bar{\boldsymbol{M}}(\boldsymbol{\theta})$, i.e. averaged over sample variance and thermal noise.

We first consider the limit $t_{\rm pix} \to \infty$, in which thermal noise is irrelevant. In this case, we denote the MFs obtained from the $s^{\rm th}$ slice with the symbol $M_i^{(s)}$ where the index i runs from 1 to 9. We compute the expectation values of the MFs as

$$\bar{M}_i(\boldsymbol{\theta}) = \frac{1}{S}\sum_{s=1}^{S} M_i^{(s)}(\boldsymbol{\theta})\,. \tag{22}$$

These quantities represent the mean measurement that would be obtained by averaging over the sample variance and many realizations of thermal noise. Of course, this is an abstract quantity that is useful to compute the Fisher matrix but cannot be measured in practice as we can only access one data set. Similarly, we approximate C with the sample covariance matrix of the S slices:

$$\hat{C}_{ij} = \frac{1}{S-1}\sum_{s=1}^{S}\left(M_i^{(s)} - \bar{M}_i\right)\left(M_j^{(s)} - \bar{M}_j\right), \tag{23}$$

where all quantities are evaluated using the fiducial set of model parameters. The corresponding correlation matrix is shown in Fig. 4. Particularly strong correlations are noticeable in the V_1 sector and among the first and last thresholds for V_2. Note that, by using equation (23), we are implicitly assuming that the measurements extracted from neighbouring slices are independent while there are certainly some correlations along the line of sight. Anyway, we have checked that considering only one slice every two or three does not change the overall structure of $\hat{\mathsf{C}}$. Equation (23) gives an unbiased (but noisy, due to the fact that S is finite) estimate of the covariance matrix. However, since matrix inversion is a highly non-linear operation, the resulting $\hat{\mathsf{C}}^{-1}$ is a biased estimate of the precision matrix. In order to (statistically) correct for this bias, we compute the Fisher matrix by replacing C^{-1} with $\Gamma\,\hat{\mathsf{C}}^{-1}$, where $\Gamma = \frac{S-13-2}{S-1} \simeq 0.865$ (Kaufman 1967; Hartlap, Simon & Schneider 2007).

We make sure that the numerical evaluation of the partial derivatives in equation (18) gives stable results when the increment of the model parameters is changed. Before differentiating, we preventively smooth the function $V_2(\tau)$ using a Savitzky–Golay filter that irons out small-scale fluctuations appearing due to the fact that the Euler characteristics is integer valued (see the right-hand panels in Fig. 2). This is a necessary step that makes the derivatives and the Fisher matrix meaningful. Our results are stable with respect to reasonable changes in the free parameters of the Savitzky–Golay method.

Considering the impact of thermal noise adds an extra level of complication. In order to extract the expected signal as a function of $t_{\rm pix}$ we proceed as follows. We consider 10 different values for $t_{\rm pix}$ and, for each of them, we generate $N = 30$ different realizations of the thermal noise over the full MDPL box. We then combine them with the 21-cm signal obtained using a particular set of parameters $\boldsymbol{\theta}$ and measure the MFs in each slice (frequency channel) of the foreground-cleaned data cubes. We denote the MFs obtained from the $s^{\rm th}$ slice and the $n^{\rm th}$ noise realization with the symbol $M_i^{(s,n)}$. We compute the expectation values of the MFs as

$$\bar{M}_i(\boldsymbol{\theta}) = \frac{1}{S\times N}\sum_{s=1}^{S}\sum_{n=1}^{N} M_i^{(s,n)}(\boldsymbol{\theta})\,, \tag{24}$$

and substitute them for $\boldsymbol{M}_{\rm mod}(\boldsymbol{\theta})$ in equation (18). However, a problem arises with the estimate of the covariance matrix. We find, in fact, that $\hat{\mathsf{C}}$ depends on the noise realization, in particular when $t_{\rm pix}$ becomes small. This has important consequences as the calculation of the Fisher matrix requires inverting $\Gamma^{-1}\,\hat{\mathsf{C}}$, which prevents us from averaging the estimates of the covariance matrix over the noise realizations. We thus compute a Fisher matrix and produce forecasts for each noise realization independently. We present our results in terms of the distribution of the parameter constraints.

4.2 Scaling with the survey area

So far, we have always considered a region of the sky with a linear transverse size that matches the comoving side of the MDPL simulation box. By considering only half or a quarter of this region, we have checked that the data covariance matrix defined in equation (23) scales proportionally to the survey area. Therefore, in order to consider a larger survey area, it is enough to correct the covariance matrix by the multiplicative factor $\zeta = \sqrt{\Omega_{\rm surv}/\,\Omega_{\rm MDPL}}$.

4.3 Constraints from $\Omega_{\rm H\,I}$

In order to compute the constraints coming from the measurements of $\Omega_{\rm H\,I}$, we use a second Fisher matrix

$$G_{ij} = \frac{1}{\sigma^2_{\Omega_{\rm H\,I}}}\frac{\partial\Omega_{\rm H\,I}}{\partial\theta_i}\frac{\partial\Omega_{\rm H\,I}}{\partial\theta_j}\,, \tag{25}$$

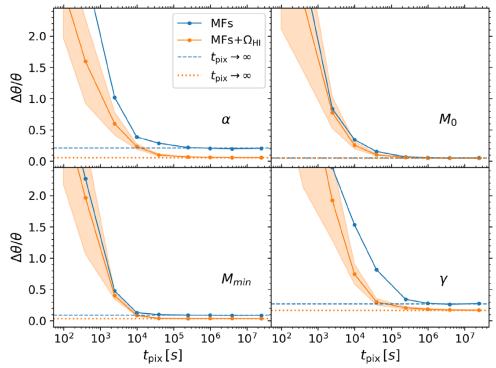

Figure 5. The relative marginalized uncertainty (68.3 per cent credibility interval) for each model parameter as a function of t_{pix}. The orange lines and shaded regions show the mean and rms scatter of the results over the 30 noise realizations. They refer to the combination of the MFs with the $\Omega_{H\,I}$ data. The blue lines show the mean uncertainty when only the MFs are considered (in order to improve readability, we do not show the scatter that is similar to the previous case). Finally, the dotted and dashed lines indicate the results obtained in the absence of thermal noise, i.e. in the limit $t_{pix} \to \infty$.

with $\sigma^2_{\Omega_{H\,I}}$ the standard error of the (observed) H I cosmic abundance. The combined constraints from the MFs and $\Omega_{H\,I}$ are obtained using $(\mathsf{F}+\mathsf{G})^{-1}$.

5 RESULTS

In this section, we present the main results of our study: a Fisher forecast for the HIHMR. As a reference, we consider a (future) wide H I intensity-mapping survey conducted with the SKA-1 MID observatory. The experiment was originally designed in order to measure the power spectrum of the maps and subsequently constrain the dark-energy equation of state at a competitive level with respect to the forthcoming generation of galaxy redshift surveys (Bull et al. 2015).

5.1 The SKA-1 MID dark-energy survey

This intensity-mapping survey is expected to cover an area of the sky of $\Omega_{surv} \simeq 20\,000$ deg ^{2} (corresponding to $\zeta \approx 5.8$) and a frequency range of $350 < \nu < 1050$ MHz (i.e. $0.35 < z < 3.06$). The total observation time should be of approximately 10 000 h. Using the same pixel size as in Section 2.5, this setup corresponds to $t_{pix} = 9.5 \times 10^4$ s, which should be enough to determine the HIHMR in a nearly optimal way.

5.2 Constraints from individual channels

Our Fisher forecasts are presented in Fig. 5. We first discuss the ideal situation corresponding to $t_{pix} \to \infty$, which we use as a reference and represent with horizontal lines in the figure. By combining measurements of the MFs and $\Omega_{H\,I}$, all parameters of the mean HIHMR are constrained to better than 10 per cent (68.3 per cent credibility) as indicated by the orange dotted lines. Of course, these constraints progressively deteriorate by decreasing t_{pix} (orange curves). However, they change very little if $t_{pix} \gtrsim 4 \times 10^4$ s and worsen significantly only if $t_{pix} \ll 10^4$ s.

Considering only the MFs (blue curves) deteriorates the constraints on α and γ by nearly a factor of 2 while it does not affect the other parameters for large values of t_{pix} (blue dashed lines). On the other hand, for lower t_{pix}, the constraints on all parameters become worse.

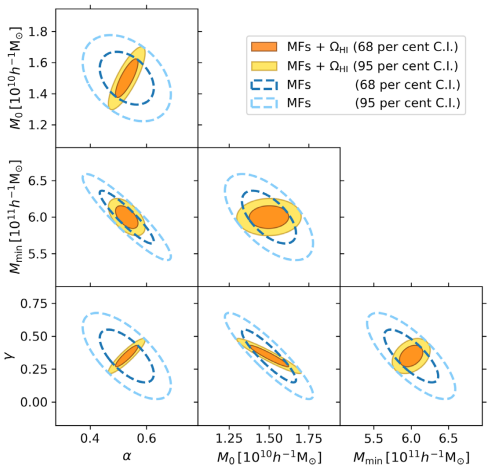

Figure 6. Joint marginalized 68.3 and 95.4 per cent credible regions for all pairs of model parameters that determine the HIHMR. The figure refers to a particular realization of thermal noise obtained assuming $t_{pix} = 3.9 \times 10^4$ s. Dashed ellipses show the results obtained from the Fisher information matrix for the MFs while shaded ellipses refer to the combination with measurements of $\Omega_{H\,I}$.

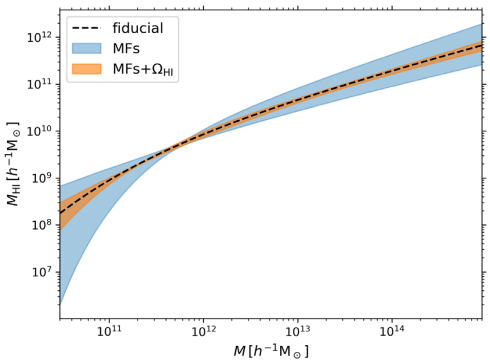

Figure 7. Central 68.3 per cent credibility interval of the mean HIHMR at $z = 1$ obtained by fitting the MFs (light blue) and their combination with estimates of $\Omega_{H\,I}$ (orange) in a single frequency channel at $z = 1$. This result is obtained by marginalizing equation (1) over the posterior distribution of the model parameters shown in Fig. 6, and thus assumes $t_{pix} = 3.9 \times 10^4$ s. The dashed curve indicates the fiducial relation.

An example of the joint (68.3 and 95.4 per cent) credible regions for all parameter pairs is shown in Fig. 6. This plot has been obtained from one particular realization of thermal noise with $t_{pix} = 3.9 \times 10^4$ s. The dashed lines show the results obtained from the MFs while the shaded regions refer to the combination with $\Omega_{H\,I}$.

In Fig. 7, we show that the MFs and $\Omega_{H\,I}$ are very informative regarding the halo occupation properties of H I. The posterior distribution of the mean HIHMR shows little scatter around the

Table 1. Fiducial values and forecast uncertainties for the model parameters that regulate the HIHMR.

	α			M_0 $[10^{10}h^{-1}\,\mathrm{M}_\odot]$			$M_{\rm min}$ $[10^{11}h^{-1}\,\mathrm{M}_\odot]$			γ
Fiducial value θ	0.53			1.50			6.00			0.35
$\Delta\theta$ single channel	0.05			0.15			0.23			0.11
	θ_0	θ_1	θ_2	θ_0	θ_1	θ_2	θ_0	θ_1	θ_2	θ_0
Fiducial value θ_i	0.53	0.20	−0.06	1.50	−1.49	1.38	6.00	−8.21	6.02	0.35
$\Delta\theta_i$ data cube (opt)	0.01	0.01	0.06	0.02	0.05	0.16	0.02	0.07	0.23	0.01
$\Delta\theta_i$ data cube (pes)	0.01	0.03	0.11	0.03	0.09	0.33	0.05	0.15	0.49	0.02

Notes. The top section gives the results from the Fisher-information analysis for one channel at $z = 1$. The bottom sections refer to fits of either second-order polynomials, $\theta(z) = \theta_0 + \theta_1\,(z-1) + \theta_2\,(z-1)^2$ or constants, $\theta(z) = \theta_0$, to the measurements in the data cubes for $0.5 < z < 1.5$. The optimistic case (opt) treats all channels as independent data points while the pessimistic one (pes) considers only one channel every five.

fiducial model. Uncertainties on $M_{\rm H\,I}$ are particularly small for halo masses $M \simeq 10^{12}\,h^{-1}\,\mathrm{M}_\odot$ (which corresponds to a few $M_{\rm min}$ and identifies the haloes containing most H I) and increase as one moves away in both directions.

5.3 Constraints from the full data cube

So far, we have focused on a single-frequency channel corresponding to a narrow redshift bin centred around $z = 1$ and constrained the HIHMR using the MFs of the two-dimensional H I intensity map. We want now to make use of the full data cube that the radio observations will provide. With this aim in mind, we imagine to repeat the measurement of the MFs in each frequency channel and then fit the redshift evolution of the model parameters. In fact, it is reasonable to expect that the parameters of the HIHMR should vary smoothly with redshift. In order to exemplify the power of this technique, we synthesize the data at $z \simeq 1$ in table 1 of Villaescusa-Navarro et al. (2018) in terms of the second-order Taylor expansions:

$$\alpha = 0.53 + 0.18\,(z-1) - 0.11\,(z-1)^2\,, \tag{26}$$

$$M_0 = [1.5 - 1.5\,(z-1) + 1.3\,(z-1)^2] \times 10^{10}\,h^{-1}\,\mathrm{M}_\odot\,, \tag{27}$$

$$M_{\rm min} = [6.0 - 8.2\,(z-1) + 5.8\,(z-1)^2] \times 10^{11}\,h^{-1}\,\mathrm{M}_\odot\,, \tag{28}$$

$$\gamma = 0.35\,, \tag{29}$$

and use them as the fiducial model for the evolution of the HIHMR in the redshift range $0.5 < z < 1.5$. This interval contains 189 channels with $\Delta\nu = 2$ MHz. We can measure the MFs in each of them and then fit the marginalized constraints on the individual model parameters with second-order polynomials of $z - 1$. Results obtained by assuming independent data points are reported in Table 1. Since the thickness of each channel along the line of sight is only a few times larger than the correlation length of the density field and redshift–space distortions can also shift H I by comparable distances, it might be inappropriate to consider the results obtained from consecutive channels as independent. We thus repeat the analysis by considering only one channel every five. Of course, the uncertainty on the model parameters slightly increases in this case (see Table 1). Note that, if we compare the results obtained for $z = 1$ from the multichannel fits to those presented in Section 5.2, we note an improvement of the error bars up to a factor of 10.

6 DISCUSSION

It has been shown that the MFs are powerful data-analysis tools to characterize –and extract information from– various cosmological datasets. However, they are suboptimal in the case of Gaussian random fields for which the whole statistical information is contained in the power spectrum. In fact, many of the practical issues associated with the MFs (dealing with a masked sky, modelling the impact of noise) are solved problems for inferences based on power spectra. A question then naturally arises: are the fluctuations in the 21-cm-brightness-temperature distribution departing from a Gaussian random field so that to justify the use of the MFs in our study? This question is particularly relevant, given the large beam of the SKA-1 MID telescopes that corresponds to a transverse size of approximately 50 comoving Mpc at $z = 1$ (that we use in combination with frequency channels that extend for nearly 13 comoving Mpc along the line of sight).

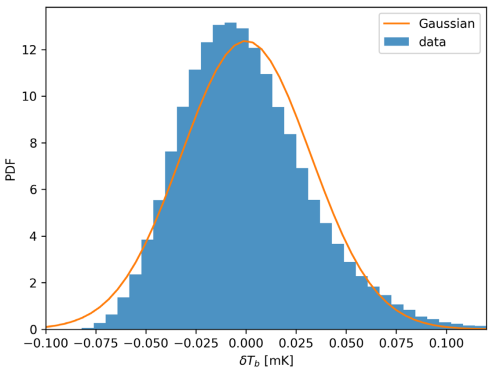

Figure 8. The PDF of the brightness-temperature fluctuations in our mock maps (shaded histogram) is contrasted with a Gaussian distribution with the same mean and variance.

In Fig. 8, we show the probability density function (PDF) of the brightness temperature extracted from our simulation box in the absence of thermal noise (i.e. assuming $t_{\rm pix} \to \infty$). In order to facilitate visual comparison with the Gaussian case, we also plot a normal distribution with the same mean and standard deviation as the simulated data (solid line). The PDF of the brightness temperature is clearly asymmetric (with a skewness of 1.04 ± 0.01) and has heavier tails than a Gaussian distribution (with a kurtosis of 0.708 ± 0.006)

It is also interesting to check how much the MFs in the mock maps depart from the expected values for a Gaussian random field. Analytical formulas have been derived for the MFs of a Gaussian random field on the Euclidean plane or on the two

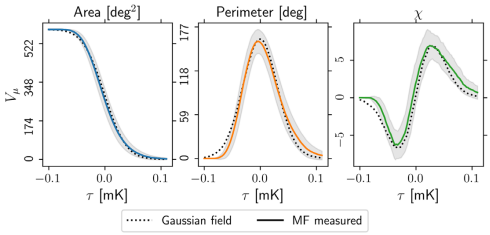

Figure 9. The mean MFs (solid) and their rms scatter (shaded region) extracted from our simulations in the absence of thermal noise are compared with the mean signal for a Gaussian random field with the same power spectrum and domain (dotted lines).

sphere (see Section 4.1). However, important corrections need to be applied[5] when the domain of the random field is a small subset of the above-mentioned spaces. Moreover, the overall normalization of the curves $V_i(\tau)$ depends on the power spectrum of the random field. For these two reasons, we compute the Gaussian MFs by averaging over many realizations obtained by shuffling the phases between the Fourier modes of the simulated data (in the absence of thermal noise, as in Fig. 3). Our results are shown in Fig. 9. While the surface area $V_0(\tau)$ covered by the excursion set is hardly distinguishable from the Gaussian case, larger differences are noticeable for the perimeter of the boundary $V_1(\tau)$ and for the Euler characteristic, in particular, for extreme values of the threshold parameter τ both on the low and high sides.

A related issue is whether the information encoded in the MFs on the large scales probed by the SKA-1 MID is enough to constrain all the parameters that influence the mean HIHMR. In fact, by reasoning in terms of the halo model (see Cooray & Sheth 2002, for a review), one expects the data to be fully in the so-called 'two-halo' regime where all the information about the HIHMR is collapsed into the linear bias parameter of the H I with respect to the mass. Effectively, this would mean that the exercise we are proposing tries to constrain several parameters from the measurement of a single quantity. This line of reasoning assumes that the bias relation between H I and matter fluctuations is linear on scales of 10–50 Mpc. It is well known, however, that this is only true to first approximation. Several non-linear corrections are needed to accurately model the spatial distribution of DM haloes, especially when one considers statistics that are more sensitive to non-Gaussian features than the power spectrum. State-of-the-art models include, at least, extra terms that scale quadratically with the matter density or linearly with the tidal field (see e.g. Desjacques, Jeong & Schmidt 2018, for a review). Our results can thus be interpreted as providing evidence that the same corrections are needed to describe the non-Gaussian features of the H I distribution. Further evidence has been recently provided by Cunnington, Watkinson & Pourtsidou (2021) in a study of the bispectrum of 21-cm intensity maps. Nevertheless, some level of degeneracy between the model parameters is indeed present in our results, as shown in Fig. 6. This is mostly broken when independent measurements of the H I abundance are considered in combination with the MFs.

[5]This is often neglected in astrophysical and cosmological applications but see Pranav et al. (2019) for a nice example.

7 SUMMARY

Determining the overall content and the spatial distribution of H I in the post-reionization Universe is pivotal to understand galaxy formation and evolution. An important step in this direction is the determination of the HIHMR that gives the mean and scatter of the total H I mass contained within a DM halo of mass M. In this paper, we have investigated the possibility of constraining parametric models of the HIHMR from 21-cm intensity maps at redshift $z \simeq 1$. In particular, we have used the geometry and topology of the maps as quantified by the MFs of the brightness-temperature isocontours.

For practical reasons, we have considered a specific parametrization for the HIHMR in which the mean H I mass at fixed halo mass is given by equation (1) and a lognormal scatter of size σ is assumed (see Section 2.2). By assuming a set of fiducial values based on previous numerical studies, we have generated mock data from a large N-body simulation and used the Fisher information matrix to derive the forecast constraints on the model parameters. As a reference case, we have considered the SKA-1 MID dark-energy survey at $z = 1$ (conducted in single-dish mode) and assumed frequency channels of width $\Delta\nu = 2$ MHz.

Our main results can be summarized as follows:

(i) After subtracting the foregrounds, the 21-cm signal is of the order of a few mK, nearly 1000 times lower than the system temperature of the telescopes. In order to beat thermal noise and reach the sensitivity necessary for imaging the H I distribution, long integration times per pixel, $t_{\rm pix}$, are thus required. By using the MFs of the 21-cm intensity map in a single-frequency channel, we find that the parameters of the HIHMR can be measured with a signal-to-noise ratio of 1 if $t_{\rm pix} \simeq 9.8 \times 10^3$ s. Nearly optimal error bars are obtained using $t_{\rm pix} \simeq$ a few $\times\ 10^4$ s. This corresponds to a total observing time of 4–5 d.

(ii) Information on the mean H I density, $\Omega_{\rm H\,I}$, is lost during the foreground removal from the intensity maps. This loss can be restored by combining the MFs with independent measures of $\Omega_{\rm H\,I}$. This addition slightly improves the constraints on some of the parameters that regulate the HIHMR, in particular on the slope α defined in equation (1).

(iii) The mean HIHMR is very tightly constrained for haloes with $M \simeq 10^{12} h^{-1}\ {\rm M}_\odot$, which contain most of the H I at $z = 1$. Uncertainties grow larger as both larger and smaller halo masses are considered.

(iv) Combining measurements of the MFs in different frequency channels provides exquisite constraints on the redshift evolution of the HIHMR (both for the mean and the scatter), especially for redshifts that lie around the centre of the data cube along the frequency axis. In this case, we forecast uncertainties on the model parameters that are an order of magnitude smaller with respect to those extracted from a single channel.

In this paper, we have used the SKA-1 MID dark-energy survey as an example of what can be achieved with forthcoming facilities. Observations at higher angular resolution with SKA precursors like the Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX; Newburgh et al. 2016) in the Southern hemisphere and the Canadian Hydrogen Intensity Mapping Experiment[6] (CHIME;

[6]As well as other proposed facilities such as the Canadian Hydrogen Observatory and Radio-transient Detector (CHORD; Vanderlinde et al. 2019), the Deep Synoptic Array-2000 (DSA-2000; Hallinan et al. 2019), and the Packed Ultra-wideband Mapping Array (PUMA; Slosar et al. 2019).

Bandura et al. 2014) in the Northern hemisphere will probe length-scales at which the H I-to-mass bias relation is more non-linear. Given long enough integration times, these experiments might be able to partially remove the degeneracy among the model parameters of the HIHMR and even constrain them more tightly. We will investigate this possibility in our future work.

ACKNOWLEDGEMENTS

We are grateful to an anonymous reviewer for drawing our attention to the topics discussed in Section 6. We thank Moutaz Haq for making the MINKFNCTS2D code public as well as Sabino Matarrese, Alkistis Pourtsidou, Marta Spinelli, and Matteo Viel for discussions during the early stages of this work. This work is carried out within the Collaborative Research Centre 956 'The Conditions and Impact of Star Formation', sub-project C4, funded by the Deutsche Forschungsgemeinschaft (DFG). CS acknowledges the Programme National Cosmologie et Galaxies (PNCG) of CNRS/INSU with INP and IN2P3 for financial support. The CosmoSim data base used in this paper is a service by the Leibniz-Institute for Astrophysics Potsdam (AIP). The MultiDark data base was developed in cooperation with the Spanish MultiDark Consolider Project CSD2009-00064.

DATA AVAILABILITY STATEMENT

The data underlying this paper will be shared on reasonable request to the corresponding author.

REFERENCES

Adler R. J., 1981, The Geometry of Random Fields. John Wiley & Sons, Chichester
Ai M., Zhu M., 2018, ApJ, 862, 48
Alonso D., Ferreira P. G., Jarvis M. J., Moodley K., 2017, Phys. Rev. D, 96, 043515
Ando R., Nishizawa A. J., Hasegawa K., Shimizu I., Nagamine K., 2019, MNRAS, 484, 5389
Bagla J. S., Khandai N., Datta K. K., 2010, MNRAS, 407, 567
Bandura K. et al., 2014, Ground-based and Airborne Telescopes V, Vol. 9145, SPIE, Bellingham, p. 914522
Barkana R., Loeb A., 2005, ApJ, 626, 1
Barnes L. A., Haehnelt M. G., 2014, MNRAS, 440, 2313
Baugh C. M. et al., 2019, MNRAS, 483, 4922
Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, ApJ, 803, 21
Chen Z., Xu Y., Wang Y., Chen X., 2019, ApJ, 885, 23
Chen T., Battye R. A., Costa A. A., Dickinson C., Harper S. E., 2020, MNRAS, 491, 4254
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Crain R. A. et al., 2017, MNRAS, 464, 4204
Crighton N. H. M. et al., 2015, MNRAS, 452, 217
Cunnington S., Harrison I., Pourtsidou A., Bacon D., 2019, MNRAS, 482, 3341
Cunnington S., Watkinson C., Pourtsidou A., 2021, preprint (arXiv:2102.11153)
Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645
Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1
Di Matteo T., Perna R., Abel T., Rees M. J., 2002, ApJ, 564, 576
Efstathiou G., 1992, MNRAS, 256, 43P
Giovanelli R. et al., 2005, AJ, 130, 2598
Gleser L., Nusser A., Ciardi B., Desjacques V., 2006, MNRAS, 370, 1329
Gott J., Richard I., Park C., Juszkiewicz R., Bies W. E., Bennett D. P., Bouchet F. R., Stebbins A., 1990, ApJ, 352, 1
Guo H., Jones M. G., Haynes M. P., Fu J., 2020, ApJ, 894, 92
Hadwiger H., 1957, Vorlesungen ueber Inhalt, Oberflache und Isoperimetrie. Springer-Verlag, Berlin
Hallinan G. et al., 2019, BAAS, 51, 7
Hall P. J., 2005, The Square Kilometre Array: An Engineering Perspective. Springer, Berlin
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Hong S. E., Ahn K., Park C., Kim J., Iliev I. T., Mellema G., 2014, J. Korean Astron. Soc., 47, 49
Hu W. et al., 2019, MNRAS, 489, 1619
Kanekar N., Sethi S., Dwarakanath K. S., 2016, ApJ, 818, L28
Kapahtia A., Chingangbam P., Appleby S., Park C., 2018, J. Cosmol. Astropart. Phys., 2018, 011
Kaufman G., 1967, Some Buyesian Moment Formulae. Report No. 6710, Center for Operations Research and Econometrics. Catholic University of Louvain, Heverlee, Belgium
Kim H.-S., Wyithe J. S. B., Baugh C. M., Lagos C. d. P., Power C., Park J., 2017, MNRAS, 465, 111
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Loeb A., Wyithe J. S. B., 2008, Phys. Rev. Lett., 100, 161301
Mantz H., Jacobs K., Mecke K., 2008, J. Stat. Mech., 2008, 12015
Marr J., Snell R., Kurtz S., 2015, Fundamentals of Radio Astronomy: Observational Methods. CRC Press, Boca Raton, FL
Matsubara T., 2003, ApJ, 584, 1
Mecke K. R., Buchert T., Wagner H., 1994, A&A, 288, 697
Minkowski H., 1903, Math. Ann., 57, 447
Monsalve R. A., Fialkov A., Bowman J. D., Rogers A. E. E., Mozdzen T. J., Cohen A., Barkana R., Mahesh N., 2019, ApJ, 875, 67
Neeleman M., Prochaska J. X., Ribaudo J., Lehner N., Howk J. C., Rafelski M., Kanekar N., 2016, ApJ, 818, 113
Newburgh L. B. et al., 2016, in Hall H. J., Gilmozzi R., Marshall H. K., eds, Proc. SPIE Conf. Ser. Vol. 9906, Ground-based and Airborne Telescopes VI. SPIE, Bellingham, p. 99065X
Novikov D., Feldman H. A., Shandarin S. F., 1999, Int. J. Mod. Phys. D, 8, 291
Novikov D., Schmalzing J., Mukhanov V. F., 2000, A&A, 364, 17
Obuljen A., Alonso D., Villaescusa-Navarro F., Yoon I., Jones M., 2019, MNRAS, 486, 5124
Oh S. P., Mack K. J., 2003, MNRAS, 346, 871
Padmanabhan H., Refregier A., 2017, MNRAS, 464, 4008
Padmanabhan H., Choudhury T. R., Refregier A., 2015, MNRAS, 447, 3745
Padmanabhan H., Refregier A., Amara A., 2017, MNRAS, 469, 2323
Papastergis E., Giovanelli R., Haynes M. P., Rodríguez-Puebla A., Jones M. G., 2013, ApJ, 776, 43
Paul N., Choudhury T. R., Paranjape A., 2018, MNRAS, 479, 1627
Péroux C., Howk J. C., 2020, ARA&A, 58, 363
Planck Collaboration et al., 2014, A&A, 571, A16
Pontzen A. et al., 2008, MNRAS, 390, 1349
Pranav P. et al., 2019, MNRAS, 485, 4167
Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610
Rao S. M., Turnshek D. A., Sardane G. M., Monier E. M., 2017, MNRAS, 471, 3428
Rees M. J., 1986, MNRAS, 218, 25P
Schmalzing J., Buchert T., 1997, ApJ, 482, L1
Schmalzing J., Gorski K. M., 1998, MNRAS, 297, 355
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Seehars S., Paranjape A., Witzemann A., Refregier A., Amara A., Akeret J., 2016, J. Cosmol. Astropart. Phys., 2016, 001
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Slosar A. et al., 2019, BAAS, 51, 53
Spinelli M., Zoldan A., De Lucia G., Xie L., Viel M., 2020, MNRAS, 493, 5434
Square Kilometre Array Cosmology Science Working Group et al., 2018, Publ. Astron. Soc. Aust, 37, e007

Switzer E. R. et al., 2013, MNRAS, 434, L46
Tomita H., 1990, Formation, Dynamics and Statistics of Patterns, Vol. 1. World Scientific, Singapore
Vanderlinde K. et al., 2019, Canadian Long Range Plan for Astronomy and Astrophysics White Papers, p. 28
Villaescusa-Navarro F. et al., 2018, ApJ, 866, 135
Villaescusa-Navarro F., Viel M., Datta K. K., Choudhury T. R., 2014, J. Cosmol. Astropart. Phys., 2014, 050
Wolz L. et al., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), p. 35
Yue B., Ferrara A., 2019, MNRAS, 490, 1928
Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622
Zoldan A., De Lucia G., Xie L., Fontanot F., Hirschmann M., 2017, MNRAS, 465, 2236

This paper has been typeset from a TEX/LATEX file prepared by the author.

Bibliography

Ahad S. L., Muzzin A., Bahé Y. M., Hoekstra H., 2023, arXiv e-prints, p. arXiv:2307.01147

Alam S., et al., 2017, MNRAS, 470, 2617

Alonso D., Ferreira P. G., Jarvis M. J., Moodley K., 2017, Phys. Rev. D, 96, 043515

Amendola L., Tsujikawa S., 2010, Dark Energy: Theory and Observations

Ando R., Nishizawa A. J., Hasegawa K., Shimizu I., Nagamine K., 2019, MNRAS, 484, 5389

Bagla J. S., Loeb A., 2009, arXiv e-prints, p. arXiv:0905.1698

Bagla J. S., Padmanabhan T., 1997, Pramana, 49, 161

Bagla J. S., Khandai N., Datta K. K., 2010, MNRAS, 407, 567

Balogh M. L., et al., 2016, MNRAS, 456, 4364

Balser D. S., Bania T. M., 2018, AJ, 156, 280

Balser D. S., Goss W. M., Bania T. M., Rood R. T., 2006, ApJ, 640, 360

Bandura K., et al., 2014, Ground-based and Airborne Telescopes V

Barkana R., Loeb A., 2001, Phys. Rep., 349, 125

Barkana R., Loeb A., 2005, ApJ, 626, 1

Barnes L. A., Haehnelt M. G., 2014, MNRAS, 440, 2313

Barnes A. T., Longmore S. N., Battersby C., Bally J., Kruijssen J. M. D., Henshaw J. D., Walker D. L., 2017, MNRAS, 469, 2263

Battye R. A., Browne I. W. A., Dickinson C., Heron G., Maffei B., Pourtsidou A., 2013, MNRAS, 434, 1239

Bauermeister A., Blitz L., Ma C.-P., 2010, ApJ, 717, 323

Baugh C. M., et al., 2019, MNRAS, 483, 4922

Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, MNRAS, 410, 1096

Begeman K. G., 1987, PhD thesis, University of Groningen, Kapteyn Astronomical Institute

Bekki K., 2009, MNRAS, 399, 2221

Belokurov V., et al., 2010, ApJ, 712, L103

Benítez-Llambay A., Navarro J. F., Abadi M. G., Gottlöber S., Yepes G., Hoffman Y., Steinmetz M., 2015, MNRAS, 450, 4207

Bernal J. L., Kovetz E. D., 2022, A&A Rev., 30, 5

Betoule M., et al., 2014, A&A, 568, A22

Bettinelli M., Hidalgo S. L., Cassisi S., Aparicio A., Piotto G., 2018, MNRAS, 476, 71

Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

Bond J. R., Centrella J., Szalay A. S., Wilson J. R., 1984, in Audouze J., Tran Thanh Van J., eds, NATO Advanced Study Institute (ASI) Series C Vol. 117, Formation and Evolution of Galaxies and Large Structures in the Universe. p. 87

Brown T. M., et al., 2014, ApJ, 796, 91

Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, ApJ, 539, 517

Bundy K., et al., 2015, ApJ, 798, 7

CONCERTO Collaboration et al., 2020, A&A, 642, A60

Cantalupo S., Porciani C., 2011, MNRAS, 411, 1678

Chardin J., Puchwein E., Haehnelt M. G., 2017, MNRAS, 465, 3429

Chen Z., Xu Y., Wang Y., Chen X., 2019, ApJ, 885, 23

Chen T., Battye R. A., Costa A. A., Dickinson C., Harper S. E., 2020, MNRAS, 491, 4254

Cid Fernandes R., Mateus A., Sodré L., Stasińska G., Gomes J. M., 2005, MNRAS, 358, 363

Cleary K. A., et al., 2022, ApJ, 933, 182

Cole A. A., et al., 2007, ApJ, 659, L17

Compostella M., Cantalupo S., Porciani C., 2014, MNRAS, 445, 4186

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1

Crain R. A., et al., 2017, MNRAS, 464, 4204

Crighton N. H. M., et al., 2015, MNRAS, 452, 217

Crites A. T., et al., 2014, in Holland W. S., Zmuidzinas J., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9153, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII. p. 91531W, doi:10.1117/12.2057207

Croton D. J., et al., 2006, MNRAS, 365, 11

Cunnington S., Harrison I., Pourtsidou A., Bacon D., 2019, MNRAS, 482, 3341

Cunnington S., Watkinson C., Pourtsidou A., 2021, arXiv e-prints, p. arXiv:2102.11153

Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645

DeBoer D. R., et al., 2017, PASP, 129, 045001

Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1

Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, IEEE Proceedings, 97, 1482

Di Matteo T., Perna R., Abel T., Rees M. J., 2002, ApJ, 564, 576

Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, ApJ, 676, 33

Di Valentino E., et al., 2021, Classical and Quantum Gravity, 38, 153001

Domínguez A., Siana B., Brooks A. M., Christensen C. R., Bruzual G., Stark D. P., Alavi A., 2015, MNRAS, 451, 839

Efstathiou G., 1992, MNRAS, 256, 43P

Eide M. B., Ciardi B., Graziani L., Busch P., Feng Y., Di Matteo T., 2020, MNRAS, 498, 6083

Einstein A., 1915, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, pp 844–847

Fabian A. C., 2012, ARA&A, 50, 455

Faisst A. L., Carollo C. M., Capak P. L., Tacchella S., Renzini A., Ilbert O., McCracken H. J., Scoville N. Z., 2017, ApJ, 839, 71

Freedman W. L., Madore B. F., Scowcroft V., Burns C., Monson A., Persson S. E., Seibert M., Rigby J., 2012, ApJ, 758, 24

Freeman K. C., 1970, ApJ, 160, 811

Friedmann A., 1922, Zeitschrift fur Physik, 10, 377

Friedmann A., 1924, Zeitschrift fur Physik, 21, 326

Furlanetto S. R., Oh S. P., 2008a, ApJ, 681, 1

Furlanetto S. R., Oh S. P., 2008b, ApJ, 681, 1

Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181

Garaldi E., Compostella M., Porciani C., 2019, MNRAS, 483, 5301

Gardner J. P., et al., 2006, Space Sci. Rev., 123, 485

Giallongo E., et al., 2015, A&A, 578, A83

Giovanelli R., et al., 2005, AJ, 130, 2598

Gleser L., Nusser A., Ciardi B., Desjacques V., 2006, MNRAS, 370, 1329

Glikman E., Djorgovski S. G., Stern D., Dey A., Jannuzi B. T., Lee K.-S., 2011, ApJ, 728, L26

Gnedin N. Y., 2000, ApJ, 542, 535

Gong Y., Cooray A., Silva M. B., Zemcov M., Feng C., Santos M. G., Dore O., Chen X., 2017, ApJ, 835, 273

Grebel E. K., Gallagher John S. I., 2004, ApJ, 610, L89

Gunn J. E., Gott J. Richard I., 1972, ApJ, 176, 1

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

Hall P. J., 2005, The Square Kilometre Array: An Engineering Perspective

Hallinan G., et al., 2019, in Bulletin of the American Astronomical Society. p. 255 (arXiv:1907.07648)

Hidalgo S. L., et al., 2011, ApJ, 730, 14

Hotinli S. C., Ferraro S., Holder G. P., Johnson M. C., Kamionkowski M., La Plante P., 2022, arXiv e-prints, p. arXiv:2207.07660

Hu W., Dodelson S., 2002, ARA&A, 40, 171

Hu W., et al., 2019, MNRAS, 489, 1619

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Huterer D., Shafer D. L., 2018, Reports on Progress in Physics, 81, 016901

Iliev I. T., Shapiro P. R., Raga A. C., 2005, MNRAS, 361, 405

Iliev I. T., Mellema G., Pen U. L., Merz H., Shapiro P. R., Alvarez M. A., 2006, MNRAS, 369, 1625

Jakobsen P., Boksenberg A., Deharveng J. M., Greenfield P., Jedrzejewski R., Paresce F., 1994, Nature, 370, 35

Janett G., Steiner O., Belluzzi L., 2019, in Belluzzi L., Casini R., Romoli M., Trujillo Bueno J., eds, Astronomical Society of the Pacific Conference Series Vol. 526, Solar Polariation Workshop 8. p. 133

Kajita T., 2016, Reviews of Modern Physics, 88, 030501

Kannan R., Vogelsberger M., Marinacci F., McKinnon R., Pakmor R., Springel V., 2019, MNRAS, 485, 117

Kannan R., Garaldi E., Smith A., Pakmor R., Springel V., Vogelsberger M., Hernquist L., 2022, MNRAS, 511, 4005

Katz H., et al., 2020, MNRAS, 494, 2200

Khullar S., Ma Q., Busch P., Ciardi B., Eide M. B., Kakiichi K., 2020, MNRAS, 497, 572

Kim H.-S., Wyithe J. S. B., Baugh C. M., Lagos C. d. P., Power C., Park J., 2017, MNRAS, 465, 111

King A., Pounds K., 2015, ARA&A, 53, 115

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

Leauthaud A., et al., 2012, ApJ, 744, 159

Ledinauskas E., Zubovas K., 2018, A&A, 615, A64

Lemaître G., 1931a, MNRAS, 91, 483

Lemaître G., 1931b, Nature, 127, 706

Lewis A., Challinor A., 2011, CAMB: Code for Anisotropies in the Microwave Background, Astrophysics Source Code Library, record ascl:1102.026 (ascl:1102.026)

Li T. Y., Wechsler R. H., Devaraj K., Church S. E., 2016, ApJ, 817, 169

Linder E. V., 2020, Phys. Rev. D, 101, 103019

Loeb A., Wyithe J. S. B., 2008, Phys. Rev. Lett., 100, 161301

Looser T. J., et al., 2023, arXiv e-prints, p. arXiv:2302.14155

Mac Low M.-M., Ferrara A., 1999, ApJ, 513, 142

Madau P., Haardt F., 2015, ApJ, 813, L8

Marr J., Snell R., Kurtz S., 2015, Fundamentals of Radio Astronomy: Observational Methods. Series in Astronomy and Astrophysics, CRC Press, `https://books.google.de/books?id=T54oCwAAQBAJ`

Martin N. F., Ibata R. A., Bellazzini M., Irwin M. J., Lewis G. F., Dehnen W., 2004, MNRAS, 348, 12

McDonald A. B., 2016, Reviews of Modern Physics, 88, 030502

McGaugh S. S., Wolf J., 2010, ApJ, 722, 248

McGee S. L., Bower R. G., Balogh M. L., 2014, MNRAS, 442, L105

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

McKinnon R., Torrey P., Vogelsberger M., 2016, MNRAS, 457, 3775

McKinnon R., Torrey P., Vogelsberger M., Hayward C. C., Marinacci F., 2017, MNRAS, 468, 1505

McQuinn M., Switzer E. R., 2009, Phys. Rev. D, 80, 063010

Mecke K. R., Buchert T., Wagner H., 1994, A&A, 288, 697

Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955

Minkowski H., 1903, Mathematische Annalen, 57, 447

Mishchenko M. I., 2003, in van Tiggelen B. A., Skipetrov S. E., eds, Wave Scattering in Complex Media: From Theory to Applications. Springer Netherlands, Dordrecht, pp 366–414

Mitchell P. D., Lacey C. G., Baugh C. M., Cole S., 2013, MNRAS, 435, 87

Monelli M., et al., 2010, ApJ, 722, 1864

Monsalve R. A., Fialkov A., Bowman J. D., Rogers A. E. E., Mozdzen T. J., Cohen A., Barkana R., Mahesh N., 2019, ApJ, 875, 67

Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nature, 379, 613

Newburgh L. B., et al., 2016a, in Hall H. J., Gilmozzi R., Marshall H. K., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9906, Ground-based and Airborne Telescopes VI. p. 99065X (arXiv:1607.02059), doi:10.1117/12.2234286

Newburgh L. B., et al., 2016b, in Hall H. J., Gilmozzi R., Marshall H. K., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9906, Ground-based and Airborne Telescopes VI. p. 99065X (arXiv:1607.02059), doi:10.1117/12.2234286

Noebauer U. M., Sim S. A., 2019, Living Reviews in Computational Astrophysics, 5

Novikov D., Feldman H. A., Shandarin S. F., 1999, International Journal of Modern Physics D, 8, 291

Ocvirk P., et al., 2016, MNRAS, 463, 1462

Oh S. P., Mack K. J., 2003, MNRAS, 346, 871

Padmanabhan H., 2019, MNRAS, 488, 3014

Padmanabhan H., Refregier A., 2017, MNRAS, 464, 4008

Padmanabhan H., Choudhury T. R., Refregier A., 2015, MNRAS, 447, 3745

Padmanabhan H., Refregier A., Amara A., 2017, MNRAS, 469, 2323

Peeples M. S., Somerville R. S., 2013, MNRAS, 428, 1766

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Planck Collaboration et al., 2016a, A&A, 594, A13

Planck Collaboration et al., 2016b, A&A, 594, A13

Planck Collaboration et al., 2020, A&A, 641, A5

Pontzen A., et al., 2008, MNRAS, 390, 1349

Press W. H., Schechter P., 1974, ApJ, 187, 425

Pritchard J. R., Loeb A., 2012, Reports on Progress in Physics, 75, 086901

Puchwein E., Springel V., 2013, MNRAS, 428, 2966

Pullen A. R., et al., 2023, MNRAS, 521, 6124

Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610

Rao S. M., Turnshek D. A., Sardane G. M., Monier E. M., 2017, MNRAS, 471, 3428

Rees M. J., 1986, MNRAS, 218, 25P

Rhode K. L., et al., 2013, AJ, 145, 149

Richards G. T., et al., 2006, The Astronomical Journal, 131, 2766

Riess A. G., et al., 1998, AJ, 116, 1009

Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, ApJ, 876, 85

Robertson H. P., 1933, Reviews of Modern Physics, 5, 62

Rood R. T., Wilson T. L., Steigman G., 1979, ApJ, 227, L97

Rood R. T., Bania T. M., Wilson T. L., 1984, ApJ, 280, 629

Rubin V. C., Ford W. Kent J., 1970, ApJ, 159, 379

Sánchez-Blázquez P., Gorgas J., Cardiel N., González J. J., 2006, A&A, 457, 809

Schmalzing J., Gorski K. M., 1998, MNRAS, 297, 355

Schumann M., 2019, Journal of Physics G Nuclear Physics, 46, 103003

Seehars S., Paranjape A., Witzemann A., Refregier A., Amara A., Akeret J., 2016, J. Cosmology Astropart. Phys., 2016, 001

Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380

Skillman E. D., et al., 2017, ApJ, 837, 102

Slipher V. M., 1917, Lowell Observatory Bulletin, 3, 59

Slosar A., et al., 2019, in Bulletin of the American Astronomical Society. p. 53 (arXiv:1907.12559)

Smith R. J., et al., 2008, MNRAS, 386, L96

Smoot G. F., et al., 1992, The Astrophysical Journal, 396

Spina B., Porciani C., Schimd C., 2021, MNRAS, 505, 3492

Spinelli M., Zoldan A., De Lucia G., Xie L., Viel M., 2020, MNRAS, 493, 5434

Springel V., 2005, MNRAS, 364, 1105

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

Square Kilometre Array Cosmology Science Working Group et al., 2018, arXiv e-prints, p. arXiv:1811.02743

Stacey G. J., et al., 2018, in Marshall H. K., Spyromilio J., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 10700, Ground-based and Airborne Telescopes VII. p. 107001M (arXiv:1807.04354), doi:10.1117/12.2314031

Steigman G., 2007, Annual Review of Nuclear and Particle Science, 57, 463

Switzer E. R., et al., 2013, MNRAS, 434, L46

Takeuchi Y., Zaroubi S., Sugiyama N., 2014, MNRAS, 444, 2236

Teyssier R., 2002, A&A, 385, 337

Theuns T., Schaye J., Zaroubi S., Kim T.-S., Tzanavaris P., Carswell B., 2002, ApJ, 567, L103

Vasiliev E. O., Sethi S. K., Shchekinov Y. A., 2019, MNRAS, 490, 5057

Villaescusa-Navarro F., Viel M., Datta K. K., Choudhury T. R., 2014, J. Cosmology Astropart. Phys., 2014, 050

Villaescusa-Navarro F., et al., 2018, The Astrophysical Journal, 866, 135

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Walker A. G., 1937, Proceedings of the London Mathematical Society, 42, 90

Walsh D., Carswell R. F., Weymann R. J., 1979, Nature, 279, 381

Weisz D. R., Dolphin A. E., Skillman E. D., Holtzman J., Gilbert K. M., Dalcanton J. J., Williams B. F., 2014, ApJ, 789, 148

Wheeler C., Phillips J. I., Cooper M. C., Boylan-Kolchin M., Bullock J. S., 2014, MNRAS, 442, 1396

Whitaker K. E., et al., 2021, Nature, 597, 485

Williams C. C., et al., 2021, ApJ, 908, 54

Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, MNRAS, 442, 2560

Wolz L., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 35 (arXiv:1501.03823)

Wright A. C., Brooks A. M., Weisz D. R., Christensen C. R., 2019, MNRAS, 482, 1176

Wu X., McQuinn M., Kannan R., D'Aloisio A., Bird S., Marinacci F., Davé R., Hernquist L., 2019, MNRAS, 490, 3177

Yue B., Ferrara A., 2019, MNRAS, 490, 1928

Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622

Zaroubi S., 2013, in Wiklind T., Mobasher B., Bromm V., eds, Astrophysics and Space Science Library Vol. 396, The First Galaxies. p. 45 (arXiv:1206.0267), doi:10.1007/978-3-642-32362-1_2

Zoldan A., De Lucia G., Xie L., Fontanot F., Hirschmann M., 2017, MNRAS, 465, 2236

Zwicky F., 1937, ApJ, 86, 217

van Haarlem M. P., et al., 2013, A&A, 556, A2

List of Figures

List of Tables