Hydrogen in the first billion years: a study of the 21-cm signal from the high-redshift Universe

Balu Sreedhar
ORCiD 0000-0002-5281-5151

January, 2024

A thesis submitted in complete fulfilment
of the requirements for the degree of

Doctor of Philosophy
Physics
The University of Melbourne

"For small creatures such as we the vastness is bearable only through love."

– Carl Sagan, *Contact*

Abstract

The history of our Universe is reflected in the state of its hydrogen (HI) atoms. After recombination (redshift $z \sim 1000$), the intergalactic medium (IGM) is composed primarily of neutral hydrogen (HI). The formation of the first stars and the first galaxies in the early Universe during Cosmic Dawn ($z \sim 30$) triggered the last major phase change of the HI. During the Epoch of Reionisation (EoR), the intense ultraviolet (UV) and X-ray radiation emitted by the first luminous sources carve out ionised hydrogen (HII) bubbles in the IGM. These HII bubbles expand and fill the whole Universe ($z \gtrsim 5$). By altering the thermal and ionisation state of the IGM, the EoR directly impacts the subsequent formation and evolution of galaxies in the Universe.

The 21-cm hyperfine spin-flip of HI is the primary probe of this period, and dedicated observational campaigns are ongoing/under construction to observe this redshifted 21-cm emission. Theoretical models must be on hand to interpret current upper limits as well as future observations. Semi-analytical models (SAMs) are well-suited for this purpose because of their computationally efficient and physically motivated prescriptions of relevant physics. Galaxy formation SAMs typically work by post-processing the dark-matter halo merger trees from dark-matter-only N-body simulations. This thesis updated the MERAXES SAM of coupled galaxy formation and reionisation in this thesis. Specifically, the explicit calculation of the spin temperature of the HI gas was implemented. This involves tracking the thermal state of the IGM, which is influenced primarily by the X-rays.

This updated version of MERAXES was deployed on an N-body simulation of side $210\ h^{-1}$ Mpc. Such large cosmological volumes are necessitated by the long mean free paths, $\sim \mathcal{O}(100\ \text{Mpc})$, of X-rays in the early Universe. At the same time, for a given

number of particles, the mass resolution of an N-body simulation is inversely proportional to its volume. Hence, the simulations will not capture the full source population. To overcome this, the dark matter merger trees are augmented by introducing low-mass haloes into the simulations. The augmented simulation is one of the large-volume simulations in the literature that is simultaneously capable of resolving all atomically cooled haloes down from $z = 20$ and is sufficiently large enough to track the impact of X-rays on the thermal state of the IGM. Taking advantage of the computational efficiency of the Meraxes SAM, the impact of the galaxy X-ray luminosity on the 21-cm statistics, i.e. the 21-cm global signal and 21-cm power spectra (21-cm PS), are explored. Exploiting the large dynamic range of the model, the thesis also shows that the magnitude of the non-Gaussian term in the sample variance of the 21-cm PS is more than twice the magnitude of the Gaussian term at scales relevant to the upcoming Square Kilometre Array (SKA).

The thesis then explores the astrophysical constraints that will be achievable with a future detection of the 21-cm PS. Using the Fisher matrix formalism, the fractional uncertainties in the model parameters enabled by a 21-cm detection spanning $z \in [5, 24]$ from a 1000 h mock observation with the SKA are forecasted. This work focused on the X-ray luminosity, ionising UV photon escape fraction, star formation and supernova feedback of the first galaxies. It is shown that it is possible to recover 5 of the 8 parameters describing these properties with better than 50 per cent precision using just the 21-cm PS. By combining UV luminosity functions with the 21-cm PS, we can improve our forecast, with 5 of the 8 parameters constrained to better than 10 per cent (and all below 50 per cent).

Declaration

This page certifies that:

- This thesis contains only original work towards a Doctor of Philosophy, except where indicated in the preface
- Due acknowledgement has been made in the text to all other material used
- This thesis is fewer than 100 000 words in length, exclusive of tables, figures, bibliographies, and appendices

Balu Sreedhar

Preface

Here and henceforth, "the author" refers to Balu Sreedhar, the author of this thesis. This thesis is an original work by the author reporting research done alone or in collaboration with other authors. This section provides a chapter-by-chapter summary of the author's contributions and the publication status of all material.

Chapter 1 is a comprehensive literature review written by the author for the work in this thesis. It is an original work of the author, with editing from B. Greig and J. S. B. Wyithe, and has not been submitted for publication.

Chapter 2 is an introduction to galaxy formation and evolution simulations with a focus on the early Universe. This chapter also introduces the MERAXES semi-analytical model. It is an original work of the author, with editing from B. Greig and J. S. B. Wyithe, and has not been submitted for publication.

Chapter 3 is based on the published work *S. Balu, B. Greig, Y. Qiu, C. Power, Y. Qin, S. Mutch, J. S. B. Wyithe, Monthly Notices of the Royal Astronomical Society, Volume 520, Issue 3, April 2023, Pages 3368–3382.* It was written primarily by the author, with scientific input and editing from the co-authors. The author was responsible for generating the simulations and data analyses. All figures and tables are the work of the author.

Chapter 4 is based on the published work *S. Balu, B. Greig, J. S. B. Wyithe, Monthly Notices of the Royal Astronomical Society, Volume 525, Issue 2, October 2023, Pages 3032–3047.* It was written primarily by the author, with scientific input and editing

from B. Greig and J. S. B. Wyithe. The author was responsible for generating the simulations and data analyses. All figures and tables are the work of the author.

Chapter 5 summarises the work in the previous chapters. It includes some exploratory future directions for the work in this thesis. It was written by the author, with editing from B. Greig and J. S. B. Wyithe.

The author also contributed to three other publications during their PhD candidature, which is not included for examination in this thesis but is listed here for completeness:

- *Yuxiang Qin, Sreedhar Balu, J. Stuart B. Wyithe, Implications of $z \gtrsim 12$ JWST galaxies for galaxy formation at high redshift, Monthly Notices of the Royal Astronomical Society, Volume 526, Issue 1, November 2023, Pages 1324–1342.* The author assisted with generating the simulations necessary for this work and contributed scientific inputs and edits to the manuscript.

- *Emanuele M. Ventura, Yuxiang Qin, Sreedhar Balu, J. Stuart B. Wyithe, Semi-analytic modelling of Pop. III star formation and metallicity evolution - I. Impact on the UV luminosity functions at $z = 9-16$, Monthly Notices of the Royal Astronomical Society, Volume 529, Issue 1, March 2024, pp.628-646.* The author assisted with generating the simulations necessary for this work and contributed scientific inputs and edits to the manuscript.

- *J. Cook, S. Balu, B. Greig, C. M. Trott, J. L. B. Line, Y. Qin, and J. S. B. Wyithe, The Impact of ionization Morphology and X-ray Heating on the Cosmological 21cm Skew Spectrum, Monthly Notices of the Royal Astronomical Society, Volume 529, Issue 3, April 2024, pp.2734-2752.* The author generated the necessary simulations for this work and contributed scientific inputs and edits to the manuscript.

The author acknowledges funding support through the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D) through project #CE170100013. The works in this thesis utilized the OzSTAR national facility at the Swinburne University of Technology. The OzSTAR program partially receives funding from the Astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation provided by the Australian Government. This thesis was also undertaken with the assistance of resources from the National Computational Infrastructure (NCI Australia), an NCRIS-enabled capability supported by the Australian Government. The author is supported by a Melbourne Research Scholarship, which includes a fee offset.

Acknowledgments

A Big Thanks...

To Stuart Wyithe & Bradley Greig for being the best supervisors one could ask for.

While writing an email beginning with Dear *Prof. Stuart Wyithe* from India on 20th June 2019, I never imagined it would be the first of many![*] I am still trying to understand why you took me on as a PhD student, especially since I messed up the calculation you wanted me to try out as part of the interview. Thanks for being a very patient supervisor, particularly when I slacked off. When starting my PhD, I was told, "Stu will always know more [about your thesis]." I am still impressed by your figuring out the bugs in my code, simply by looking at my plots.

Brad is officially listed as my co-supervisor and has a supervisory load of only 20%. But he has definitely given his 100% for my PhD. Thanks for being supportive when I was stuck, shooting down my bad science ideas and helping me interpret my results. A huge thanks again for your patient feedback on my drafts (papers, CV, job applications, talk slides, this thesis, and so on).

To Chris Power, for being an awesome collaborator and for all our discussions.

Thanks for finding time in your busy schedule to help me debug the codes, often ending up doing it yourself. Thanks for teaching me the nuances of setting up and running cosmological simulations, hosting me several times at the University of Western Australia and visiting me in Melbourne to help out with the projects. It's always a pleasure to chat with you, from the optimal softening length for force calculations in zoom simulations to India's political and social history.

To Yuxiang Qin for being my best friend and collaborator during the PhD. Thanks for all those impromptu Zoom calls at odd times and for helping me with my projects. And for trying to cheer me up when the going has been tough.

[*] Though it has since then become *Hi Stu* and occasionally the emails are like one-word texts.

To Simon Mutch for helping me set up and run MERAXES*.*
To Michele Trenti for helping me with my European travel.
To Rachel Webster for all those chats.
To Kim Dorrel & Kathy Palmer for keeping the astro-floor running and entertaining my last-minute emails for travel plans and reimbursement requests.
To Jeffrey McCullum for being a very helpful PhD committee chair.
To all the people on the astro-floor... too numerous to list: Justin, Giovanni, Kevin, Behzad, Eduardo, Emmanuele Maria, Hilmi, Aadarsh, Nicolo, Christian, Stephi, Michael, Aman,...
To Joe Jacob & D.C.V. Mallik for guiding me towards the PhD.
To S.V.M. Satyanarayana & T.N. Vasudevan for being two of the best physics teachers. Generations of students will miss you.
To Charlotte Stephens for being my good friend outside academia. Thanks for trying to cheer me up whenever I have been going through a rough patch; for all the coffee walks and Vic market runs.
To Reshma Anna Thomas for being my best friend. For all the practice talks and lending an ear whenever I felt down. And for pushing me to seek help during the lowest point of my life this past year.
To Profs Andrei Mesinger, Charlotte Mason, Rachel Somerville, and Adam Lidz for hosting me at their institutes.
To my sisters, Athira & Arathi, for lending support from afar. And also to my nieces Aparna (Ponnu) & Arpana (Kunjatta)!
To Amma & Achachan for everything. Thanks for encouraging me and letting me pursue what I am passionate about. Now that I am older, I am starting to appreciate more all the hardships you had to go through while I was growing up. This thesis is dedicated to you both.

No thanks to the heartbreaks, Covid-19 and lockdowns.

If someone asks me about my experience during the PhD, I will be hard-pressed to find anything more apt than the opening lines of *A Tale of Two Cities* by Charles Dickens*:

"It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way."

Parkville, Melbourne
5th January 2024

*I like to focus on the *wisdom*, *Light*, *hope*, and *Heaven* parts!

Contents

List of figures

List of Tables

"Begin at the beginning," the King said, gravely, "and go on till you come to an end; then stop."

Lewis Carroll, *Alice in Wonderland*

"The story so far: In the beginning, the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move."

Douglas Adams, *The Restaurant at the End of the Universe*

CHAPTER 1

Introduction

1.1. IN THE BEGINNING...

Arguably, one of the most fundamental astronomical observations is that the *night sky is dark*, and this is termed Olber's paradox. This seemingly innocuous observation does have profound cosmological implications: Our Universe cannot be infinite and static, with a constant number density of stars. If this were so, every line of sight on the night sky would end at the surface of a star – the night sky would, therefore, be as bright as during the day. We now know that the resolution of this paradox lies in the finite age as well as the expansion of the Universe.

The quest to understand our place in the cosmos and the evolution of the Universe is as old as humankind. In this chapter, I first give a brief historical overview (section 1.2) of our evolving and current picture of the Universe. This section closely follows Page et al. (2009), Peebles (2020, 2022), and Longair (2020, 2023). The interested reader is urged to consult these for an interesting and comprehensive overview. In section 1.3, I introduce the relevant concepts of the ΛCDM concordance model of cosmology, and then, in section 1.4, I summarise how the large-scale structures that we observe in the Universe formed from the uniform and smooth initial condition that emerge from the Big Bang. Section 1.5 introduces the Epoch of Reionisation, a period in cosmic history

that is the main focus of this thesis, and lastly, in section 1.6, I expound upon how the 21-cm line of the neutral hydrogen (HI) atoms from the EoR is a rich and promising probe of the early Universe.

1.2. Historical overview

In the final decades of the sixteenth century, over a period of 20 years, Tycho Brahe accurately measured the positions and motions across the sky of the Sun, Moon, planets, and hundreds of stars. He passed on this torch of accurate and precise astronomical observations to his assistant Johannes Kepler (he of the three planetary laws' fame). Kepler managed to interpret the observational data within the *heliocentric framework* put forward by Copernicus.

In 1664, a young Isaac Newton, from Kepler's period-radius law, derived his inverse-square law of gravity, which gives the gravitational force of attraction $\boldsymbol{F}_{12}$ between two bodies of masses M_1 and M_2 situated at $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ as

$$\boldsymbol{F}_{12} = -\frac{GM_1M_2}{r_{12}^2}\hat{\boldsymbol{r}}_{12} \tag{1.1}$$

where G is Newton's gravitational constant, and $r_{12} = |\boldsymbol{r}_1 - \boldsymbol{r}_2|$ is the separation between the masses and $\hat{\boldsymbol{r}}_{12}$ is the unit vector along $\boldsymbol{r}_{12}$*. Newton was probably the first to realise that a finite Universe is not stable. He concluded that the Universe must be infinite since otherwise, it would be unstable under gravity, and all the matter content would collapse to the centre. At the same time, an infinite Universe that is filled with stars is also gravitationally unstable.

To digress a bit, this instability was investigated further in the first half of the twentieth century. James Jeans, in 1902, derived the condition for the onset and growth of instability in a static medium. Later, in the 1930s - 1940s, Georges Lemaître, Howard Robertson and Evgenii Lifshitz derived the same for the case of an expanding Universe.

*Throughout this thesis, vectors are shown in **bold** (e.g. $\boldsymbol{x}$, $\boldsymbol{r}$), and dots above a quantity represent its time-derivative (such as $\dot{\boldsymbol{x}}$, $\dot{r}$).

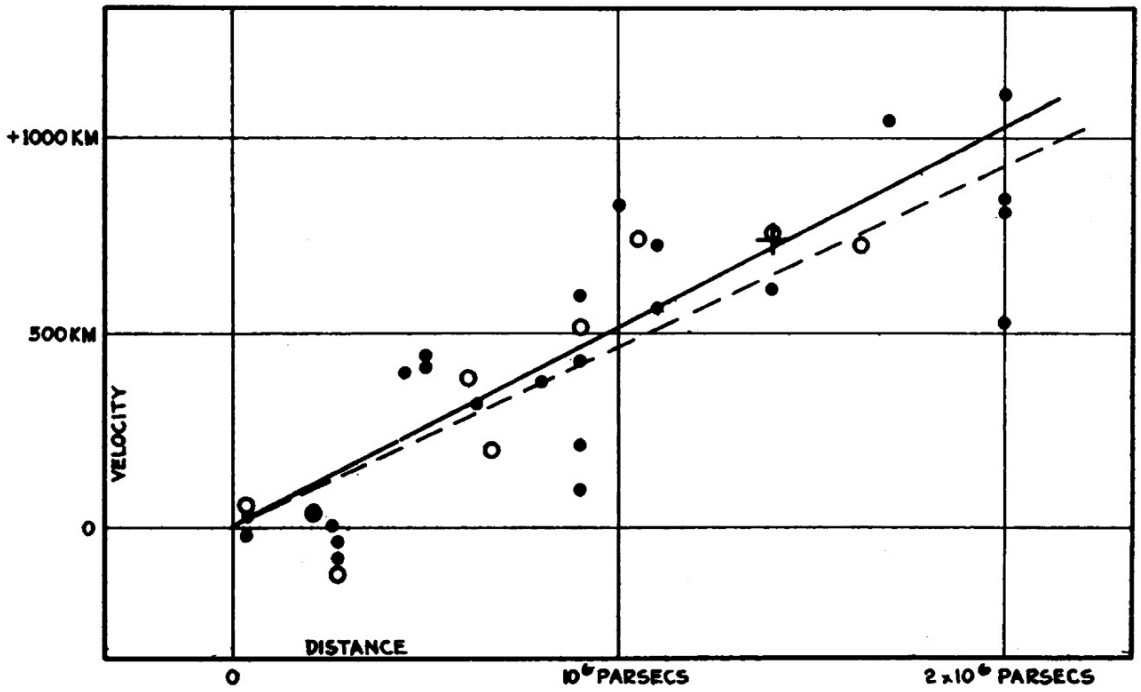

Figure 1.1: One of the earliest recessional velocity versus distance relations among extragalactic *nebulae*. The slope of the linear fit gives the Hubble constant. *Image credit:* Hubble (1929).

1.2.1. Island Universes

The end of the eighteenth century saw the first careful analysis, by the Herschel siblings, William and Caroline, of the observed stellar distributions. The ensuing developments in optical telescopes and spectrographs, coupled with the applications of statistical methods to astronomical data, brought about much more scientific rigour to the question of the shape and extent of the Galaxy. The so-called 'Great Debate' in astronomy centred around *the size of our own Galaxy* and *the nature of the observed spiral nebulae – are they part of the Galaxy*, or *are they separate 'island universes'?*

Progress came next from Edwin Hubble's observations (Hubble, 1925) of the Cepheid variables in the M31 (Andromeda galaxy). Using the period-luminosity relation* of the Cepheid variables, Hubble conclusively proved that M31 is extragalactic in nature. The next milestone in extragalactic astronomy and cosmology came with the discovery by Hubble (1929) that these extragalactic systems are all receding away from us†, with re-

*Discovered by Henrietta Leavitt in 1912 (Leavitt & Pickering, 1912).

†Interestingly, Hubble was not the first one to note the Doppler shifts of the *nebulae* and to attribute recessional velocities to them. See here for the contributions of Vesto Slipher; Slipher (1913)

cessional velocities being directly proportional to their distance from us

$$v = H_0 r \tag{1.2}$$

where v is the speed with which a galaxy at a distance r recedes from the observer. This is termed *Hubble's law*[*] and H_0 is known as the Hubble's constant. Figure 1.1, taken from Hubble (1929), shows one of the earliest velocity versus distance relations for the galaxies. The straight-line fit to the data points demonstrates their linear relation, and the slope of the line gives H_0. This observational fact established without doubt the expansion of the Universe.

1.2.2. *The expanding Universe*

More than a decade before the discovery of the expanding Universe, Einstein published his General Theory of Relativity (GR; Einstein 1915b; *translation* Einstein 1915a), wherein spacetime geometry and gravity were intimately interlinked. Einstein then set about applying his relativistic theory of gravity to models of the Universe as a whole in a self-consistent manner (Einstein, 1917; O'Raifeartaigh et al., 2017). For stable static solutions for the Universe, he introduced the cosmological constant Λ.

Within a few years, Aleksander Friedman (Friedman 1922; *translation* Friedman 1999) showed that the field equations of GR allow both static and expanding solutions. These solutions were rediscovered by Georges Lemaître in 1927 (Lemaître 1927; *translation* Lemaître 1931). It was shown that the field equations without the cosmological constant admit solutions in which the Universe is expanding uniformly. The cosmological constant thus fell out of mainstream cosmology until it was revived in the last decade of the twenty-first century.

1.2.3. *Big Bang*

It was straightforward to come to the conclusion that if the present Universe is expanding, then in the past, it would have been arbitrarily small. Alpher, Bethe & Gamow (1948) presented the *primordial nucleosynthesis* and was able to explain the cosmic proportion

[*]Taking cognisance of the contributions of Georges Lemaître, the International Astronomical Union recently voted the recommendation to rename this as Hubble-Lemaître law.

of the observed elements of hydrogen, helium, and other elements. The thermal history of the Universe was also worked out during this period, and a relic radiation of temperature roughly 5 K was predicted. This so-called *Big Bang* model of the Universe was firmly established by the detection of this relic radiation, the *Cosmic Background Radiation*, by Penzias & Wilson (1965). The radiation has, by now, because of the expansion of the Universe, redshifted to the microwave regime of the electromagnetic spectrum and is thus termed the Cosmic Microwave Background (CMB).

The CMB is the afterglow of the Big Bang and is composed of an (almost) uniform distribution of photons that fill the Universe. This is one of the best pieces of evidence for the isotropy of the Universe and also serves as a snapshot of the Universe from almost 400,000 after the Big Bang. The matter and radiation in the Universe was in thermal equilibrium after the Big Bang, with the matter being in a completely ionised plasma state. CMB was emitted during the epoch of recombination when the matter had cooled down to start forming stable neutral atoms. The CMB appears to come from a spherical surface around the observer of radius which is the distance that CMB each photon has travelled since its last scattering at the epoch of recombination. This surface is termed the last scattering surface. The most accurate measure of the CMB temperature (Fixsen, 2009), at the time of writing, gives

$$T_{\rm CMB} = 2.72548 \pm 0.00057 \text{ K} \tag{1.3}$$

However, the CMB is not completely uniform. The quantum fluctuations in the early Universe leave their imprint on the density content of the Universe, and this is reflected in the CMB temperature anisotropies of the order of

$$\frac{\delta T}{T_{\rm CMB}} = 10^{-5}, \tag{1.4}$$

where δT is the temperature fluctuation. Figure 1.2 shows this fluctuation on the sky and is from the *Planck* mission (Planck Collaboration et al., 2016). The small value of the fluctuations implies that the young Universe was extremely smooth. One can ask about the distribution of the fluctuation in the CMB as a function of the angular scale on the sky. This is quantified by the CMB power spectrum and is shown in Figure 1.3.

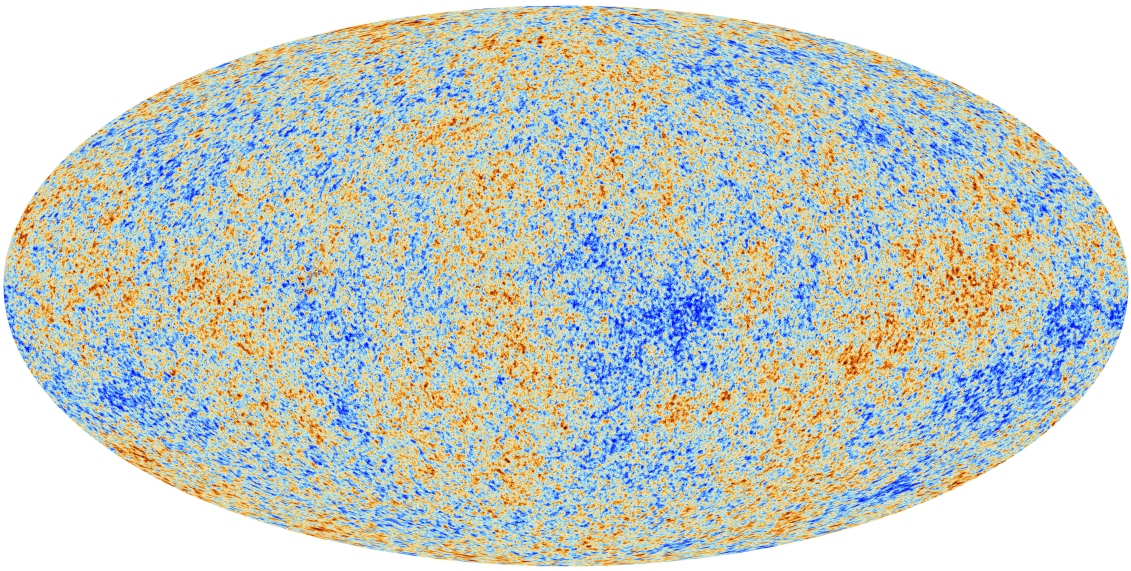

Figure 1.2: Cosmic microwave background anisotropies as measured by the Planck Collaboration et al. (2016). The image shows the tiny temperature fluctuations ($\Delta T/T \sim 10^{-5}$) representing regions of slightly different densities. The density fluctuations result in the formation of all the structures in the Universe in the present day. *Image Credit*: ESA and Planck Collaboration.

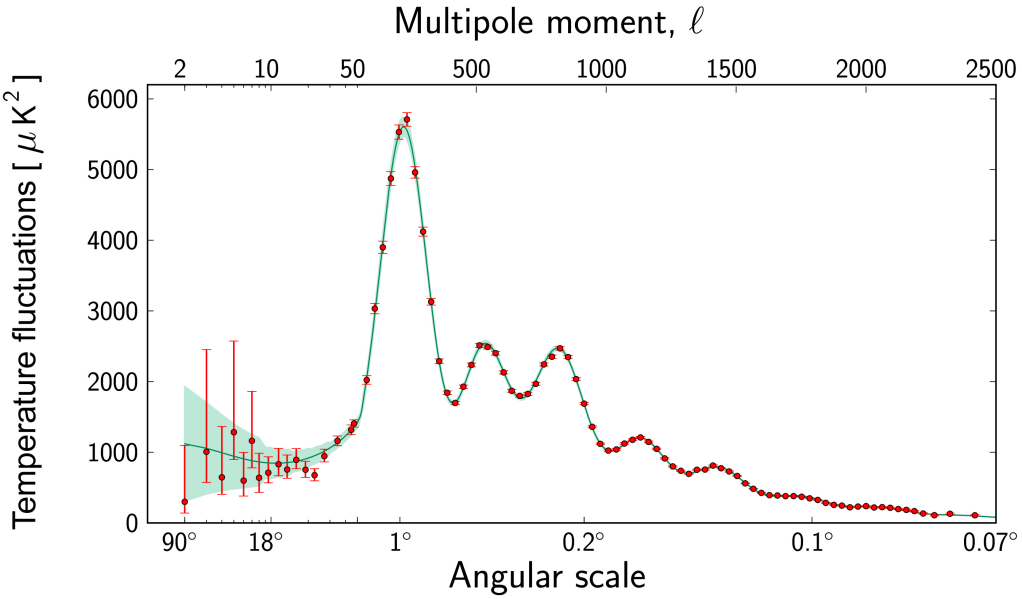

Figure 1.3: The angular power spectrum, quantifying the angular coherence of the CMB temperature fluctuations (Figure 1.2), as a function of the angular scales (l) on the sky. *Image Credit*: ESA and Planck Collaboration.

1.2.4. *Galaxies*

The next major step was the elucidation of the formation of the structures that we see around in the Universe out of the smooth and uniform Universe that we infer from the CMB. From a theoretical point of view, one wants to add small perturbations in the energy density of the Universe and then study their evolution under gravity. The case for a static medium was solved by Jeans (Jeans, 1997), who showed that a perturbation would collapse if it exceeds the Jeans' length of

$$\lambda_{\mathrm{J}} = c_{\mathrm{s}} \sqrt{\frac{\pi}{G\rho_0}}, \tag{1.5}$$

where c_{s} is the sound speed in a medium of density ρ_0. On scales larger than λ_{J}, the perturbations grow exponentially as the gravitational force is able to overcome the pressure gradient.

This analysis was extended to the scenario of a uniformly expanding medium by Lemaître (1933); Tolman (1934); Lifshitz (1946). Due to the expanding nature of the background Universe, the growth of the perturbation is algebraic instead of exponential in this case. This implies that fluctuations in the density, from which the structures in the Universe formed, could not have originated out of infinitesimal random perturbations in the early Universe.

Cosmologists at the beginning of the second half of the twentieth century quickly showed that the formation of the galaxies and galaxy clusters necessitate the corresponding perturbations in the distant past to be of the order of (Novikov, 1964)

$$\Delta = \frac{\delta\rho}{\rho} = 10^{-4}, \tag{1.6}$$

where $\delta\rho$ is the density fluctuation in a medium of mean density rho.

Harrison (1970) and Zeldovich (1972) proposed a form for the power spectrum of the density perturbations in the early Universe. The *Harrison–Zeldovich spectrum of initial perturbations* is a scale-free power spectrum of the form $P(k) \propto k^n$ with $n = 1$. The density perturbations would have resulted in associated fluctuations of similar order in the temperature. This was subsequently verified from the CMB anisotropies (see Figure 1.2). The level of fluctuations in the CMB temperature was able to put constraints on the

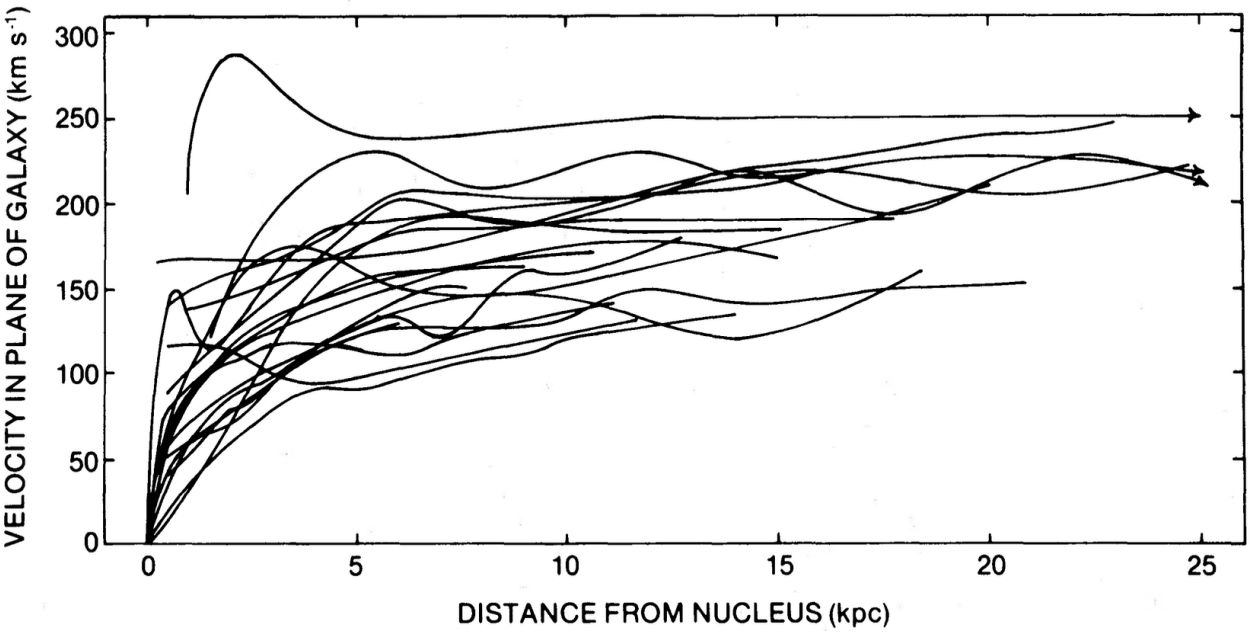

Figure 1.4: Galaxy rotation curves of 21 spiral galaxies. The form of the observed curve can only be explained by proposing the presence of non-luminous *dark* matter. *Image credit:* Rubin et al. (1980)

matter density of the Universe. Cosmological models with low matter-density parameters were in tension with the observed abundance of structures, paving the way for the next revolution in cosmology.

1.2.5. Cold dark matter

Dark matter constitutes about one-fifth of the matter density of the Universe. The 'dark' refers to the fact that it only interacts gravitationally and thus cannot be detected in electromagnetic radiation. Zwicky (1933) (*translation* Andernach & Zwicky 2017) hypothesised their existence from observations of the Coma cluster. The existence of dark matter was put on firmer ground during the 1970s - 1980s by the works of Vera Rubin from observations of the galactic rotation curves of the galaxies. Figure 1.4 from Rubin et al. (1980) shows the rotation curves of 21 spiral galaxies. The deviation of observed rotation curves from Keplerian predictions, especially as one moves away from the centre of the galaxies, needs the postulating of non-luminous matter to reconcile with theoretical models.

The true particle nature of dark matter is an open problem in physics, with indications that physics beyond the standard model is required. The current model of cosmology

demands 'cold' dark matter, with the adjective referring to the fact that the particles composing the dark matter are non-relativistic. A Universe dominated by cold dark matter is characterised by the bottom-up or hierarchical formation of large-scale structures. This means that the small-scale structures form first. When the Universe becomes matter-dominated, fluctuations in all scales grow. After the recombination epoch, the baryonic matter followed the resulting dark matter distribution. Press & Schechter (1974) analysed this scenario and gave a prescription of how the halo mass functions evolve with time.

1.3. ΛCDM model

In these sections, we give detailed and quantitative descriptions of the topics covered in section 1.2.

Over the past few decades, we have come to a consensus on the theoretical framework for the observed Universe, often called the Concordance Model of Cosmology. The model has now undergone stringent observational tests, ushering in an era of precision cosmology. In the next subsections, we turn our attention to the basics of this model.

Einstein introduced the first modern physical theory of cosmology rooted within the framework of GR in Einstein (1917). Starting from the *cosmological principle* for his cosmological model, he showed that a finite and static universe is unstable under GR.

1.3.1. Cosmological principle

In order to develop his cosmological model, Einstein elevated the Copernican Principle (the assertion that Earth is not a special place) to the cosmological principle. The cosmological principle states that our position within the Universe is not privileged and that the matter and energy distribution of our Universe is homogeneous and isotropic on cosmological scales.

1. *Homogeneity of space* states that the Universe is the same from one point to the other (i.e. translational symmetry).

2. *Isotropy of space* which implies that the Universe is the same in different directions (i.e. rotational symmetry).

The *isotropy* of the Universe is, by now, a well-tested hypothesis: by the isotropic distribution of radio sources, optically-selected galaxies, X-ray background, and the cosmic microwave background. On the other hand, *homogeneity* is not straightforward to verify. However, a Universe that is isotropic around two distinct points necessarily has to be homogeneous Peacock (1999). It has also been shown that our Universe cannot be significantly inhomogeneous around our location in the Universe.

It has to be pointed out that the cosmological principle is not one that can be applied locally. The principle is applicable only at cosmological distances, $\sim \mathcal{O}(100\ \mathrm{Mpc})$, as is typical of statements in cosmology.

1.3.2. *Friedmann–Lemaître–Robertson–Walker metric*

The metric fixes everything that one needs to know about the underlying space-time. Robertson and Walker (Robertson, 1935; Walker, 1937) derived the spacetime metric for all isotropic, homogeneous, uniformly expanding models of the Universe. Note that the form of this metric is fixed by consideration of the cosmological principle and that on small scales, it should reduce to the Minkowski spacetime of the special theory of relativity and is hence applicable to all cosmological models where the space is homogeneous and isotopic and is expanding.

$$ds^2 = -c^2 dt^2 + a^2(t)\left[\frac{1}{1-kr^2}dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \tag{1.7}$$

where ds^2 is the space-time interval, $a(t)$ is an unknown function, known as the scale factor, that embodies the expansion of the Universe, (r, θ, ϕ) is the comoving spherical coordinates. $k \equiv (-1, 0, +1)$ fixes the global curvature of the Universe.

1.3.3. *Cosmological redshift, z*

Cosmological redshift refers to the shift towards longer wavelengths of the spectral lines due to the expansion of the Universe. The cosmological redshift, z, is given by

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_e} \tag{1.8}$$

where λ_0 is the observed wavelength and λ_e is the rest-frame emitted wavelength of

the same photon. The redshift is thus the fractional change in the wavelength of the photon.

We now turn to the propagation of an electromagnetic wave in an expanding Universe. Light always follows a *null geodesic*, i.e. $ds^2 = 0$, in spacetime during its travel. Equation 1.7 thus gives

$$c\,dt = \pm a(t)\frac{dr}{\sqrt{1 - kr^2/R^2}} \tag{1.9}$$

The scale factor and redshift are related as:

$$a(t) = \frac{1}{1+z} \tag{1.10}$$

This implies that redshift can thus be interpreted as a measure of the scale factor of the Universe when the radiation was emitted by the source.

1.3.4. Coordinates

We differentiate between *comoving* ($\boldsymbol{x}$) and *proper* ($\boldsymbol{r}$) coordinates. Due to the expansion of the Universe, the distance between any two given points will change with cosmic time, and the associated coordinates are the proper coordinates. On the other hand, if one measures the coordinates with a system whose grid lines also expand with that of the Universe, then, in the absence of peculiar motion, the coordinates will remain fixed in time. Comoving coordinates are the ones in which an observer moves along the Hubble expansion.

The two are related by

$$\boldsymbol{r}(t) = a(t)\boldsymbol{x} \tag{1.11}$$

where $a(t)$ is the scale factor of the Universe.

1.3.5. *Hubble-Lemaitre law*

Taking the time-derivative of equation 1.11, we have

$$\begin{aligned} \dot{\boldsymbol{r}} &= \dot{a}(t)\boldsymbol{x} + a(t)\dot{\boldsymbol{x}} \\ &= \dot{a}(t)\boldsymbol{x} \\ &= \frac{\dot{a(t)}}{a(t)}\boldsymbol{r} \qquad\qquad (\because \dot{\boldsymbol{x}} = 0) \\ \implies \dot{\boldsymbol{r}} &= H(t)\boldsymbol{r}, \end{aligned} \tag{1.12}$$

where $H(t)$ is termed the Hubble parameter and H_0 is its present value. We have thus recovered the Hubble-Lemaitre law. Note that this is a direct consequence of any isotropic, homogeneous and expanding space and does not require an appeal to GR or the FLRW metric.

1.3.6. *Friedmann equations*

The field equations of GR applied to the FLRW metric yield the Friedmann equations:

$$\frac{\dot{a}^2 + kc^2}{a^2} = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} \tag{1.13}$$

$$\frac{2\ddot{a}}{a} + \frac{\dot{a}^2 + kc^2}{a^2} = -\frac{8\pi G P}{c^2} + \frac{\Lambda}{3} \tag{1.14}$$

relating the expansion rate of the Universe (scale factor a) to its constituents via the Universe's energy density (ρ) and pressure (P), and Λ is the cosmological constant. There are three unknowns in these equations: scale factor (a), pressure (P), and mass-energy density (ρ). Once the value of k is specified and we assume an equation of state, i.e., $P = f(\rho)$, the Friedmann equations yield $a(t)$, $P(t)$, $\rho(t)$ giving us the temporal evolution of the size, pressure, and density of the Universe.

It is convenient to define the critical density of the Universe, which is the matter-energy density required to get a flat Universe (i.e. $k = 0$).

$$\rho_{\text{crit}}(t) = \frac{3H^2(t)}{8\pi G} \tag{1.15}$$

We can thus define the ratio of the density to the critical density for different components of the contents of the Universe as

$$\Omega_i(t) = \frac{\rho_i(t)}{\rho_{\rm crit}(t)} \tag{1.16}$$

where $i = [m,\ \Lambda,\ r]$ for, respectively, matter (including dark matter), cosmological constant, and radiation.

The Friedmann equation now becomes

$$\frac{H(z)}{H_0} = \left[\Omega_{\rm r}(1+z)^4 + \Omega_{\rm m}(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda\right]^{1/2} \tag{1.17}$$

where the subscript '0' refers to the values of these parameters in the present day (also note that I have switched to redshift z instead of time t).

Defining $\Omega_0 = \Omega_m + \Omega_\Lambda + \Omega_r$ we have

$$\Omega_k \equiv -\frac{k}{H_0^2} = 1 - \Omega_0 \tag{1.18}$$

and

$$\Omega_{\rm b} + \Omega_{\rm DM} + \Omega_{\rm r} + \Omega_\Lambda\Omega_{\rm k} = 1. \tag{1.19}$$

The eight free parameters of the ΛCDM theory are the density parameters in ordinary matter ($\Omega_{\rm b}$), dark matter ($\Omega_{\rm DM}$), the CMB ($\Omega_{\rm r}$), and neutrinos with negligible rest masses (Ω_ν), Hubble's constant ($H_0 \equiv 100h$)*, the primeval Gaussian process amplitude and power law index (n_s), and the optical depth for the scattering of the CMB by intergalactic plasma (τ_e).

1.4. Structure Formation

We now turn to the question of the formation of structures in the Universe. A completely homogeneous and isotropic Universe cannot form structures from symmetry arguments. Luckily, the small quantum-level fluctuations during the early Universe, amplified during inflation*, seeded the inhomogeneities needed to eventually form structures in the

*The Hubble's constant is often expressed in literature as $H_0 = 100h$ km s^{-1} Mpc^{-1}. See Croton (2013) for a detailed exposition on *little h*.

*Inflation, first proposed in Guth (1981), refers to a very short duration period in the history of the Universe during which it underwent exponential expansion.

Universe. The causal mechanism for this is the gravitational instabilities of the density inhomogeneities in the matter distribution. This is the small thermal inhomogeneity that is observed in the CMB sky.

1.4.1. Linear Perturbation theory

The overdensity $\delta(\boldsymbol{x}, t)$ as a function of comoving coordinates $\boldsymbol{x}$ and time t is

$$\delta(\boldsymbol{x}, t) = \frac{\rho(\boldsymbol{x}, t) - \overline{\rho}(t)}{\overline{\rho}(t)} \tag{1.20}$$

where ρ is the position and time dependent matter density and $\overline{\rho}$ is the mean density at cosmic time t. The further evolution of these density perturbations can be analysed using linear perturbation theory. The density perturbations will give rise to what is called *peculiar velocities* $\boldsymbol{v}$, which is the deviation away from the Hubble flow. The total velocity of a density element will, therefore, be $\boldsymbol{v}_{\text{total}} = H(t)\boldsymbol{r} + \boldsymbol{v}$.

The evolution of the matter-fluid is then given by the continuity equation

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot [(1+\delta)\mathbf{u}] = 0, \tag{1.21}$$

and the Euler equation

$$\frac{\partial \mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}(\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{a}\nabla\phi, \tag{1.22}$$

in comoving coordinates. Here, ϕ is the gravitational potential in the Poisson equation

$$\nabla^2\phi = 4\pi G\overline{\rho}a^2\delta. \tag{1.23}$$

In the linear regime (i.e. $\delta << 1$) equations 1.21 and 1.22 can be combined to

$$\frac{\partial^2 \delta}{\partial t^2} + 2H\frac{\partial \delta}{\partial t} = 4\pi G\overline{\rho}\delta. \tag{1.24}$$

This second-order linear partial equation admits two independent solutions, of which only one grows with time. This 'growing mode' solution, $D(z)$, will thus come to dominate the evolution with time of the density perturbations. The density perturbations will thus retain their shapes in comoving coordinates and grow by a factor that is pro-

portional to $D(z)$ given by

$$D(z) \propto \frac{\left(\Omega_\Lambda a^3 + \Omega_k a + \Omega_\mathrm{m}\right)^{1/2}}{a^{3/2}} \int_0^a \frac{a'^{3/2} da'}{\left(\Omega_\Lambda a'^3 + \Omega_k a' + \Omega_\mathrm{m}\right)^{1/2}}, \tag{1.25}$$

and is normalised with respect to its value at $z = 0$.

There is still the issue of the nature of the initial conditions for linear evolution. For this, we move to the Fourier domain by defining

$$\tilde{\delta}(\boldsymbol{k}, t) = \int \mathrm{d}^3\boldsymbol{x}\delta(\boldsymbol{x}, t)\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{1.26}$$

where $\boldsymbol{k}$ is the Fourier mode corresponding to a spatial scale λ (i.e. $k = 2\pi/\lambda$).

For a Gaussian random field, a complete statistic description is contained in the power spectrum $P(k, t)$ defined as

$$\left\langle \tilde{\delta}\left(\boldsymbol{k}, t_i\right) \tilde{\delta}\left(\boldsymbol{k}', t_i\right) \right\rangle = (2\pi)^3 P\left(k, t_i\right) \delta\left(\boldsymbol{k} - \boldsymbol{k}'\right) \tag{1.27}$$

The three-dimensional Dirac-delta function $\delta(\boldsymbol{k} - \boldsymbol{k'})$ attests to the statistical independence of the different $\boldsymbol{k}$-modes*.

Inflation produces density fluctuations that follow a power spectrum of the form $P(k) \propto k^{n_s}$ where $n_s \sim 1$ is called the primordial power spectral index. This 'scale-free' primordial power spectrum is also known as the Harrison-Zeldovich power spectrum.

It is clear from the definition of $P(k)$ that it evolves with redshift as

$$P(k, z) \propto k^{n_s} D(z). \tag{1.28}$$

Strictly speaking, this is valid at small k-values or large scales (this simple evolution is a consequence of linear growth). On small scales or large k-values, one has to take into account the additional effects of the non-linear evolution due to the gravitational instability of perturbations.

As $P(k)$ evolves in z, the smaller k-modes start entering the non-linear regime. The size or mass of the collapsed structures being formed out of gravitational collapse thus gets bigger with time. This, in short, is the essence of the bottom-up 'hierarchical mode

*This essentially states that the values are Gaussian-distributed

of structure formation'. It is, therefore, instructive to smooth the density field $\delta\boldsymbol{x}$ with top-hats of different sizes (we only consider real-space spherical top-hats of radius R).

In addition, the normalisation of the power spectrum of fluctuations is not set by inflation and has to be determined observationally. The variance of the smoothed density field when smoothed with a real-space top-hat of radius R is given by

$$\sigma^2(M) = \sigma^2(R) = \int_0^\infty \frac{\mathrm{d}k}{2\pi^2} k^2 P(k) \left[\frac{3j_1(kR)}{kR}\right]^2 \tag{1.29}$$

Here, $M = \frac{4\pi\overline{\rho}_m R^3}{3}$ is the matter contained in radius R with $\overline{\rho}_m$ being the current average matter density, and $j_1(x) = (sinx - xcosx)/x^2$ is the first-order Bessel function. $P(k)$ is normalised using the value of $\sigma \equiv \sigma(R = 8\ h^{-1}\ \mathrm{Mpc})$.

1.4.2. Non-linear evolution

When the fluctuations in the δ become of the order unity, the corresponding k-mode becomes non-linear and has to be treated in a fully non-linear fashion. This is a non-trivial problem; analytic solutions exist only for very few highly symmetric cases. One of these is the case of symmetric spherical collapse case.

Consider a spherically symmetric overdensity of radius r. The dynamic of this spherical top-hat region is given by the Newtonian equation

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \Omega_\Lambda H_0 r - \frac{GM}{r^2}, \tag{1.30}$$

where M is the mass contained inside the sphere. The first term on the right is a correction factor to compensate for the cosmological constant. The overdense regions will collapse to a gravitationally stable structure in due course. The initial evolution of the overdensity follows the linear formula until it reaches order unity. Thereafter, the spherical region separates from the background Universe, and its dynamics becomes independent of the regions outside it. The top hat will collapse to a dark matter halo when the perturbation, when extrapolated using linear evolution, has reached a critical overdensity given by

$$\delta_{\mathrm{crit}}(z) = \frac{1.686}{D(z)}. \tag{1.31}$$

Note that $D(z = 0) = 1$.

1.4.3. *Halo mass functions*

The formation of a stable gravitationally bound mass of a group of particles (be it stars, dark matter, etc.) is termed virialisation. In the case of dark matter, the process is achieved via violent relaxation (Lynden-Bell, 1967). This is characterised by the rapid restructuring of the collapsing particles such that the pressure balances the gravitational force. The velocities of the constituent particles are also randomised, and the structure as a whole does not undergo any changes. The virialised object will obey the virial theorem: The total kinetic energy (T) and potential energy (V) of the system are related by

$$2T + V = 0 \tag{1.32}$$

Apart from the mechanism of the formation of the virialised structures, a successful model of structure formation should also be able to predict the abundance of these entities. The first attempts along these lines were by Press & Schechter (1974). Assuming the density perturbations to be a realisation of a Gaussian random field, they followed the gravitational evolution of these perturbations in the linear regime and spherically symmetric collapse when the perturbations go over to the non-linear regime. If one smooths the overdensity field over a mass scale m (equation1.29) giving a δ_m field, and if this is assumed to be a Gaussian distribution of zero mean and variance $\sigma^2(m)$ then the probability that $\delta_m > \delta$ for some value of δ is

$$p(\delta, m) = \int_{\delta}^{\infty} \mathrm{d}\delta_m \frac{1}{(2\pi)^{1/2}\sigma(m)} \exp\left[-\frac{\delta_m^2}{2\sigma^2(m)}\right] = \frac{1}{2}\,\mathrm{erfc}\left[\frac{\delta}{\sqrt{2}\sigma(m)}\right] \tag{1.33}$$

where $\mathrm{erfc}(x)$ is the error function.

If this is now interpreted as the fraction of dark matter that are part of virialised haloes of mass M at redshift z, and that $\delta = \delta_{\mathrm{crit}}$ which is the critical overdensity for spherical collapse under linear theory, the mass fraction in haloes above mass M at z is given by

$$F(> m \mid z) = \mathrm{erfc}\left[\frac{\delta_{\mathrm{c}}(z)}{\sqrt{2}\sigma(m)}\right] \tag{1.34}$$

Note that there is an additional factor of 2, which compensates for the 'cloud-in-cloud' problem as shown by Bond et al. (1991).

In order to get the abundance of the haloes as a function of their mass known as the *halo mass function*, we differentiate equation1.34. If dn is the comoving number density of dark matter haloes of mass in the range m and $m + dm$,

$$\frac{\mathrm{d}n_{\mathrm{h,PS}}(m,z)}{\mathrm{d}m} = -\sqrt{\frac{2}{\pi}}\frac{\bar{\rho}_{\mathrm{m}}}{m}\frac{\mathrm{d}[\ln\sigma(m)]}{\mathrm{d}m}\nu_{\mathrm{c}}\exp\left(-\frac{\nu_{\mathrm{c}}^2}{2}\right) \tag{1.35}$$

where $\nu_c = \delta_{\mathrm{crit}}(z)/\sigma(m)$ quantifying standard deviations that the the critical over density has on a particular mass scale m.

There are a number of updates to the Press-Schechter (PS) HMF in the literature (for example, Tinker et al., 2008; Watson et al., 2013). Sheth, Mo & Tormen (2001) gives a form that agrees with numerical simulations, particularly at lower z

$$\frac{\mathrm{d}n_{\mathrm{h,ST}}(m,z)}{\mathrm{d}m} = -\sqrt{\frac{2}{\pi}}\frac{\bar{\rho}_{\mathrm{m}}}{m}\frac{\mathrm{d}[\ln\sigma(m)]}{\mathrm{d}m}A\left(1+\frac{1}{\hat{\nu}_{\mathrm{c}}^{2p}}\right)\hat{\nu}_{\mathrm{c}}\exp\left(-\frac{\hat{\nu}_{\mathrm{c}}^2}{2}\right) \tag{1.36}$$

where $\hat{\nu}_{\mathrm{c}} = \sqrt{a}\delta_{\mathrm{crit}}/\sigma(m)$ with $a = 0.73$, $A = 0.353$ and $p = 0.175$. The Sheth-Mo-Tormen (SMT) HMF reduces to the PS HMF for $A = 0.5$, $a = 1$, and $p = 0$. Figure 1.5 shows the PS and SMT HMFs for $z = 30, 20, 15, 10, 7, 5$ and it is evident that the PS HMF underpredicts the halo number density, particularly at the high mass end.

At the same time, it is worth pointing out that none of these formalisms deal with the question of how the haloes are related to one another. In the cold dark matter cosmology, the haloes form in a hierarchical fashion primarily via accretion and merging. Hence, large, massive haloes are formed from the merging of smaller, less massive haloes that form early.

The next question is the spatial distribution of these dark matter haloes. Galaxy surveys have established that the galaxy distribution has extended large-scale features such as cosmic filaments, cosmic voids and regions of higher number density (see Figure 1.6 for the distribution of galaxies around us from the SDSS survey (York et al., 2000; Almeida et al., 2023)). This distribution is termed the cosmic web and is also borne out in N-body simulations (Springel et al., 2005; Vogelsberger et al., 2020).

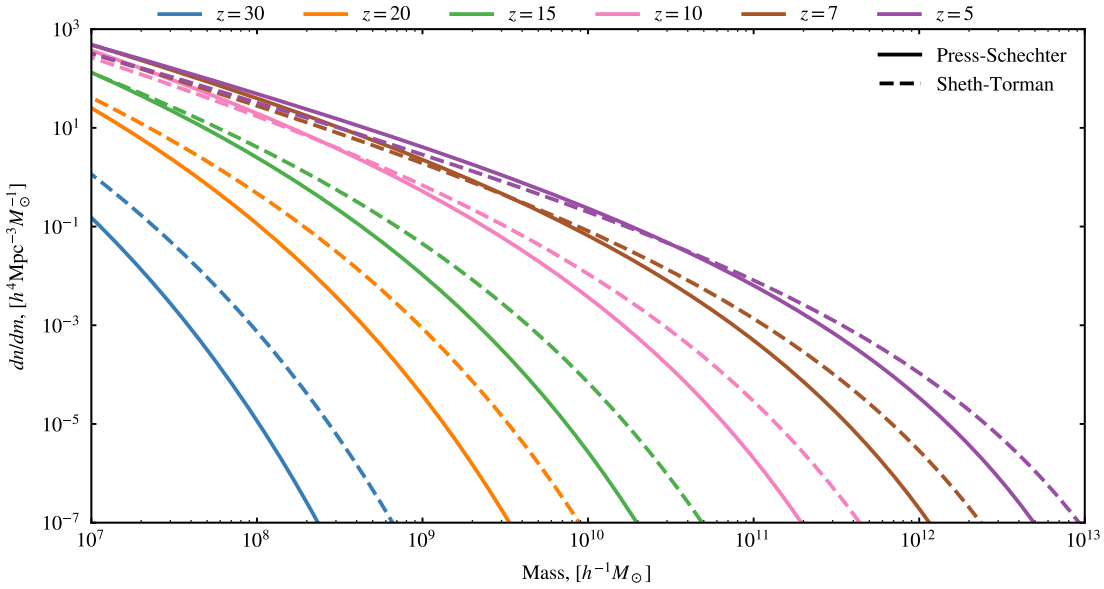

Figure 1.5: Figure shows the halo mass functions (HMFs) for $z \in [5, 7, 10, 15, 20, 30]$. The solid curves represent the Press & Schechter (1974) HMFs, and the dashed curves represent Sheth et al. (2001) HMFs.

1.4.4. Baryonic matter

So far, we have ignored the baryonic matter in the Universe and have focused exclusively on the gravitational evolution and distribution of the dark matter. This is justifiable in that dark matter constitutes the majority of the mass-energy density of the matter content of the Universe. Nevertheless, the only observable part of the Universe is the luminous baryonic content, primarily in the form of stars and galaxies.

One of the earliest theoretical works along these lines was White & Rees (1978) who combined the insights from the PS formalism with the understanding of the cooling of gas to produce a model of galaxy formation and evolution. First, the dissipationless dark matter evolves under the gravitational force to produce virialised dark matter haloes. Second, the baryons will cool down towards the centre of the potential well of these haloes to form galaxies. The baryonic gas has to cool efficiently to achieve densities that can result in nuclear fusion. This is usually achieved by the baryonic gas cloud fragmenting into smaller clouds that will subsequently result in the formation of stars. Thermal energy needs to be lost by the gas cloud for this condensation to occur. This is accomplished primarily via radiative cooling, resulting in the lowering of the internal energy of the system.

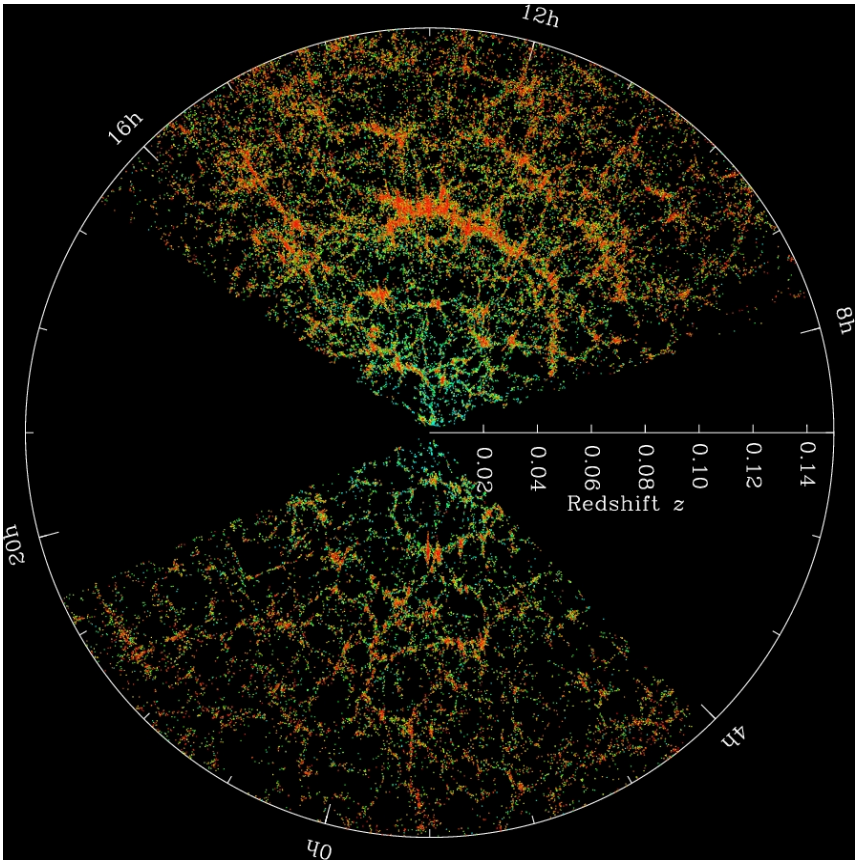

Figure 1.6: Distribution of more than one million galaxies from the SDSS survey. Each dot in the map is a galaxy, and the colour is indicative of local galaxy density (green is less dense, and red is high density). *Image credit:* SDSS survey

These ideas were first implemented in White & Frenk (1991) in what was arguably the first semi-analytical models of galaxy formations. This model had imprints of most of the important physical processes relevant for galaxy formation, namely, cold dark matter, gas cooling, star formation, feedback and stellar populations (Cole, 1991; Kauffmann et al., 1993; Cole et al., 1994).

This first generation of stars and galaxies had a profound impact on the intergalactic medium (IGM), altering its ionisation state. The efficiency of the cooling of the IGM is very sensitive to the ionisation state of the IGM. These first galaxies, therefore, impacted all subsequent galaxy formation and evolution.

1.5. The Epoch of Reionisation

The history of the Universe is written in HI. After the last scattering ($z \sim 1100$), the Universe rapidly became neutral, i.e., most of the baryonic content of the Universe was in the form of neutral hydrogen and helium atoms. The ensuing period is termed the 'Cosmic Dark Ages' as there were no luminous sources. The term is even more apt when one considers that even if there were any light emitters, all the photons blue-ward of the Lyman limit are scattered/absorbed very strongly by HI, rendering the IGM opaque.

We then have the 'Cosmic Dawn' roughly 500 million years after the Big Bang ($z \sim 60 - 30$), when the first large-scale structures started forming under the influence of gravity. The baryons in the IGM fall into these gravitationally bound dark matter haloes, and this process culminates in the formation of the first stars and galaxies. The Universe had the first luminous sources. The first stars formed out of the pristine metal-free gas, and these population III stars would have been more massive than their metal-polluted counterparts. Massive stars are naturally hotter and hence will have an emission that peaks in shorter wavelengths compared to the Sun.

The intense ionising UV photons that are emitted by the stars in the first galaxies will carve out ionised HII bubbles in the IGM (kind of similar to the Stromgren spheres in the ISM around the O & B stars). As more and more galaxies start appearing, as well as with the growth of a galaxy, these HII bubbles will increase in size and fuse to fill the entire IGM. The Universe has thus become ionised once again. This period is termed the Epoch of Reionisation (EoR). Figure 1.7 shows a schematic timeline of the Universe with a focus on the EoR. This thesis is concerned with this last evolutionary stage of hydrogen in the Universe, during the EoR.

A number of observational probes have been utilised to probe this period. The primary one is the 21-cm hyperfine transition of the HI atoms, which is the primary topic of this thesis. Before going into the details of the HI 21-cm signal, I will first summarise the other probes of the EoR.

1.5.1. Observational probes of the EoR

In this subsection, I list two observational probes of the EoR.

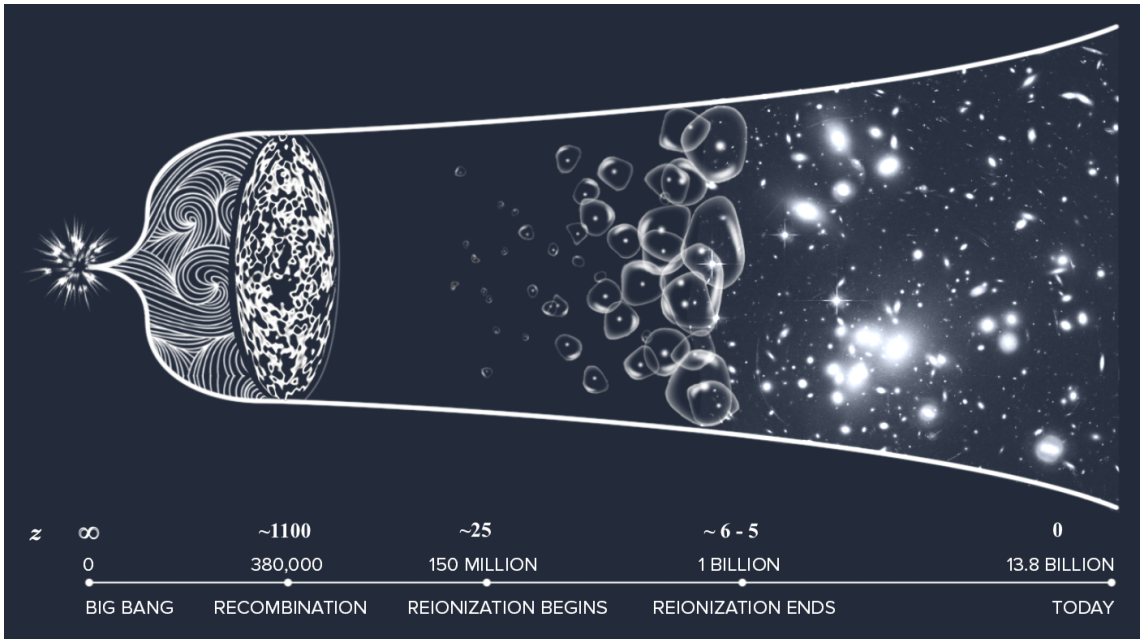

Figure 1.7: A timeline of our Universe. *Image credit:* Aman Chokshi, University of Melbourne.

Gunn-Peterson effect

The Lyman-α (Lyα) photon ($n = 2$ to $n = 1$ transition of HI) is a highly resonant line of neutral hydrogen. The optical depth of a Lyα photon emitted at redshift z and a neutral hydrogen fraction of $x_{\rm HI}$ was worked out by Gunn & Peterson (1965) (also see Scheuer 1965) to be

$$\tau_{\rm GP}(z) = 1.8 \times 10^5 h^{-1} \Omega_M^{-1/2} \left(\frac{\Omega_b h^2}{0.02}\right) \times \left(\frac{1+z}{7}\right)^{3/2} \left(\frac{n_{\rm HI}}{n_{\rm H}}\right), \tag{1.37}$$

The large value of this optical depth, as is evident from equation1.37, implies that the IGM will be opaque, due to scattering, to the Lyα line even with trace amounts of HI (as small as $x_{\rm HI} \sim 10^{-5}$). As the light emitted by sources at cosmological distances gets redshifted into the Lyα frequency, it will be absorbed by the HI in the IGM. At low redshifts (towards the end of reionisation), HI is present in HI clouds along the line-of-sight. The photons that redshift to the Lyα frequency will, therefore, get scattered by these intervening HI. The spectra of these sources will thus have the so-called Lyα-forest that are absorption features blueward of the Lyα frequency. For sources that are further away, where every photon blue-ward will be absorbed, we have the so-called *Gunn-Peterson trough* where there will be a sudden drop in the spectra. The trough was

first detected in the SDSS quasar spectra by Becker et al. (2001). The $z = 6.28$ quasar spectra from their work (reproduced here as Figure 1.8) shows the absorption trough extending from $z = 5.95$ to $z = 6.15$. The other panels progressively show the nature of the effect moving towards that of the Lyα-forest. This clearly demonstrates the end of EoR when the Universe moved towards an ionised state.

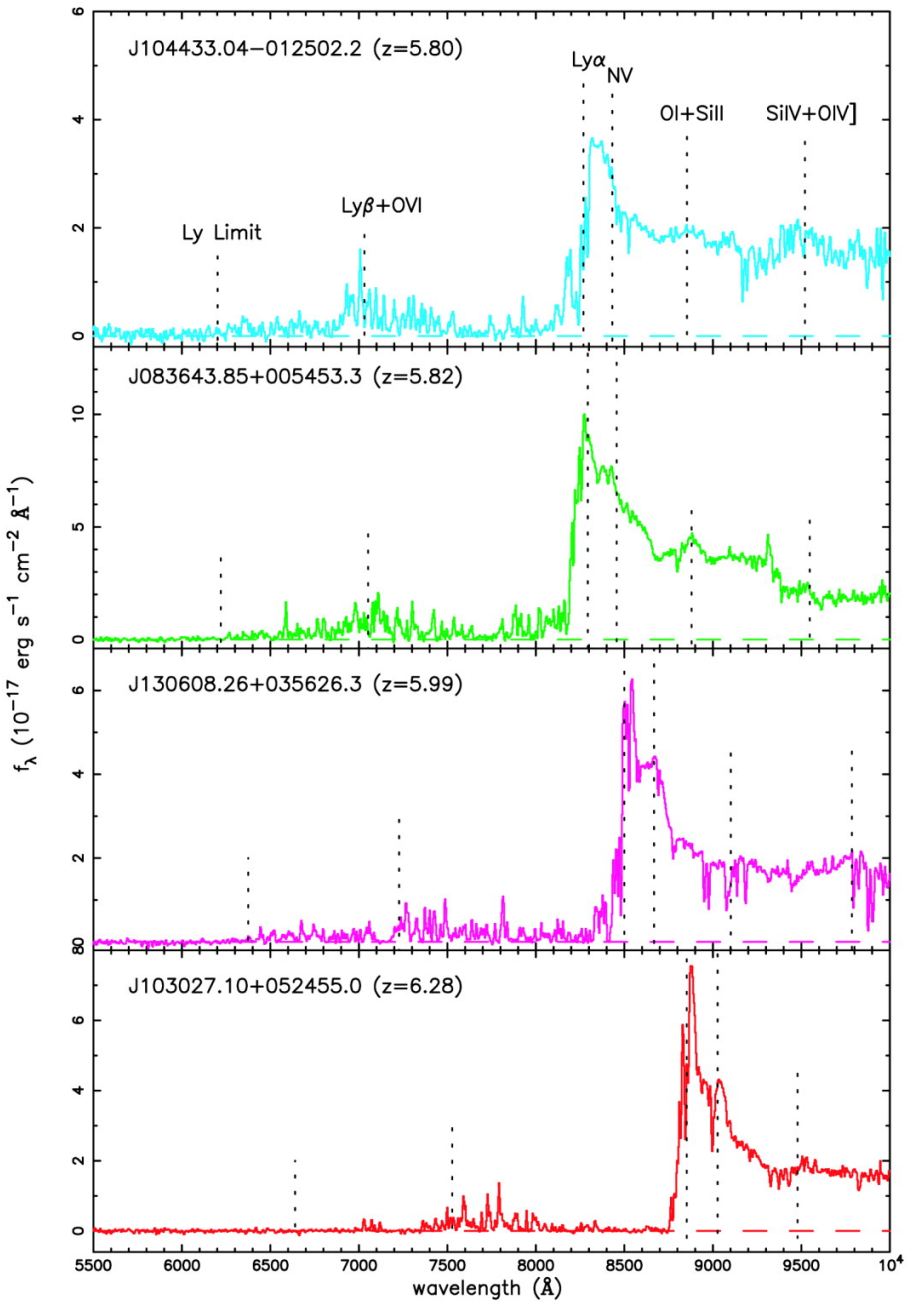

Figure 1.8: Spectra of high-redshift (> 5.8) quasars clearly demonstrating the Gunn-Peterson effect. The presence of the Gunn-Peterson trough in the spectra of the $z = 6.28$ quasar (last panel) is clear evidence for reionisation. The dashed lines indicate prominent emission lines and Lyman limit. *Image credit:* Becker et al. (2001)

CMB optical depth

Another primary probe that constrains the timing and duration of the EoR is the integrated optical depth, τ_e of the CMB photons due to Thompson scattering off the free electrons in the IGM. Assuming helium is singly ionised and is produced at the same rate as hydrogen (HII),

$$\begin{aligned}\tau_{\rm e} = &\int_{z=0}^{\infty} \frac{c\,{\rm d}t}{{\rm d}z}(1+z)^3\sigma_{\rm T} \\ &\times\left[Q_{\rm HII}^{\rm m}\langle n_{\rm H}\rangle + (Q_{\rm HeII}^{\rm m} + 2Q_{\rm HeIII}^{\rm m})\langle n_{\rm He}\rangle\right]{\rm d}z,\end{aligned} \tag{1.38}$$

where $\sigma_T = 6.652\times10^{-25}$ cm^2 is Thompson scattering cross-section, $Q_X^{\rm m}$ denotes the mass-weighted global fraction of species X, $\langle n_H\rangle = 1.88\times10^{-7}(\Omega_b h^2/0.022)$ cm^{-3} is the average comoving number density of Hydrogen, and $\langle n_{He}\rangle = 0.148\times10^{-7}(\Omega_b h^2/0.022)$ cm^{-3} is that of Helium (Wyithe & Loeb, 2003).

1.6. 21-cm Cosmology

The hyperfine spin-flip transition is one of the primary probes employed in astronomy for detecting the HI in the Universe. The 21-cm line has led to a number of paradigms in astronomy. The 21-cm line is the *only* known astrophysical probe that can give information about the Universe between the time of the last scattering and the formation of the first luminous sources. We now turn to the physics of this spectral line.

1.6.1. The physics of the 21-cm line

First predicted by van de Hulst (1945) during the Second World War, the hyperfine transition of HI atoms is a quantum-mechanically forbidden transition. In fact, the higher energy excited state has a mean lifetime of $\sim 10^7$ years. The saving grace here is that, to first order, hydrogen comprises the whole baryonic content of the Universe. This directly results in a non-zero detectable signal.

The $n = 0$ electronic ground state of HI is non-degenerate with respect to the spin-alignment of the proton and electron. The spin of the electron (m_e) and the proton (m_p) are coupled to each other and lead to the lifting of the degeneracy. The singlet anti-

parallel state $\left|m_p = 1, m_e = -1\right\rangle$ and the excited triplet parallel spin state $\left|m_p = 1, m_e = 1\right\rangle$. This *hyperfine* splitting corresponds to an energy difference of $E_{21} = 5.87 \times 10^{-6}$ eV. The rest-frame frequency is ~ 1420 MHz, corresponding to a wavelength $\lambda \sim 21$ cm.

1.6.2. 21-cm brightness temperature, T_b

The hyperfine level populations are given by

$$\frac{n_1}{n_0} = 3\, e^{-\frac{E_{21}}{k_{\rm B} T_{\rm S}}} \tag{1.39}$$

where n_1 is the number of atoms in the triplet state (hence the factor of 3), n_0 is the number of atoms in the ground state, $k_{\rm B}$ is the Boltzmann constant, and $T_{\rm S}$ is termed the spin-temperature that quantifies the level populations. It is evident from the equation that the level populations are dependent only on $T_{\rm S}$. It is customary in the literature to rewrite the above formula in terms of $T_* = 0.07$ K defined as $k_{\rm B} T_* = E_{21}$.

The spontaneous decay rate for the hyperfine transition is $A_{10} = 2.9 \times 10^{-15}\ s^{-1}$. The spin states will thus reach thermal equilibrium with the CMB, i.e. $T_{\rm S} = T_{\rm CMB} = 2.73\ (1+z)$ K, on a time-scale of

$$t_{\rm eq} = \frac{T_*}{A_{10}\ T_{\rm CMB}} \simeq \frac{3 \times 10^5}{1+z}\ \text{yr}. \tag{1.40}$$

Note that this value is less than the age of the Universe for all redshifts that concern 21-cm cosmology.

The key physical parameter to understand is the spin temperature $T_{\rm S}$. Different astrophysical phenomena variably impart distinct signatures in the evolution of $T_{\rm S}$. The IGM is only observable when $T_{\rm S}$ is different from $T_{\rm CMB}$. This is usually achieved when $T_{\rm S}$ is coupled to the kinetic temperature $T_{\rm K}$ of the HI gas.

The fundamental quantity of interest in 21-cm EoR cosmology is the *brightness temperature,* $T_{\rm b}$ of the 21-cm line. Consider a cloud of HI . The *specific intensity* or *brightness* I_ν is the energy carried by a ray emerging from this cloud at frequency ν in a given direction, per unit area, frequency, solid angle, and time. I_ν thus has units of erg s^{-1} cm^{-2} sr^{-1} Hz^{-1}. Radio astronomers historically represent I_ν in the temperature scale by considering the temperature of a blackbody (of spectrum B_ν) such that $I_\nu = B_\nu(T_{\rm b})$. This

analysis is considerably simplified by realising that the blackbody spectrum can be approximated by the Rayleigh-Jeans formula at the frequencies that we are interested in as:

$$T_{\rm b}(\nu) \approx \frac{I_\nu c^2}{2k_{\rm B}\nu^2} \tag{1.41}$$

where c is the speed of light.

The emergent brightness temperature along a column of the HI cloud is:

$$T_{\rm b}(\nu) = T_\gamma\ (1 - e^{-\tau_\nu}) + T_{\rm S}(\nu)\ e^{-\tau_\nu} \tag{1.42}$$

where $\tau_\nu = \int \alpha_\nu\ {\rm d}s$ is the optical depth with α_ν the absorption coefficient along the proper distance s. T_γ is the incident background radiation. The optical depth is given by:

$$\begin{aligned}\tau_{\nu_0} &= \frac{3}{32\pi}\frac{hc^3 A_{10}}{k_{\rm B} T_S \nu_0^2}\frac{x_{\rm HI} n_H}{(1+z)\left({\rm d}v_\parallel/{\rm d}r_\parallel\right)} \\ &\approx 0.0092(1+\delta)(1+z)^{3/2}\frac{x_{\rm HI}}{T_S}\left[\frac{H(z)/(1+z)}{{\rm d}v_\parallel/{\rm d}r_\parallel}\right],\end{aligned} \tag{1.43}$$

where δ denotes baryonic overdensity, $x_{\rm HI}$ is the neutral fraction, and ${\rm d}v_\parallel/{\rm d}r_\parallel$ is the line-of-sight gradient of the proper velocity.

The differential brightness temperature from redshift z, the primary observational signature in 21-cm cosmology, is given by an offset against the background temperature as

$$\begin{aligned}\delta T_{\rm b}(\nu) &= \frac{T_{\rm b}(\nu) - T_\gamma}{1+z} \\ &= \frac{T_{\rm S} - T_\gamma}{1+z}(1 - e^{-\tau_{\nu_0}}) \approx \frac{T_{\rm S} - T_\gamma}{1+z}\tau_{\nu_0} \\ &\approx 27 x_{\rm HI}(1+\delta_{nl})\left(\frac{H}{dv_r/dr + H}\right)\left(1 - \frac{T_\gamma}{T_{\rm S}}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_M h^2}\right)\left(\frac{\Omega_{\rm b}h^2}{0.023}\right)\ {\rm mK},\end{aligned} \tag{1.44}$$

For the rest of this work, we will set $T_\gamma = T_{CMB}$.

δT_b saturates when $T_S >> T_\gamma$; the HI 21-cm signal, in this case, will be in emission at a level that is independent of T_S. On the other hand, the IGM will be seen in absorption (at arbitrarily large values) when $T_S < T_\gamma$. Any astrophysical processes that can thus impact T_S will hence leave their imprints on the 21-cm signal.

1.6.3. *The evolution of the spin temperature, T_S*

The three (competing) astrophysical processes that determine the evolution of the spin temperature T_S are

1. absorption and stimulated emission from and to the CMB background, respectively,
2. collisions with the other hydrogen atoms as well as free electrons and protons, and
3. scattering of Lyα photons, termed the Wouthusyen-Field effect.

The rates corresponding to these effects are sufficiently high that the spin temperature is given by the equilibrium balance as:

$$T_S^{-1} = \frac{T_{CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}, \tag{1.45}$$

where x_α and x_c are, respectively, the coupling coefficient of Lyα scatterings and atomic collisions, T_α is the *colour* temperature of the Lyα radiation field, and T_K is the gas kinetic temperature.

It is instructional to pause here to reflect that among the physical quantities introduced so far that have units of temperature, only T_{CMB} and T_K are true thermodynamic temperatures. δT_b is a measure of the radiation intensity, T_S is a measure of the hyperfine level populations, and T_α is used to describe the photon distribution around the Lyα transition.

2

1.6.4. *21-cm global signal, $\overline{\delta T}_b$*

The zeroth-order information that can be obtained from the 21-cm sky is its mean value. This sky-averaged quantity is termed the *21-cm global signal.*

Different cosmic epochs leave distinct features in the 21-cm signal.

- $z \gtrsim 200$: After recombination, the gas is thermally coupled to the CMB and therefore $T_K = T_{CMB}$. The relatively high density of the IGM results in the $T_S = T_{CMB}$. As a result, there is no 21-cm signal from this epoch

- $200 \gtrsim z \gtrsim 40$: The IGM cools adiabatically during this period so that $T_{\rm K} < T_{\rm CMB}$. The IGM will now be seen in absorption as the gas is still collisionally coupled to give $T_{\rm S} = T_{\rm K}$

- $40 \gtrsim z \gtrsim z_*$: The spin temperature $T_{\rm S} = T_{\rm CMB}$ as the IGM density falls below the threshold for effective collisional coupling. The signal, therefore, decays to zero. z_* denotes the redshift when the first sources turn on.

- $z_* \gtrsim z_\alpha$: The Lyα photons emitted by these sources will couple $T_{\rm S}$ to $T_{\rm K}$, resulting in an absorption signal. By z_α, the gas will be strongly coupled.

- $z_\alpha \gtrsim z_{\rm h}$: The X-rays in the early Universe are now capable of influencing the HI, primarily by heating up the gas. This will drive up the $T_{\rm K}$, and consequently $T_{\rm S}$. By $z_{\rm h}$, the HI gas would be heated by the X-rays. The signal transitions from absorption to emission.

- $z_{\rm h} \gtrsim z_{\rm r}$: The EoR is well in progress. Heating from X-rays drives the $T_{\rm S} >> T_{\rm CMB}$, and the signal saturates. As reionisation proceeds, the neutral fraction drops and by $z_{\rm r}$, the signal will also fade away to zero.

1.6.5. 21-cm power spectrum, $\Delta^2_{21}(k)$

There is considerably more information to be gleaned from the fluctuations in the 21-cm sky as a function of scale. A primary statistic in this regard is the 21-cm power spectrum.

For a three-dimensional differential brightness temperature field, $\delta T_{\rm b}(\boldsymbol{x})$, we can define the fractional perturbation as

$$\delta_{21}(\boldsymbol{x}) = \frac{\delta T_{\rm b}(\boldsymbol{x}) - \overline{\delta T}_{\rm b}}{\overline{\delta T}_{\rm b}}, \tag{1.46}$$

where $\overline{\delta T}_{\rm b}$ is the average value. Note that we have dropped the dependence of the signal on redshift z.

We are now interested in the Fourier transform of this zero-mean $\delta_{21}(\boldsymbol{x})$ random field defined such that:

$$\delta_{21}(\boldsymbol{x}) = \int \frac{{\rm d}^3\boldsymbol{k}}{(2\pi)^3}\, \tilde{\delta}(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{1.47}$$

The 21-cm power spectrum $P_{21}(\boldsymbol{k})$ is defined as the 2-point correlation function of this $\tilde{\delta}_{21}(\boldsymbol{k})$.

$$\left\langle \tilde{\delta}_{21}(\mathbf{k}_1)\,\tilde{\delta}_{21}(\mathbf{k}_2) \right\rangle \equiv (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2)\, P_{21}(k_1) \tag{1.48}$$

We will be using the so-called dimensionless power spectrum* defined as

$$\Delta^2_{21}(k) = \frac{1}{(2\pi)^3} 4\pi k^3 P_{21}(k) \tag{1.49}$$

$\Delta^2_{21}(k)$ is a measure of the variance per logarithmic interval in wavenumber, k, of the δT_b field. This can also be viewed as the variance in the δT_b when the field is smoothed on a scale of $x = 2\pi/k$.

1.6.6. Observational frontier

It is evident that the 21-cm signal from the early Universe can shed light on a number of galaxy astrophysical processes of interest. The two primary 21-cm statistics are the 21-cm global signal and the 21-cm power spectra. There are a number of radio-frequency experiments that are trying to detect these.

To date, there has only been one claimed detection of the 21-cm global signal by the EDGES† team (Bowman et al., 2018). Figure 1.9 shows the EDGES signal, which is unexpectedly strong in absorption at $z \approx 17$. On face value, this would require the IGM to be below the adiabatic cooling limit (Barkana et al., 2018; Muñoz & Loeb, 2018; Fialkov & Barkana, 2019). Another plausibility is to posit a strong radio background in excess of the CMB at high redshift (Barkana, 2018). Others have reanalysed the EDGES signal with different models for the instrument systematics (Singh & Subrahmanyan, 2019; Sims & Pober, 2020). Additionally the signal has not been independently verified by the other experiments such as SARAS‡ which has, in fact, a null result at the same frequency and is in direct tension with the EDGES signal. A new experiment that will achieve first-light soon is REACH§.

The 21-cm PS will be detected by radio interferometers. The faintness of the 21-cm

*Though note that the *dimensionless* 21-cm power spectrum has units of $[temperature^2]$.
†Experiment to Detect the Global EoR Signature
‡Shaped Antenna measurement of the background Radio Spectrum.
§Radio Experiment for the Analysis of Cosmic Hydrogen.

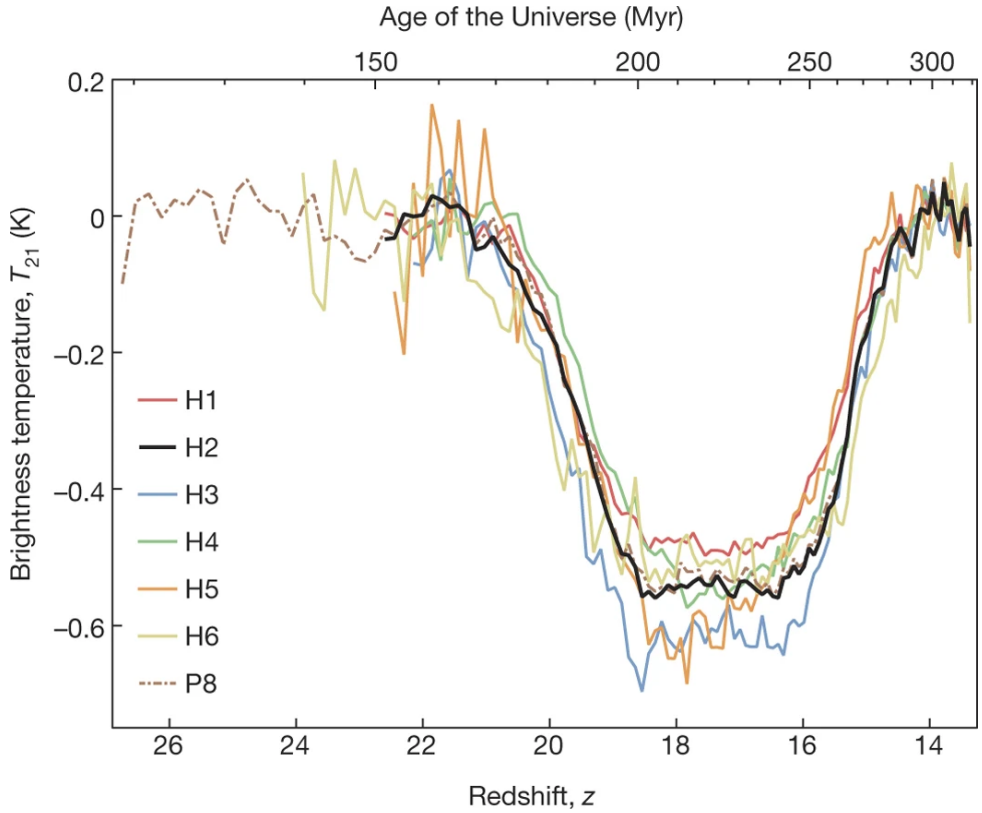

Figure 1.9: The absorption signal reported by the EDGES Bowman et al. (2018) team. The curves represent different hardware and analysis configurations as well as different foreground models. *Image credit:* Bowman et al. (2018)

signal results in the detection of the 21-cm PS, a non-trivial task. Still, the upper limits have continued to improve with advances in foreground removal strategies, a better understanding of the instrument's systematics, etc. At the time of writing, the telescopes actively reporting 21-cm PS upper limits are MWA* in Australia, LOFAR† in Netherlands, HERA‡ in South Africa, and OVRO-LWA§ in the United States of America. Figure 1.10 shows the latest upper limits on the 21-cm power spectra in the literature (Mertens et al., 2020; Rahimi et al., 2021; The HERA Collaboration et al., 2022) along with some theoretical curves (Mesinger et al., 2016). It is seen clearly that most of the theoretical curves are at least a few orders of magnitude lower than the best upper limits.

*Murchison Widefield Array.
†Low Frequency Array.
‡Hydrogen Epoch of Reionization Array
§Owens Valley Radio Observatory Long Wavelength Array

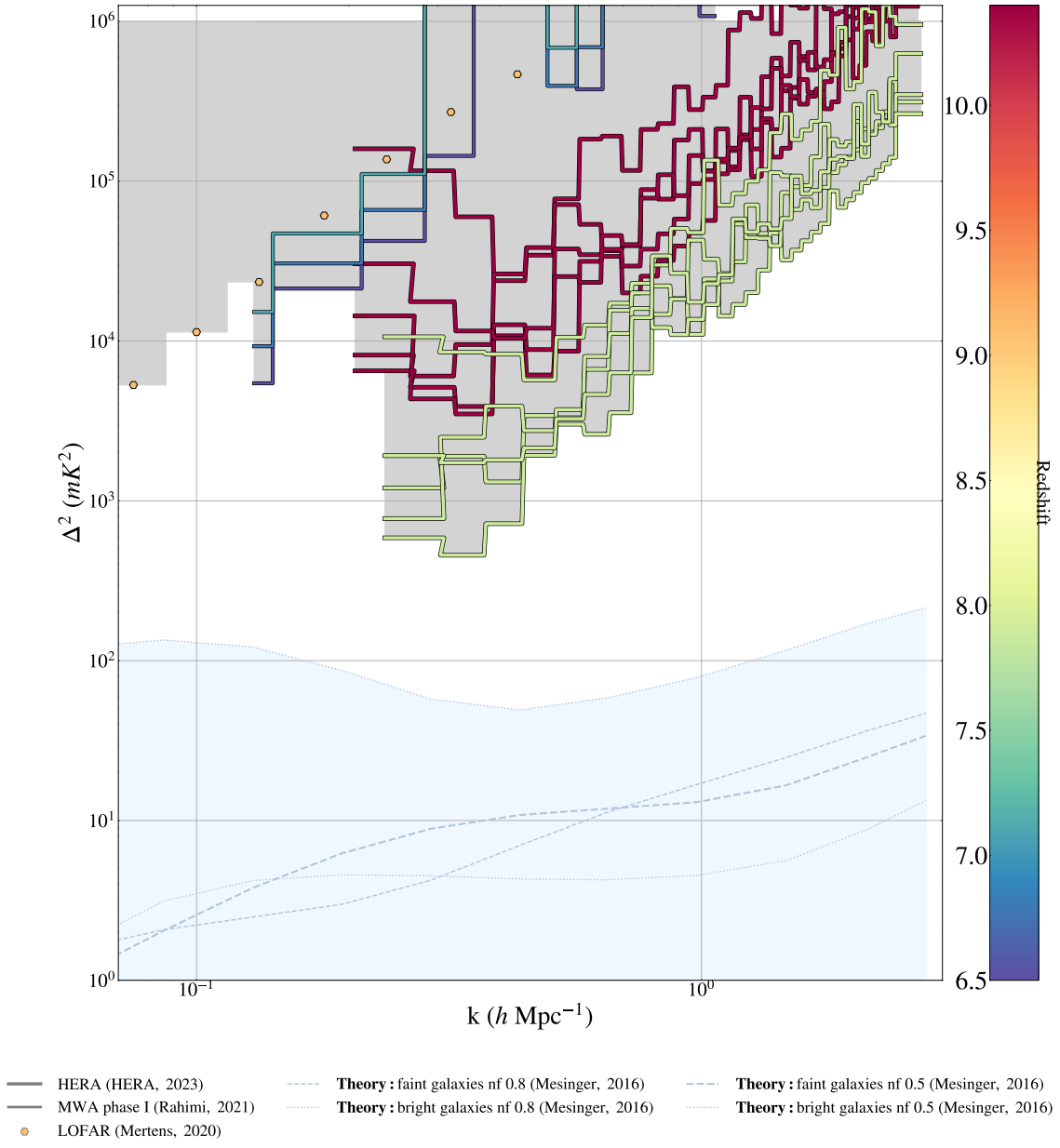

Figure 1.10: The latest 21-cm power spectrum (21-cm PS) upper limits from LOFAR (Mertens et al., 2020), MWA (Rahimi et al., 2021) and HERA (The HERA Collaboration et al., 2022). Shown also are theoretical 21-cm PS from Mesinger et al. (2016). *Modified from: github.com/EoRImaging/eor_limits*

1.7. Outline of the thesis

It is imperative that there are efficient theoretical models to take advantage of the current upper limits as well as a future detection of the 21-cm signal. The present thesis is work in this direction. Below I give a brief summary of how this thesis is structured.

In chapter 2, I introduce and briefly summarise the MERAXES semi-analytical model. A considerable part of this thesis involved updates to MERAXES , both in terms of new physics implementation as well as ones to reduce the memory overhead and speed up the calculations.

In chapter 3, I deployed MERAXES on the biggest simulation volume to date. In this work, I also explored the impact of the mass resolution and the X-rays on the predictions of the 21-cm signal from the early Universe. In addition, taking advantage of the large simulation volume, I also explored the non-Gaussianity of the 21-cm field via the cosmic variance of the 21-cm power spectra.

In chapter 4, I use the fiducial simulation from the previous chapter to forecast constraints on the astrophysics of the early galaxies that will be enabled by the detection of the 21-cm power spectra. Specifically, the analysis assumes a 1000 h observation with the upcoming Square Kilometre Array.

Chapter 5 summarises and concludes this thesis.

If you wish to make an apple pie from scratch, you must first invent the Universe.

Carl Sagan, *Cosmos*

CHAPTER 2

Modelling the high-redshift Universe

This thesis relies extensively on the semi-analytical galaxy formation model MERAXES that self-consistently couples high-z galaxy formation and evolution to reionisation. The present chapter summarises the relevant astrophysical implementation within MERAXES and the updates that were introduced into the model as part of this thesis. The reader is referred to Mutch et al. (2016), Qin et al. (2017) and Qiu et al. (2019) for more details about the model.

2.1. SIMULATIONS OF THE EOR

In this section, I briefly summarise the different kinds of simulations that are used to model the EoR. Simulations of the EoR broadly fall into three categories: hydrodynamical simulations, semi-analytical models, and semi-numerical models*. This classification is based purely on the amount of realism that goes into these models in terms of the physics implementation. Each has its own strengths and weaknesses. A recent comprehensive review on simulations of the EoR can be found in Gnedin & Madau (2022).

*I would be remiss if I do not mention the analytical works that have been developed to study reionisation (few examples being Miralda-Escude et al., 2000; Furlanetto et al., 2004; Madau & Fragos, 2017). Even though they lack detailed physics they have been greatly informative in driving the field forward. For instance, they help reduce the parameter space that can then be explored by more detailed models.

2.1.1. *Hydrodynamical simulations*

Hydrodynamical models, the most physics-rich of the simulations, involve accurately modelling the dynamics of dark matter and baryons in a cosmological volume. In addition to efficient implementations for gravitational force calculations, hydrodynamical simulations are also characterised by detailed implementations for star formation, supernova feedback, and metal enrichment, to name a few. Crucially, they can capture the feedback of baryons onto large-scale structure formation and shell-crossing. The main drawback of hydrodynamical simulations is their immense computational overhead. This precludes their use to explore the complex underlying parameter space of astrophysics and/or cosmology. Some examples of hydrodynamical simulations are Cosmic Dawn (CoDa; Ocvirk et al., 2020), Cosmic Reionisation On Computers (CROC; Gnedin, 2014), SPHINX (Rosdahl et al., 2018), and THESAN (Kannan et al., 2022).

2.1.2. *Semi-numerical models*

Semi-numerical models fall on the other end of the spectrum of cosmological simulations. They usually involve evolving Gaussian random fields (that are generated as a realisation from the primordial power spectrum) at high redshifts (typically ~ 100). This three-dimensional density field is then evolved down to low redshifts under second-order perturbation theory to the non-linear regime. The peaks in the evolved density fields are then identified as sites of galaxy formation. The main advantage of the semi-numerical models is that they are computationally cheap, thereby enabling them to sample much larger simulation volumes. Semi-numerical models are capable of efficient and fast exploration of the parameter space. Some examples of semi-numerical models of reionisation are SIMFAST21 (Santos et al., 2010; Hassan et al., 2016), 21CMFAST (Mesinger et al., 2011; Murray et al., 2020), and SCRIPT (Choudhury & Paranjape, 2018).

2.1.3. *Semi-analytical models*

Semi-analytical models (SAMs) of galaxy formation and evolution are able to give us the best of both worlds – the computational efficiency of the semi-numerical simulations plus the non-linear evolution under the gravity of the N-body simulations. The basic idea is that one 'paint' galaxies onto dark matter haloes in post-processing. SAMs usu-

ally take in dark matter halo merger trees as inputs. These are then populated by subgrid, but still physically motivated, prescriptions for different astrophysical processes, including radiative cooling, star formation, supernova and AGN feedback, metal enrichment, galaxy mergers, etc.

This thesis is concerned with the galaxy SAM Meraxes. Examples of other SAMs in the literature are Reionization using Semi-Analytic Galaxy Evolution (RSAGE; Seiler et al., 2019), Visbal et al. (2020), and seminumerical rAdiative tranSfer coupling of galaxy formaTion and Reionization in N-body dArk mattErsimUlationS (ASTRAEUS; Hutter et al., 2021).

2.2. Galaxy formation and evolution within Meraxes

We now move onto the specifics of the galaxy physics that is implemented within Meraxes.

2.2.1. *Baryonic infall*

At every snapshot of the simulation, we increase the baryonic content of each Friends-of-Friends (FoF) group in proportion to its virial mass $M_{\rm vir}$ and the universal baryonic fraction, $f_{\rm b} = \Omega_b/\Omega_m$. The freshly accreted mass is therefore given by

$$m_{\rm infall} = f_{\rm mod} f_{\rm b} M_{\rm vir} - \sum_{i=0}^{N_{\rm gal}-1} m^i_{\rm star} + m^i_{\rm cold} + m^i_{\rm hot} + m^i_{\rm ejected}, \tag{2.1}$$

where the summation is over every $N_{\rm gal}$ galaxy in the FoF halo. In equation 2.1, $m_{\rm star}$, $m_{\rm cold}$, $m_{\rm hot}$, and $m_{\rm ejected}$ denotes, respectively, the baryonic mass locked up in the form of stars, cold gas, hot gas, and ejected mass in the halo. In a manner, the accurate calculations of these mass reservoirs are the main component of Meraxes (or any SAM for that matter). The factor $f_{\rm mod}$, termed *baryonic fraction modifier*, accounts for the varied physical processes that can suppress the amount of baryonic infall that is available for galaxy formation. These can include, for example, the local UV background that can alter the cooling properties of the IGM thus increasing the local Jeans mass (Dijkstra et al., 2004), as well as the feedback from reionisation. The accurate, temporally and

spatially dependent, calculation of $f_{\rm mod}$ is what sets MERAXES apart from other SAMs in the literature.

2.2.2. *Gas cooling*

All infalling baryons are added to the *hot gas reservoir* of the halo where they are shocked to the halo virial temperature. This gas needs to cool down to the *cold gas disc* for it to take part in galaxy formation. The cooling of the hot gas follows White & Frenk (1991), where the cooling time $t_{\rm cool}$ is given by

$$t_{\rm cool}(r) = \frac{1.5\bar{\mu}m_{\rm p}k_{\rm B}T}{\rho_{\rm hot}(r)\Lambda(T,Z)}, \tag{2.2}$$

where $\bar{\mu}m_{\rm p}$ is the average particle mass, k_B is the Boltzmann constant, $\rho_{\rm hot}(r)$ is the density profile for the hot gas which is assumed to be a singular isothermal profile, and Λ is the temperature (T) and metallicity (Z) dependent cooling function (Sutherland & Dopita, 1993). Following Croton et al. (2006) we compare $t_{\rm cool}$ to the dynamical time t_{dyn}^{FoF} of the halo. All material with $t_{dyn}^{FoF} < t_{\rm cool}$ is moved to the cold gas reservoir of the halo.

2.2.3. *Star formation*

MERAXES 's star formation prescriptions follow Croton et al. (2006) and Guo et al. (2011) closely. All gas that cools down from the hot halo is deposited onto a rotationally supported exponential disc of scale radius

$$r_{\rm disc} = 3R_{\rm vir}\lambda/2, \tag{2.3}$$

where $R_{\rm vir}$ is the virial radius of the halo and λ is the spin parameter of the halo (Bullock et al., 2001).

A critical mass $m_{\rm crit}$ needs to be achieved for the cold gas disc to undergo fragmentation and start to form stars

$$m_{\rm crit} = \Sigma_{\rm SF}\left(\frac{V_{\rm max}}{100\ {\rm km\ s^{-1}}}\right)\left(\frac{r_{\rm disk}}{10{\rm kpc}}\right) \times 10^{10}{\rm M}_{\odot}, , \tag{2.4}$$

where $V_{\rm max}$ is the maximum circular velocity of the halo. The star formation rate (SFR; $\dot{m}_{\rm star}$) is then given by

$$\dot{m}_{\rm star} = \alpha_{\rm SF} \frac{m_{\rm gas} - m_{\rm crit}}{t_{\rm disc}^{\rm dyn}}, \tag{2.5}$$

where $t_{\rm disc}^{\rm dyn}$ is the dynamical time of the disc and $\alpha_{\rm SF}$ is the star formation efficiency.

The starbursts from galaxy mergers are implemented according to Somerville et al. (2001) as

$$\frac{m_{\rm burst}}{m_{\rm cold}} = \alpha_{\rm burst} \left(\frac{m_{\rm gal}}{m_{\rm parent}} \right)^{\gamma_{\rm burst}}, \tag{2.6}$$

where $m_{\rm gal}$ and $m_{\rm parent}$ are the masses of the infalling and parent galaxies. The free parameters $\alpha_{\rm burst}$ and $\gamma_{\rm burst}$ are set to 0.57 and 0.7, respectively, following Croton et al. (2006)

2.2.4. *Supernova feedback*

The primary effect of the Supernova feedback is to inject energy into the cold gas component of the halo. This results in a portion of the cold gas mass being moved to the hot halo component. For extremely energetic supernovas or for galaxies, usually low-mass ones, with a shallow gravitational potential, the gas can even be ejected from the halo altogether.

MERAXES calculates the mass of cold gas that is reheated by supernova energy injection as

$$\Delta m_{\rm reheat} = \begin{cases} \eta \Delta m_{\rm new}, & \Delta E_{\rm SN} \geq \Delta E_{\rm hot} \\ \frac{\Delta E_{\rm SN}}{1/2 V_{\rm vir}^2}, & \Delta E_{\rm SN} < \Delta E_{\rm hot} \end{cases} \tag{2.7}$$

where

$$\Delta E_{\rm hot} = \frac{1}{2} \eta \Delta m_{\rm new} V_{\rm vir}^2, \tag{2.8}$$

η is the mass loading factor, $\Delta m_{\rm new}$ is the mass of newly formed stars, $V_{\rm vir}$ is the virial velocity of the halo. $\Delta E_{\rm SN}$ is the energy injected by the supernova into the ISM and is given by

$$\Delta E_{\rm SN} = \epsilon \times \int_{t}^{t+\Delta t} {\rm d}t' \int_{0}^{\infty} {\rm d}\tau \frac{{\rm d}\varepsilon}{{\rm d}\tau} \psi(t'-\tau), \tag{2.9}$$

where ϵ is the energy coupling efficiency, t (Δt) is the simulation time (time-step), $({\rm d}\varepsilon/{\rm d}\tau)$ is the rate of energy released by stars of age between τ and $\tau + {\rm d}\tau$ as Type-II supernova, and $\psi(t)$ is the SFR. The metallicity dependent ε is calculated directly from the stellar population code STARBURST99 (Leitherer et al., 1999). Throughout this work, we assume a Kroupa (2002) initial mass function.

Due to the high cadence of the simulation, stars can live for a while and only go supernova after several snapshots. Therefore the mass of the newly formed stars $\Delta m_{\rm new}$ is calculated by weighting the average star formation history with the supernova energy as

$$\Delta m_{\rm new} = \frac{\int_{t}^{t+\Delta t} {\rm d}t' \int_{0}^{\infty} {\rm d}\tau \frac{{\rm d}\varepsilon}{{\rm d}\tau} \psi(t'-\tau)}{\int_{0}^{\infty} {\rm d}\tau \frac{d\varepsilon}{{\rm d}\tau}}. \tag{2.10}$$

The free parameters η and ϵ are implemented as

$$\epsilon = \begin{cases} \epsilon_0 \left(\frac{1+z}{4}\right)^{\alpha_{\rm eject}} \left(\frac{V_{\rm max}}{70{\rm kms}^{-1}}\right)^{-1}, & V_{\rm max} \geq 70{\rm kms}^{-1} \\ \\ \epsilon_0 \left(\frac{1+z}{4}\right)^{\alpha_{\rm eject}} \left(\frac{V_{\rm max}}{70{\rm kms}^{-1}}\right)^{-3.2}, & V_{\rm max} < 70{\rm kms}^{-1} \end{cases}, \tag{2.11}$$

and,

$$\eta = \begin{cases} \eta_0 \left(\frac{1+z}{4}\right)^{\alpha_{\rm reheat}} \left(\frac{V_{\rm max}}{60{\rm kms}^{-1}}\right)^{-1}, & V_{\rm max} \geq 60{\rm kms}^{-1} \\ \\ \eta_0 \left(\frac{1+z}{4}\right)^{\alpha_{\rm reheat}} \left(\frac{V_{\rm max}}{60{\rm kms}^{-1}}\right)^{-3.2}, & V_{\rm max} < 60{\rm kms}^{-1} \end{cases} \tag{2.12}$$

Throughout this thesis, we set $\alpha_{\rm reheat} = 2$ and $\alpha_{\rm eject} = 0$. The values of the other parameters are discussed in chapter 3.

2.2.5. *Galaxy UV Luminosities and dust models*

MERAXES also has implementations to calculate the UV luminosities for every galaxy. This is accomplished by generating a full spectral energy distribution (SED) for every galaxy and applying a tophat filter centered on $\lambda = 1600$ Å as

$$L_{\rm UV}(t) = \int_{1550}^{1650} d\lambda \int_0^t dt \int_{Z_{\rm min}}^{Z_{\rm max}} dZ \psi(t-\tau, Z) S_\lambda(\tau, Z) T_\lambda(\tau) \tag{2.13}$$

Here, Z is the metallicity, $\psi(t-\tau, Z)$ is the star formation rate, and $S_\lambda(\tau, Z)$ is the luminosity of a stellar population of age τ, and $T_\lambda(\tau)$ is the dust transmission function. S_λ is calculated from Starburst99 (Leitherer et al., 1999).

The dust model employed throughout this thesis follows the two-phase model of Charlot & Fall (2000). The stars are assumed to form out of dense birth clouds that live for $t_{\rm BC}$. Hence, the photons emitted by the young stars (of age less than $t_{\rm BC}$) will be attenuated by both the birth cloud and the ISM in the galaxy. For older stars, the birth cloud is not taken into account, as they would have dissipated by then. $T_\lambda(\tau)$ is then given by

$$T_\lambda(t) = \begin{cases} e^{-(\tau_{\rm ISM}+\tau_{\rm BC})\Gamma_\lambda} & t < t_{\rm BC} \\ e^{-\tau_{\rm ISM}\Gamma_\lambda} & t \geq t_{\rm BC} \end{cases}, \tag{2.14}$$

where

$$\Gamma_\lambda = {\rm e}^{-az} \left(\frac{m_{\rm cold}}{10^{10} h^{-1} {\rm M}_\odot} \right)^{\gamma_{\rm GCD}} \left(\frac{r_{\rm disc}}{h^{-1}{\rm kpc}} \right)^{-2} \left(\frac{\lambda}{1600} \right)^n \tag{2.15}$$

The free parameters are a, $\gamma_{\rm GCD}$, n, $\tau_{\rm ISM}$, and $\tau_{\rm BC}$. Their values are discussed in chapter 3

2.3. Reionization within Meraxes

Meraxes is designed to self-consistently simulate the EoR. This is achieved by a reionisation implementation that is modelled after the semi-numerical reionisation code 21cmFAST (Mesinger et al., 2011; Murray et al., 2020).

2.3.1. HI reionisation

The IGM ionisation state is computed via an excursion-set formalism (Furlanetto et al., 2004). The essential idea is to compare the number of ionising photons within a sphere of radius R, centred at location $\boldsymbol{x}$ at redshift z, against the total number of baryons in

the same volume. The volume is flagged as ionised when

$$N_{\mathrm{b}*}(\boldsymbol{x}, z|R) N_{\gamma} f_{\mathrm{esc}} \geq N_{\mathrm{atom}}(\boldsymbol{x}, z|R)(1+\bar{n}_{\mathrm{rec}})(1-\bar{x}_{e}), \tag{2.16}$$

where $N_{\mathrm{b}*}(\boldsymbol{x}, z|R)$ is the number of stellar baryons in the volume, N_{γ} is the number of ionising photons per stellar baryons, f_{esc} is the escape fraction of these photons, and $N_{\mathrm{atom}}(\boldsymbol{x}, z|R)$ is the number of neutral baryons in the same sphere. The last two terms on the right-hand-side of equation 2.16 accounts for, respectively, the recombination of the ionised hydrogen and ionisation caused by X-rays.

2.3.2. *Coupling reionisation to galaxy formation and evolution*

MERAXES has been designed to explore the effect of reionisation on the galaxy population. MERAXES achieves this by temporally and spatially coupling reionisation to the formation and evolution of the galaxies. A key enabling factor in this direction is the simultaneous loading to memory of all the merger tree data at each snapshot. This *horizontal* merger tree loading mechanism allows the processing of the entire simulation volume one snapshot at a time – a crucial aspect for a realistic treatment of reionisation. For reionisation calculations, we grid the simulation volume.

As already mentioned in 2.2.1, MERAXES couples the galaxy population to reionisation via the baryon fraction modifier, f_{mod} following the 1-D simulations of Sobacchi & Mesinger (2013).

An algorithmic schema of how MERAXES achieves this is as follows:

1. MERAXES loads all the haloes in the snapshot into memory and evolves all the galaxies.

2. Cartesian grids of stellar mass and average star formation rates are generated.

3. Using the excursion-set formalism, the average UV background (UVB; $\bar{J}_{21}$) and HI grids are generated. The $\bar{J}_{21}$ field is computed as

$$\bar{J}_{21} = \frac{(1+z)^2}{4\pi} \lambda_{\mathrm{mfp}} h \alpha f_{\mathrm{bias}}\, \bar{\epsilon} \tag{2.17}$$

where λ_{mfp} is the radius of the HII bubbles, $\alpha = 5$ (Thoul & Weinberg, 1996), h

is the Planck's constant, $f_{\rm bias} = 2$ is to account for the increased clustering of the haloes hosting galaxies (Mesinger & Dijkstra, 2008) and $\bar{\epsilon}$ is the ionising emissivity.

4. In order to couple the photoionisation feedback from reionisation to the galaxy growth, the amount of freshly accreted gas into the haloes is regulated by the $f_{\rm mod}$. MERAXES constructs grids of $f_{\rm mod}$ as follows:

$$f_{\rm mod} = 2^{-M_{\rm filt}/M_{\rm vir}}, \tag{2.18}$$

where the filtering mass $M_{\rm filt}$ is calculated as

$$M_{\rm filt} = M_0 J_{21}^a \left(\frac{1+z}{10}\right)^b \left[1 - \left(\frac{1+z}{1+z_{\rm ion}}\right)^c\right]^d, \tag{2.19}$$

where $[M_0, a, b, c, d] \equiv [2.8 \times 10^9\ {\rm M}_\odot, 0.17, -2.1, 2.0, 2.5]$, and z_{ion} is the redshift at which the halo was first ionised (for further details see Sobacchi & Mesinger, 2013). The J_{21} is given by

$$J(\nu) = J_{21} \left(\frac{\nu}{3.2872 \times 10^{15}\ {\rm Hz}}\right)^{-\alpha} \times 10^{-21} {\rm ergs}^{-1}\ {\rm Hz}^{-1} (\text{ proper cm })^{-2} {\rm sr}^{-1}, \tag{2.20}$$

where α is the same as that in equation 2.17.

5. This $f_{\rm mod}$ will be applied in step 1 of the next snapshot for all the galaxies in every voxel. The steps are repeated for each snapshot of the simulation.

2.4. CALIBRATING THE SIMULATIONS

SAMs generally have a large number of free model parameters that have to be tuned to match various observations. For galaxy formation SAMs, these are usually the luminosity and stellar mass functions. In this section, I give a brief summary of these two.

2.4.1. *Luminosity functions*

A luminosity function (LF), $\phi(L)$, quantifies the number density of galaxies as a function of their luminosities (L). The number of galaxies of luminosity between L and $L + dL$ in a representative unit volume of the Universe is given by $\phi(L)dL$. Schechter (1976) proposed a convenient approximation for $\phi(L)$, which is a power law with an exponential cut-off at high luminosity as

$$\phi(L)dL = \phi^* \left(\frac{L}{L^*}\right)^{-\alpha} e^{-L/L^*} \mathrm{d}\left(\frac{L}{L^*}\right) \tag{2.21}$$

where ϕ^* is the normalisation constant and L^* is the characteristic luminosity.

Luminosity functions are modelled using SAMs, usually in post-processing, by assuming a spectral energy distribution and are computed based on the star formation histories of the galaxies. See section 2.2.5 and Qiu et al. (2019) for the specific details of the implementation within Meraxes . In the second column of figure 3.2, I show the calibrated UV LFs from Meraxes .

2.4.2. *Stellar mass functions*

Stellar mass functions (SMFs) quantify the number density of galaxies as a function of their stellar mass and represent a key measure for the properties of the galaxy population. For instance, a higher star formation efficiency will move more into the form of stars and will have an influence on almost all of the galaxies. On the other hand, a greater supernova feedback efficiency will have a more pronounced effect on the low mass end where the galaxies will not be able to hold on to the ejected gas in their (relatively) shallower potential well. SMFs trace the stellar mass assembly and the evolution of the star formation rate (SFR) in the Universe. The third column of figure 3.2 I show the calibrated SMFs from Meraxes .

2.5. Bottlenecks of SAMs

As mentioned, SAMs post-process dark matter halo merger trees. Traditionally, the merger trees are extracted from dark matter-only N-body simulations. The underly-

ing N-body simulation thus fixes the cosmological volume as well as the halo masses that can be explored by the SAMs. Hence, one of the main limitations of SAMs is the input merger trees. N-body simulations are computationally expensive, constituting a major share of the computational resources of the modelling campaigns with SAMs.

Simulations of the EoR require large volumes (necessitated by the long mean-free paths of the X-rays, for example). At the same time, for a fixed particle number, the mass resolution (i.e. the mass of the smallest halo that can be resolved as gravitationally bound object) of an N-body simulation decreases as the simulation volume increases. There is, therefore, the danger of not being able to resolve the full galaxy population in large simulations.

2.6. Contribution of this thesis to Meraxes

The calculation of the spin temperature requires the explicit modelling of the thermal state of the IGM. A primary driver of the thermal state of the IGM, and hence the spin temperature, are the X-rays, which are efficient at injecting heat into the system. Past works with Meraxes (see for example Geil et al., 2016) did not track the explicit evolution of T_S. This was, in effect, the $T_S >> T_{CMB}$ regime of equation 1.44. This is a justifiable assumption, particularly during and towards the end of the EoR. Since the sources had started ionising the IGM, the brightness temperature fluctuations trace the ionisation field. Geil et al. (2016) explored the 21-cm signal and the 21-cm statistics, particularly the 21-cm power spectra, in this limit. In this thesis, Meraxes was updated to include the impact of X-rays, thereby enabling the explicit calculation of the thermal state of the IGM.

Throughout this thesis, I make use of a dark matter-only N-body simulation of side-length 210 h^{-1} cMpc that was generated as part of the *Genesis* suite of simulations of ASTRO3D. Such large volumes are necessary for capturing the impact of the X-rays on the 21-cm signal. As mentioned, this simulation is not able to resolve the full galaxy population. A solution is to supplement the merger trees extracted from N-body simulations via Monte-Carlo methods by introducing the low-mass haloes in the post-process. In this thesis, I used DarkForest Qiu et al. (2020), a Monte-Carlo algorithm-based code that augments the N-body simulation in post-processing. This augmented simulation is

capable of resolving all atomically cooled haloes in a 210 h^{-1} Mpc simulation down from $z = 20$.

In file: ./meraxes/src/core/meraxes.c func: main line: 25

Task: 0 is exiting.

CHUCK_NORRIS_EXCEPTION: "It works on my machine" always holds true for Chuck Norris.

Running cleanup...

Freeing FOF groups and halos...

Freeing hdf5 related stuff...

– Meraxes run-time error

CHAPTER 3

Thermal and Reionisation History within a Large-Volume Semi-Analytic Galaxy Formation Simulation

This chapter is based on

S. Balu, B. Greig, Y. Qiu, Chris Power, Y. Qin, S. Mutch, J. S. B. Wyithe,

Monthly Notices of the Royal Astronomical Society, 520, 3, 2023, 3368

reformatted with the following changes only:

- **The text is styled and restructured to match the rest of this thesis.**
- **Where necessary, bibliographic records are updated.**

Abstract

We predict the 21-cm global signal and power spectra during the Epoch of Reionisation using the Meraxes semi-analytic galaxy formation and reionisation model, updated to include X-ray heating and thermal evolution of the intergalactic medium. Studying the formation and evolution of galaxies together with the reionisation of cosmic hydrogen using semi-analytic models (such as Meraxes) requires N-body simulations within large volumes and high mass resolutions. For this, we use a simulation of side-length $210h^{-1}$ Mpc with 4320^3 particles resolving dark matter haloes to masses of $5 \times 10^8 h^{-1}\mathrm{M}_\odot$. To

reach the mass resolution of atomically cooled galaxies, thought to be the dominant population contributing to reionisation, at $z = 20$ of $\sim 2 \times 10^7 h^{-1} \mathrm{M}_\odot$, we augment this simulation using the DarkForest Monte-Carlo merger tree algorithm (achieving an effective particle count of $\sim 10^{12}$). Using this augmented simulation, we explore the impact of mass resolution on the predicted reionisation history as well as the impact of X-ray heating on the 21-cm global signal and the 21-cm power spectra. We also explore the cosmic variance of 21-cm statistics within $70^3\ h^{-3}\ \mathrm{Mpc}^3$ sub-volumes. We find that the midpoint of reionisation varies by $\Delta z \sim 0.8$ and that the cosmic variance on the power spectrum is underestimated by a factor of $2-4$ at $k \sim 0.1 - 0.4\ \mathrm{Mpc}^{-1}$ due to the non-Gaussian nature of the 21-cm signal. To our knowledge, this work represents the first model of both reionisation and galaxy formation, which resolves low-mass atomically cooled galaxies while simultaneously sampling sufficiently large scales necessary for exploring the effects of X-rays in the early Universe.

3.1. Introduction

The formation of the first luminous objects during the cosmic dawn resulted in the ionisation of the cosmic HI gas, rendering the intergalactic medium (IGM) transparent to UV photons. This period, termed the Epoch of Reionisation (EoR), constitutes the last major phase change of hydrogen in the Universe and had an impact on subsequent galaxy formation and evolution (Barkana & Loeb 2001). A promising probe of this period is the 21-cm hyperfine spin-flip transition of HI, which is sensitive to the evolution of the thermal and ionisation states of the IGM (Furlanetto et al. 2006b).

A number of low-frequency radio telescope arrays are in operation or are planned to detect this signal. Current instruments (MWA*, LOFAR†, HERA‡) aim to detect the signal statistically via the 21-cm power spectrum (21-cm PS; Morales & Wyithe 2010). While a detection has not yet been made, in recent years there has been significant progress in lowering the available upper limits (Mertens et al. 2020, Trott et al. 2020, The HERA Collaboration et al. 2022). In addition, the evolution of the all-sky averaged 21-cm global signal (21-cm GS) is being sought with experiments such as EDGES (Bow-

*www.mwatelescope.org
†www.astron.nl/telescopes/lofar
‡reionization.org/

man et al. 2018) and SARAS (Singh et al. 2022). In the near future, the Square Kilometre Array (SKA; Koopmans et al. 2014)* will provide an unprecedented ability to place observational constraints on the physics of this era by enabling the production of detailed 3-D 21-cm maps showing the distribution and evolution of the cosmic HI.

For interpreting current and future observations, it is important that realistic simulations of the early Universe are available, and many authors have contributed to this effort (see Gnedin & Madau 2022 for a recent review). Simulations of the EoR are made challenging by the large range of scales involved. The main drivers controlling the ionisation and thermal states of the HI are, respectively, the intense UV and X-ray photons from star-forming galaxies (see Mesinger 2019 and references therein). X-ray photons have mean-free paths of the order of 10s – 100s of Mpc in the high-z Universe, while the typical individual HII bubble sizes are $\sim 10-15$ Mpc (Wyithe & Loeb 2004, Furlanetto et al. 2006a). It has also been shown that simulation volumes of sidelength $\gtrsim 100h^{-1}$ Mpc are needed for convergent reionisation histories (Iliev et al. 2014) while $\gtrsim 200h^{-1}$ Mpc are needed for convergent 21-cm PS (Kim et al. 2016, Kaur et al. 2020). These considerations necessitate simulations capable of resolving structures from a few Mpc in volumes of $\gtrsim$ 100s Mpc on a side.

At the same time, realistic EoR modelling requires the ability to resolve haloes down to at least the hydrogen cooling limit corresponding to a halo virial temperature of $T_{\rm vir} \sim 10^4$K and virial mass (Barkana & Loeb 2001)

$$M_{\rm vir}(z) \sim 4.4\times10^3\left(\frac{T_{\rm vir}}{1+z}\right)^{3/2} h^{-1}\ {\rm M}_\odot \tag{3.1}$$

These so-called atomically cooled haloes (see Figure 3.1) provide sites where gas efficiently cools via atomic line transitions to form stars. Thus, to realistically simulate a representative volume of the early Universe, one requires large simulation volumes as well as sufficiently high mass resolutions.

Several techniques have been developed to simulate the EoR (Gnedin & Madau 2022). Semi-numerical simulations (e.g. Santos et al. 2010, Mesinger et al. 2011, Maity & Choudhury 2022a) typically associate ionising photon sources with the density peaks of evolved Gaussian random fields. As these models do not require running computationally ex-

*www.skao.int

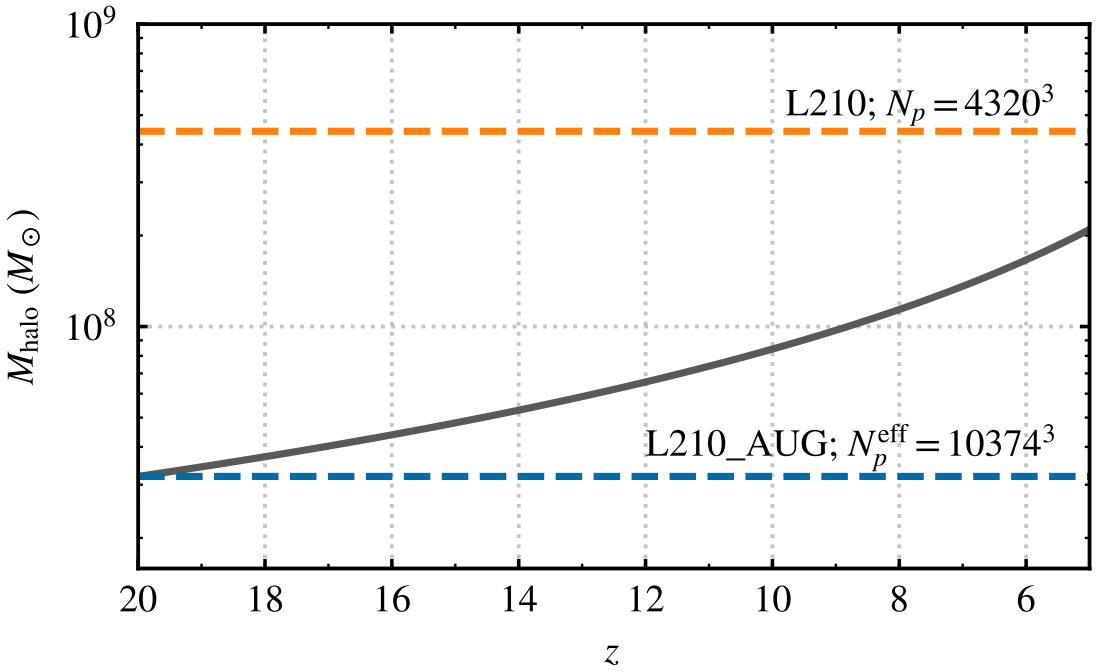

Figure 3.1: The mass of atomically cooled haloes (dark grey curve) as a function of redshift z along with the representative halo mass resolution of the L210 (orange dashed) and L210_AUG (blue dashed) simulations.

pensive N-body simulations, they are able to achieve very large volumes (Greig et al. 2022b) as well as efficiently explore the available parameter space (e.g. Greig & Mesinger 2015). Their main drawback is the absence of detailed physics, which self-consistently models a realistic galaxy population. On the other hand, achieving high resolution in large-volume hydrodynamical simulations is computationally expensive (see, for example, Gnedin 2014; Ocvirk et al. 2016; Rosdahl et al. 2018; Kannan et al. 2022). There have also been radiative transfer simulations that are uncoupled to the hydrodynamics, such as in Kulkarni et al. (2019) and the so-called hybrid models, such as by Oñorbe et al. (2019) and Puchwein et al. (2023). However, the computational overhead associated with hydrodynamical simulations precludes their use in parameter exploration.

Semi-analytic models (SAMs; see Somerville & Davé 2015 for a review) of galaxy formation (e.g. Cole et al. 2002, Benson 2012, Croton et al. 2016, Cora et al. 2018, Lagos et al. 2018) typically take merger trees from comparatively cheaper dark matter-only N-body simulations and evolve key baryonic components which describe the physical processes involved in galaxy formation, growth and evolution using simple but physically motivated prescriptions. Importantly, being based on N-body trees, the galaxies retain their association with the large-scale structure. These galaxy SAMs then provide

a realistic galaxy population at a fraction of the cost of full hydrodynamical simulations. Coupling a galaxy SAM with a semi-numerical reionisation code can provide the best of both worlds: large-volume simulations of reionisation with a self-consistent, realistic population of galaxies. In this work, we use MERAXES (Mutch et al. 2016), developed as part of the DRAGONS (Dark-ages Reionization And Galaxy formation Observables from Numerical Simulations) program, which couples a galaxy SAM model designed for galaxies in the high-z Universe during the EoR with the semi-numerical code 21CMFAST for simulating the reionisation process*. Additionally, for the first time, we implement the evolution of the neutral hydrogen gas spin temperature into MERAXES, taking into account heating by X-ray photons.

We run our updated MERAXES on a new dark matter-only N-body simulation, which has a volume of $210^3\ h^{-3}\ \mathrm{Mpc}^3$ with 4320^3 particles. This is the largest volume on which MERAXES has been deployed (previously $67.8^3\ h^{-3}\ \mathrm{Mpc}^3$; Qiu et al. 2019). To achieve sufficient mass resolution (atomic cooling limit at $z = 20$ of $\sim 2 \times 10^7 h^{-1} \mathrm{M}_\odot$) within our simulations, we use DARKFOREST – a Monte Carlo algorithm-based code introduced in Qiu et al. (2020). This provides a unique dataset modelling both individual galaxy formation and evolution during reionisation in volumes large enough for exploring the effects of X-rays on the 21-cm signal from the cosmic dawn and the EoR. Importantly, this is the first time such a large volume coupled reionisation and galaxy SAM has been performed to study the 21-cm signal into the cosmic dawn. With our large volume, we are able to explore the impact of cosmic variance across the 21-cm statistics.

The chapter is organised as follows: section 3.2 introduces the N-body simulations utilised in this work as well as its augmentation; section 3.3 provides a brief summary of the MERAXES SAM and the calibration of its input model parameters. We analyse the resultant 21-cm signal from this model in section 3.4 and explore the cosmic variance across a broad range of statistics in section 3.5, and conclude in section 3.6. In this thesis we use the best-fit parameters from the Planck Collaboration et al. (2016): $h = 0.6751$, $\Omega_\mathrm{m} = 0.3121$, $\Omega_\mathrm{b} = 0.0490$, $\Omega_\Lambda = 0.6879$, $\sigma_8 = 0.8150$, and $n_s = 0.9653$. All quantities quoted are in comoving units unless otherwise stated.

*A few other recent examples of SAMs incorporating reionisation calculations in the literature are Seiler et al. 2019, Visbal et al. 2020, Hutter et al. 2021.

3.2. N-BODY SIMULATIONS AND THEIR AUGMENTATION

In this section, we introduce the N-body simulation used in this work, as well as an outline of the augmentation pipeline.

3.2.1. L210 Simulation

We use the L210_N4320 (hereafter L210) box of the GENESIS suite of N-body simulations (Power et al. in prep). This simulation is 210 h^{-1} Mpc on a side and consists of 4320^3 dark matter particles of mass $m_{\rm p} = 9.95 \times 10^6\ h^{-1}{\rm M}_\odot$. The halo mass resolution* is $\sim 5 \times 10^8\ h^{-1}{\rm M}_\odot$ based on a minimum of 50 particles. The simulation was evolved from $z = 99$ down to $z = 5$ using the SWIFT code (Schaller et al. 2018), and the haloes were identified via friends-of-friends by the VELOCIRAPTOR halo-finder (Elahi et al. 2019a). Halo catalogues are saved over 120 snapshots evenly distributed in dynamical time between redshifts 30 and 5. The merger trees were generated using TREEFROG (Elahi et al. 2019b).

3.2.2. L210_AUG Simulation

To increase the mass resolution of the L210 simulation from $\sim 5 \times 10^8\ h^{-1}{\rm M}_\odot$ to the atomic hydrogen cooling limit at $z = 20$ ($\sim 2 \times 10^7\ h^{-1}{\rm M}_\odot$), we augment it by extending the merger trees to lower mass haloes. This is achieved using DARKFOREST (Qiu et al. 2020), a Monte-Carlo (MC) based algorithm, which we summarise below. We call this new simulation L210_AUG, which provides a unique dataset for exploring galaxy formation physics and its impact on the timing and morphology of the EoR. Figure 3.1 shows the mass resolution of both L210_AUG (orange dashed) and L210 (blue dashed) along with the atomic cooling limit (dark grey curve) for the relevant redshifts. We point out that the augmentation algorithm works backward in time (in our case, from $z = 5$).

DARKFOREST uses an updated prescription of Benson et al. (2016) for augmenting merger trees and works on what are termed "simple branches" – merger tree branches that are composed of a halo and all of its immediate progenitors. To add new haloes to

*In this work, the term *halo mass resolution* refers to the mass of the smallest gravitationally bound dark matter halo that is identified in the simulation by the halo finder.

the existing merger trees, new simple branches are generated using the algorithm outlined in Parkinson et al. (2008) which employs a conditional mass function with extra parametrisation (to take care of the differences between the analytic halo mass functions (HMFs) and the ones from N-body simulations) derived from the Extended Press Schechter theory (Bower 1991, Bond et al. 1991, Lacey & Cole 1993). Each halo is split into two (binary splits) in small internal time steps: we choose these time steps, dz_1, such that $|dz_1| << z_1/z_2$ where z_1 is the redshift of the halo and z_2 is the redshift of its immediate progenitors. This construction is repeated until we have an MC merger history. These new MC branches, by construction, have a higher mass resolution than the N-body trees. Building on the methods employed in Benson et al. (2016), the new branches are used to augment the existing N-body merger tree.

For this, we first define a mass threshold, $M_{\rm cut}$, which serves as a dynamic boundary between the N-body and MC halo populations in the final augmented merger tree, thus helping us to "average out" the differences between these two populations. If all the newly added haloes in the generated MC simple branch are less massive than $M_{\rm cut}$, then those haloes are attached to the original N-body simple branch. As a result, the augmented simple branch will have both the MC haloes for $M_{\rm halo} < M_{\rm cut}$ and the original N-body haloes for $M_{\rm halo} > M_{\rm cut}$. The final augmented merger tree with these MC branches grafted onto it will thus have both N-body as well as MC haloes with the $M_{\rm cut}$ serving as the barrier separating the N-body and MC haloes. The resultant "hybrid" merger tree will have the same mass resolution as the MC simple branches. Benson et al. (2016) used a constant value for $M_{\rm cut}$. We allow $M_{\rm cut}$ to take values $\in [M_{\rm cut}^{\rm min}, M_{\rm cut}^{\rm max}]$. For every simple branch, the augmentation starts with $M_{\rm cut} = M_{\rm cut}^{\rm min}$, but incrementally increases it if the MC simple branch is not deemed fit* to be attached to the N-body simple branch. Qiu et al. (2020) used $[M_{\rm cut}^{\rm min}, M_{\rm cut}^{\rm max}] \equiv [100m_{\rm p}, 2500m_{\rm p}]$ where $m_{\rm p}$ is the particle mass of the unaugmented simulation. Instead, we use a $M_{\rm cut}^{\rm min} = 120m_{\rm p}$ taking into account the higher resolution of the L210 simulation while $M_{\rm cut}^{\rm max}$ remains unchanged. This choice was made so that we can retain a larger fraction of the N-body haloes in the L210_AUG simulation. The mass resolution of the augmented simulation is

*This can happen for instance, if the number of MC haloes are less than that of the N-body haloes and/or if the difference in mass of the MC haloes of $M_{\rm halo} > M_{\rm cut}$ and the corresponding N-body haloes in the simple branch are larger than a precision parameter. Note that these MC haloes (with masses above $M_{\rm cut}$) do not end up in the final merger tree but are used solely as a check on the augmentation algorithm. See Qiu et al. (2020) for further details.

a free parameter of DarkForest, which we have chosen to be the atomic cooling limit at $z \sim 20$.

In the first column of Figure 3.2, we show the HMFs of the augmented (L210_AUG in blue) and unaugmented (L210 in orange) boxes at $z = 8$, 7, 6, and 5. There is a turn-over of the HMF of L210 prior to the resolution limit because of the incomplete identification of haloes by the halo-finder. This further motivates the need for augmentation to obtain all the haloes down to the desired mass limit. The HMFs of L210_AUG extend out to the desired mass resolution with the smallest haloes resolved in the augmented tree having a mass of $\sim 2 \times 10^7\ h^{-1}\ \mathrm{M}_\odot$.

Since reionisation morphology depends on the location of photon sources, it is important that the positions of the MC haloes are assigned appropriately. DarkForest determines the positions and assigns velocities to the newly added MC haloes. We apply the non-linear halo bias prescription from Ahn et al. (2015) on the input dark matter density field from L210 to generate a halo density field. This is normalised and used as a one-dimensional probability distribution from which the MC haloes are assigned their positions by random sampling. The MC haloes are placed uniformly within a voxel in such a way that the number of haloes follows a Poisson distribution. The accuracy of this random sampling method (which we assert by comparing the 2-point correlation functions between the MC and N-body haloes in the same mass ranges) has been shown to depend on the grid size. We performed a convergence test to determine the resolution providing the best performance and use 512^3 cells for our calculations. This is also partly motivated by the resulting grid size of 0.4 h^{-1} Mpc being smaller than the HII bubble sizes (Furlanetto et al. 2006a). This method is used to assign positions to every new MC that is not a progenitor of another MC halo (i.e. it has just been resolved for the first time).

The evolution of the MC haloes' position with time is based on their peculiar velocity field, $\boldsymbol{v}(\boldsymbol{x}, t)$, using the linear continuity equation as

$$\nabla \boldsymbol{v}(\boldsymbol{x}, t) = -\frac{1}{\Delta t}[\mathcal{D}(\boldsymbol{x}, t_1) - \mathcal{D}(\boldsymbol{x}, t_2)], \tag{3.2}$$

where $\mathcal{D}(\boldsymbol{x}, t) = b(M, t)\delta_{DM}(\boldsymbol{x}, t)$ is the halo density field with $b(M, t)$ the linear halo bias (Tinker et al. 2010) and $\delta_{DM}(\boldsymbol{x}, t)$ is dark-matter overdensity field, and Δt is the

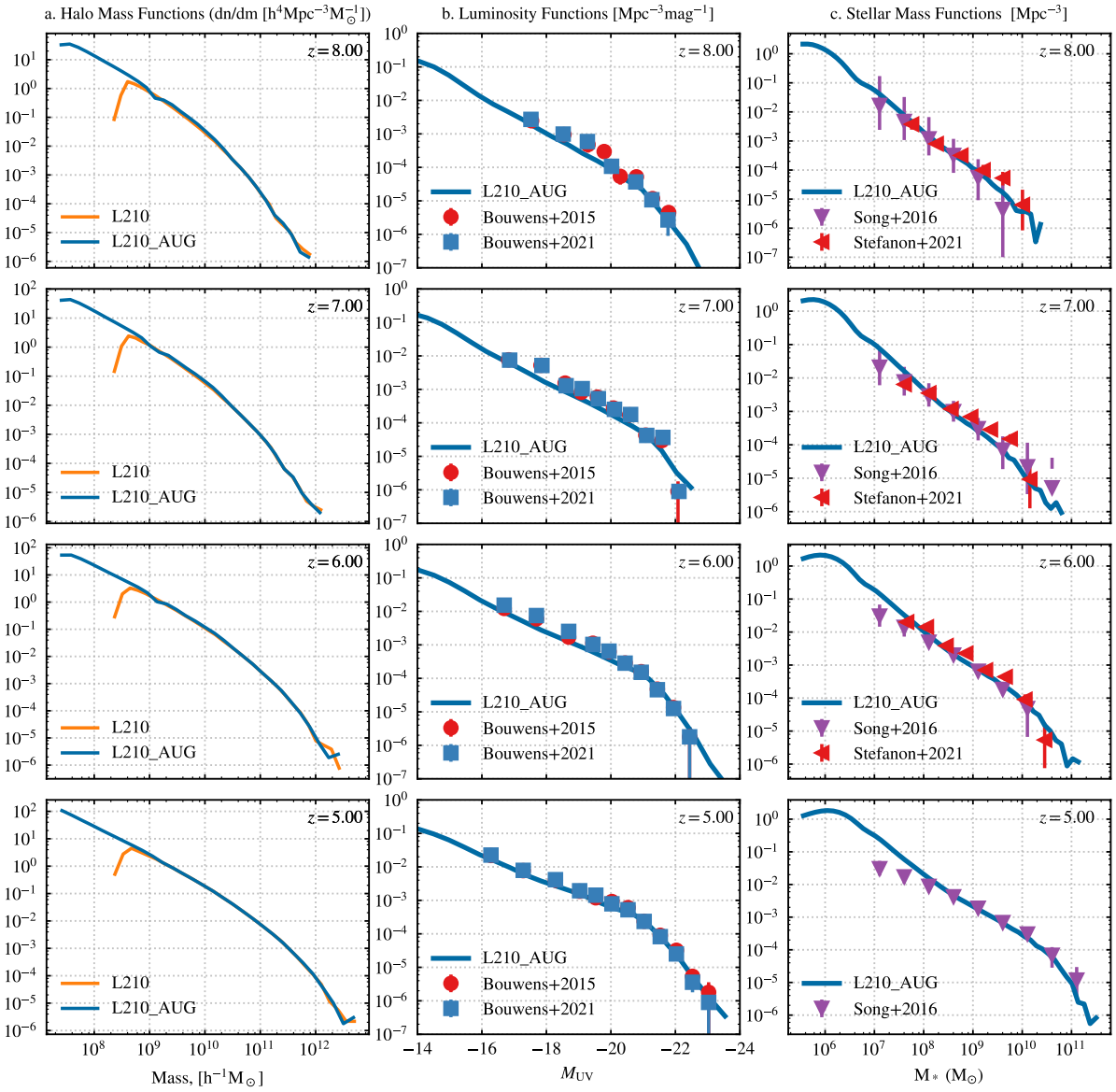

Figure 3.2: We show the HMF (first column), the LF (middle column), and SMF (right column) at $z \sim 8$, 7, 6, & 5 for both L210 (orange curve) and L210_AUG (blue curve). The first column shows the impact of the augmentation, highlighting the mass of the smallest haloes that are resolved. The final two columns demonstrate the calibration of the MERAXES using existing data. For the LF we use Bouwens et al. (2015) & Bouwens et al. (2021), and for the SMF we use Song et al. (2016) & Stefanon et al. (2021).

time-step between the snapshots. Once again, we find that the choice of grid sizes for determining the halo density fields affects the accuracy of the halo positions. Qiu et al. (2020) used an evolving (with z) grid resolution. Based on their results, we use a 256^3 grid resolution at $z = 5 - 6$, 64^3 at $z = 6 - 8$, and 32^3 grids at $z > 8$ after compensating for the differences in the simulation volumes.

As detailed in Qiu et al. (2020), we run a number of tests to ensure that the MC haloes are introduced without compromising the accuracy of the underlying L210 simulation. Specifically, while evolving the position of the MC haloes, the 2-point correlation of the halo positions and the velocity distribution of the haloes are checked to ensure they are consistent. The interested reader is referred to Qiu et al. (2020) for a detailed explanation of the augmentation algorithm.

3.3. Modelling Galaxies and the Epoch of Reionisation

3.3.1. Galaxy formation using Meraxes

SAMs enable fast and efficient modelling of galaxies and their properties within cosmological volumes. In this work, we use the Meraxes (Mutch et al. 2016) SAM, which was specifically designed to study the interplay and feedback between galaxy formation and evolution and reionisation. Since its introduction Meraxes has undergone several updates. These include AGN feedback (Qin et al. 2017) as well as updates to supernova feedback, recycling and chemical enrichment of the ISM, and reincorporation of the ejected gas (Qiu et al. 2019).

Meraxes includes detailed, physically motivated prescriptions for processes including baryonic infall into a dark matter halo, radiative cooling of this infalling gas, star formation, supernova feedback which can heat up the cold gas, mass recycling whereby the ejected material from a supernova can participate in star formation again, metal enrichment of the interstellar medium (ISM), and reincorporation of the gas that is ejected from the galaxy but still bound to the dark matter halo. The dynamical time of a typical galactic disc at high redshift is ~ 10s of Myr (which is similar to the lifetime of massive stars). Our N-body simulations have therefore been constructed with high cadence (with

a mean value of 10 Myr between $z \sim 30-5$), and MERAXES also includes time-dependent feedback based on the star formation history.

At each snapshot, the baryonic content of a dark matter halo increases up to $f_b f_{\rm mod} M_{\rm vir}$ in the form of pristine primordial gas. Here, $f_b = \Omega_{\rm b}/\Omega_{\rm m}$ is the baryon fraction of the Universe and $f_{\rm mod}$ is the baryon fraction modifier which couples the feedback of reionisation to galaxy formation. This newly acquired baryonic gas is deposited into a shock-heated and quasi-static hot-gas reservoir of the galaxy. The fraction of this hot gas, which has a cooling time less than the dynamical time of the halo, cools radiatively to a much colder gas cloud. This cold gas then participates in star formation following the Kauffmann (1996) model. The cold gas reservoir can also be depleted by the feedback from supernovae and active galactic nuclei (Qin et al. 2017).

Qiu et al. (2019) introduced a dust model into MERAXES , facilitating the computation of dust attenuated luminosity functions (LFs). The implementation is based on a dust attenuation model from Charlot & Fall (2000). Within a Bayesian framework, Qiu et al. (2019) explored three parameterisations for dust in MERAXES linked to the SFR, dust-to-gas (DTG) ratio, and gas column density (GCD). In this work, we use the DTG model, which depends on the cold gas's metallicity and mass.

To model reionisation and investigate the role of photoionisation feedback on the high-z galaxies, MERAXES includes a modified version of 21CMFAST (described in the next section; Mesinger et al. 2011, Murray et al. 2020). At each snapshot, once the galaxies are identified, and all the associated gas reservoirs are updated appropriately, MERAXES models their impact on the HI in the IGM.

3.3.2. *Reionisation in MERAXES*

Reionisation is incorporated self-consistently in MERAXES using the computationally efficient semi-numerical code 21CMFAST . Using perturbation theory, 21CMFAST generates evolved density and velocity fields, which are then converted to stellar mass and star formation rate (SFR) grids using a simple parameterisation to describe the galaxies. In MERAXES the first two fields come directly from the N-body simulations thus retaining the non-linear effects of structure formation while the stellar mass grids are computed realistically by MERAXES making use of the full galaxy properties. In this work, we extend the reionisation calculations of MERAXES to additionally follow the evolution of the

spin temperature, T_S, of HI by incorporating the heating and ionisation of the IGM by X-rays following the same approach taken within 21CMFAST. For all of the reionisation calculations, we use a grid resolution of 1024^3 corresponding to a cell resolution of $\sim 0.2\ h^{-1}$ Mpc, which is smaller than the typical size of HII regions during the EoR (Wyithe & Loeb 2004). In this section, we describe the implementation of reionisation and thermal evolution in MERAXES.

HI reionisation

The ionisation state of the IGM is determined directly from the stellar mass grids following the excursion-set formalism (Furlanetto et al. 2004). Here, the total integrated number of ionising photons is compared to the number of neutral atoms plus recombinations within spheres of radius R, centred at location $\boldsymbol{x}$ and redshift z. A simulation cell is flagged as ionised if it satisfies the condition in equation 2.16. Sobacchi & Mesinger (2014) have shown that recombinations inside Lyman limit systems can significantly reduce the sizes of HII regions. Following their implementation in 21CMFAST, through a sub-grid prescription, we account for recombinations via the $\overline{n}_{\rm rec}$ term, which is the mean number of recombinations. We decrease R from a maximum of 50 Mpc, which is the mean-free path in the IGM post-reionisation (Songaila & Cowie 2010, Becker et al. 2021), down to the size of a voxel, $R_{\rm cell}$.

The local ionisation state of the IGM is used to evaluate the value of $f_{\rm mod}$ for all the galaxies in the volume. The amount of fresh gas accreted in the next snapshot by the host haloes of the galaxies is then suppressed by a factor of $f_{\rm mod}$, thus enabling MERAXES to couple galaxy evolution with reionisation. This gives us a reionisation scenario that is self-consistent and regulated by the UV background (UVB). An exploration of the interplay between the galaxies and reionisation and its impact on the 21-cm PS (though only in the post-heating regime) with MERAXES is given in Geil et al. (2016).

Spin temperature field

The 21-cm signal depends upon the spin temperature T_S, which quantifies the population ratio of the two HI hyperfine energy levels. T_S is sensitive to the thermal state of the

IGM, which is influenced by the X-ray photons and is given by

$$T_S^{-1} = \frac{T_{\rm CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_c T_{\rm K}^{-1}}{1 + x_\alpha + x_{\rm c}}, \tag{3.3}$$

where $T_{\rm CMB}$, T_α, and $T_{\rm K}$ are the CMB, colour, and gas kinetic temperatures respectively, x_α is the Wouthuysen-Field (WF) coupling constant (Wouthuysen 1952 & Field 1958) and $x_{\rm c}$ is the collisional coupling coefficient. We take $T_\alpha = T_{\rm K}$, and x_c is computed as

$$x_c = \frac{0.0628\ {\rm K}}{A_{10} T_\gamma} \sum_i n_i \kappa^{iH}(T_{\rm K}), \tag{3.4}$$

where A_{10} is the Einstein spontaneous emission coefficient, and $i \in$ [HI, free electrons (e), free protons (p)] and the κ's refers to the corresponding collisional coefficients. We compute x_α as

$$x_\alpha = 1.7 \times 10^{11} (1+z)^{-1} S_\alpha J_\alpha, \tag{3.5}$$

where S_α is an order-of-unity correction factor involving atomic physics and J_α (pcm^{-2} s^{-1} Hz^{-1} sr^{-1} where "p" denotes proper units) is the Lyα background flux. We follow the method outlined in Mesinger et al. (2011) for computing S_α and J_α.

Following 21CMFAST, we compute the gas kinetic temperature $T_{\rm K}$ and the ionised fraction $x_{\rm e}$ at position $\boldsymbol{x}$ and redshift z as:

$$\frac{{\rm d}x_e(\boldsymbol{x}, z)}{{\rm d}z} = \frac{{\rm d}t}{{\rm d}z}[\Gamma_{ion} - \alpha_{\rm A} C x_e^2 n_b f_{\rm H}], \tag{3.6}$$

$$\frac{{\rm d}T_{\rm K}(\boldsymbol{x}, z)}{{\rm d}z} = \frac{2}{3k_{\rm B}(1+x_e)} \frac{{\rm d}t}{{\rm d}z} \sum_p \epsilon_p + \frac{2T_K}{3n_b} \frac{{\rm d}n_b}{{\rm d}z} - \frac{T_K}{1+x_e} \frac{{\rm d}x_e}{{\rm d}z}, \tag{3.7}$$

where $n_b = \overline{n}_{b,0}(1+z)^3[1+\delta_{nl}(\boldsymbol{x}, z)]$ is the total baryonic number density (H+He), $\epsilon_p(\boldsymbol{x}, z)$ is the heating rate per baryon for process p (in erg s^{-1}), $\Gamma_{\rm ion}$ is the ionisation rate per baryon, α_A is the case-A recombination coefficient, C is the clumping factor on the scale of the simulation cells ($C \equiv \langle n^2 \rangle / \langle n \rangle^2 = 2$; Sobacchi & Mesinger 2014), k_B is the Boltzmann constant, and f_H is the hydrogen number fraction. Equation 3.7 governs the thermal evolution of the gas and incorporates contributions from Compton heating (first term), adiabatic cooling and heating due to Hubble expansion and structure formation respectively (second term), and the change in internal energy of the system due to the

changing number of particles (third term).

Both $\Gamma_{\rm ion}$ and ϵ_p depend on the angle-averaged specific X-ray intensity $J(\boldsymbol{x}, E, z)$. For a voxel with location $\boldsymbol{x}$ at redshift z, the X-ray intensity at energy E, $J(\boldsymbol{x}, E, z)$, is computed by integrating the comoving X-ray specific emissivity $\epsilon_X(\boldsymbol{x}, E_e, z^{'})$ back along the light-cone as

$$J(\boldsymbol{x}, E, z) = \frac{(1+z)^3}{4\pi} \int_z^{\infty} dz' \frac{cdt}{dz^{'}} \epsilon_X e^{-\tau}, \tag{3.8}$$

where $e^{-\tau}$ accounts for the attenuation of the X-ray photons by the IGM, i.e. the probability that an X-ray photon emitted at redshift $z^{'}$ survives till z and is calculated as

$$\tau_{\rm X}(\nu, z, z') = \int_{z'}^{z} {\rm d}\hat{z} \frac{c\ {\rm d}t}{{\rm d}\hat{z}} \bar{x}_{\rm HI}(\hat{z}) \bar{n}(\hat{z}) \tilde{\sigma}(z, \hat{\nu}), \tag{3.9}$$

where $\tilde{\sigma}(z, \hat{\nu}) \equiv f_{\rm H}(1-\bar{x}_{\rm e})\sigma_{\rm H} + f_{\rm He}(1-\bar{x}_{\rm e})\sigma_{\rm HeI} + f_{\rm He}\bar{x}_{\rm e}\sigma_{\rm He}$ is the photo-ionization cross-section that is weighted over the different species (H & He) and is evaluated at $\tilde{\nu} = \nu(1+\hat{z})/(1+z)$. The integral in equation 3.9 is not trivial to evaluate due to the dependence on redshift and frequency. To make it computationally tractable, we approximate the $e^{-\tau}$ as a step function with $e^{-\tau} = 1$ for $\tau \leq 1$ and 0 for $\tau > 1$.

We relate the comoving X-ray specific emissivity $\epsilon_X(\boldsymbol{x}, E_e, z^{'})$, evaluated in the emitted frame where $E_e = E(1+z^{'})/(1+z)$, to the star formation rate density SFRD$(\boldsymbol{x}, E_e, z^{'})$ in the voxel

$$\epsilon_X(\boldsymbol{x}, E_e, z') = \frac{L'_X}{\rm SFR} \times {\rm SFRD}(\boldsymbol{x}, E_e, z'), \tag{3.10}$$

where $L'_X/{\rm SFR}$ is the specific X-ray luminosity per unit star formation that is capable of escaping the galaxy in units of [erg s^{-1}M$_{\odot}^{-1}$yr].

Unlike 21cmFAST, where the SFRD is calculated from the density field and collapse fraction, we compute the SFRD directly from Meraxes, making use of our galaxies' properties. Feedback from reionisation can thus alter the star formation rates of galaxies. $L_X/{\rm SFR}$ is assumed to follow a power-law of the form

$$L'_X/{\rm SFR} \propto E^{-\alpha_X}, \tag{3.11}$$

where E is the photon energy and is normalised with respect to the soft-band X-ray

luminosity as

$$L_{X<2\,\mathrm{keV}}/\mathrm{SFR} = \int_{E_0}^{2\,\mathrm{keV}} dE_e\; L'_X/\mathrm{SFR}. \tag{3.12}$$

Here, E_0 is a threshold energy that fixes the lowest energy of X-ray photon capable of escaping the galaxy by producing a sharp cut-off in the X-ray spectrum that accounts for where the X-rays are absorbed by the high column density gas within the galaxy.

We thus have three free parameters characterising the X-ray properties of the galaxies: $L_{X<2\,\mathrm{keV}}/\mathrm{SFR}$, E_0, and α_X. In this work we only vary $L_{X<2\,\mathrm{keV}}/\mathrm{SFR}$ keeping the other two fixed*. We set $\alpha_X = 1$ consistent with the observations of high-mass X-ray binaries in the local Universe (Mineo et al. 2012, Fragos et al. 2013, Pacucci et al. 2014). Motivated by Das et al. (2017) we adopt a value of $E_0 = 0.5$ keV throughout this work. The physical interpretation of the upper limit of 2 keV in the integral of equation (3.12) is that X-ray photons with higher energies have mean-free paths longer than the Hubble length and thus do not interact with the IGM.

Brightness temperature field

The 21-cm brightness temperature field is given by (Furlanetto et al. 2006b):

$$\begin{aligned}\delta T_\mathrm{b}(\nu) &= \frac{T_\mathrm{S} - T_\gamma}{1+z}(1 - e^{-\tau_{\nu_0}}) \\ &\approx 27 x_{\mathrm{H\,I}}(1+\delta_{nl})\left(\frac{H}{dv_r/dr + H}\right)\left(1 - \frac{T_\gamma}{T_\mathrm{S}}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_M h^2}\right)\left(\frac{\Omega_\mathrm{b} h^2}{0.023}\right)\,\mathrm{mK},\end{aligned} \tag{3.13}$$

where T_γ is the background radiation (usually assumed to be the CMB) impinging upon the HI cloud, τ_{ν_0} is the optical depth at the 21-cm transition frequency ν_0, $1+\delta_{nl}$ is the density contrast in the dark matter field ($\delta_{nl} = \rho/\bar{\rho} - 1$), $H(z)$ is the Hubble parameter at the redshift z, and dv_r/dr is the radial derivative of the line-of-sight component of the peculiar velocity.

Below $z \sim 25$ we have three broad periods reflected in the 21-cm signal (Pritchard & Loeb 2012):

1. WF coupling (Lyα pumping): The radiation from the first stars and galaxies begins

*See Greig & Mesinger (2017) for an exploration of E_0 and α_X. The reader is also cautioned that a direct comparison with this work is not straightforward since they do not have a realistic galaxy population.

to couple T_S to T_K via the WF effect. This drives the global signal ($\overline{\delta T_b}$) into the absorption regime.

2. X-ray heating: During the Epoch of Heating (EoH), the IGM is heated by X-rays. The T_S, which is still tightly coupled to the T_K, increases above T_{CMB} and the 21-cm GS shows an emission feature. The 21-cm signal also becomes insensitive to the spin temperature ($T_S >> T_\gamma$ in equation (1.44)).

3. Reionisation: As reionisation proceeds, the 21-cm signal goes to zero.

All of these epochs are reflected in the 21-cm GS and 21-cm PS. X-rays can have a significant impact on the timing and extent of these periods, most notably the EoH.

As is evident from equation (3.13), the 21-cm signal depends on the ionisation, density, velocity, and spin temperature fields. We compute the HI 21-cm signal from the EoR by efficiently computing 3-D grids of 21-cm T_S and ionisation fields while the velocity and density fields are sourced from the N-body simulation.

Most studies in the literature that include galaxy formation focus on the post-heating regime with the simplification that $T_S >> T_\gamma$ (Geil et al. 2016; also see Greig & Mesinger 2017 for a detailed analysis of the impact of this assumption). While likely valid during the late stages of the EoR, when the luminous sources have managed to couple the spin temperature to the kinetic temperature, for observations into the Dark Ages and the EoH, this assumption breaks down. The main drivers of heating of the cosmic HIare X-ray photons (Furlanetto et al. 2006a, McQuinn 2012). Large-scale simulations with low mass resolution are unable to simulate the effects of X-rays since the build-up of the stellar mass is delayed (as we demonstrate in section 3.4). We use MERAXES combined with our augmented N-body simulations for calculations of the full brightness temperature field, including contributions from heating, the spin temperature, recombinations, and peculiar velocities.

3.3.3. Calibration

Due to the numerous physical processes involved in galaxy formation and evolution, SAMs generally contain a large number of free parameters. In order to determine values for these parameters to ensure that a realistic galaxy population is produced requires

Parameter	Value	Description
$\alpha_{\rm SF}$	0.10	Star formation efficiency
$\Sigma_{\rm SF}$	0.01	Critical mass normalisation
η_0	7.0	Mass loading normalisation
ϵ_0	1.5	Supernova energy coupling normalisation
$\gamma_{\rm DTG}$	0.65	Galaxy metallicity scaling of optical depth
$f_{\rm esc,0}$	0.14	Escape fraction normalisation
$\alpha_{\rm esc}$	0.2	Escape fraction redshift scaling

Table 3.1: The fiducial input parameters and their values used for the simulations listed in Table 3.2. The first set of these ($\alpha_{\rm SF}$, $\sum_{\rm SF}$, η_0, ϵ_0, & $\gamma_{\rm DTG}$) are calibrated to the observed LFs and SMFs and control the galaxy properties of Meraxes, while $f_{\rm esc}$ & $\alpha_{\rm esc}$ is calibrated with respect to the reionisation constraints. See section 3.3.3 for more details.

calibrating the model against a number of existing observations. There are two different sets of calibrations involved in Meraxes – one for the different galaxy formation parameters and the other for the reionisation calculations*. We calibrate the L210_AUG simulation by varying the following subset of input parameters of Meraxes: $\alpha_{\rm SF}$, $\sum_{\rm SF}$, η_0, ϵ_0, $\gamma_{\rm DTG}$, $f_{\rm esc}$, & $\alpha_{\rm esc}$ (see Table 3.1 for details). We use the same parameter values for all of our simulations (listed in Table 3.2).

Qiu et al. (2019) calibrated the galaxy formation parameters of Meraxes against observed luminosity functions (LFs) and colour-magnitude relations at $z \sim 4-7$. In this work, we calibrate our simulations with respect to the LFs and the stellar mass functions (SMFs) in the $z \sim 5-8$ range. We find that, except for the $\gamma_{\rm DTG}$, the galaxy parameters from Qiu et al. (2019) give a good fit to the data. Figure 3.2.b shows the dust attenuated luminosity functions for redshifts 8, 7, 6, & 5 along with the observational data points (Bouwens et al. 2015 and Bouwens et al. 2021). Figure 3.2.c shows the stellar mass functions for the corresponding redshifts with observations from Song et al. (2016) and Stefanon et al. (2021). We had to re-calibrate the $\gamma_{\rm DTG}$ (0.65 instead of 1.20; see Table 2 of Qiu et al. 2019) parameter which governs the manner in which dust optical depth scales with the cold gas metallicity of the galaxy. The reason is that we extend our calibrations to brighter regions of the LFs than were available to Qiu et al. (2019) because of their smaller simulation size. We summarise the parameters of Meraxes along with

*Reionisation feedback affects low-mass galaxy formation but not properties constrained by observed LFs (e.g. Mutch et al. 2016).

Name	Mass resolution $[h^{-1}\mathrm{M}_\odot]$	$L_{X<2\ \mathrm{keV}}$/SFR $[\mathrm{erg\ s^{-1}\ M_\odot^{-1}\ yr}]$	Comments
L210	2.98×10^8	3.16×10^{40}	Fiducial simulation
L210_AUG	2.12×10^7	3.16×10^{40}	Augmented fiducial simulation.
L210_AUG_LOWX	2.12×10^7	3.16×10^{38}	Same as L210_AUG but with 1/100th of the galaxy X-ray luminosity
L210_AUG_HIGHX	2.12×10^7	3.16×10^{42}	Same as L210_AUG but with 100× of the galaxy X-ray luminosity
L210_NR	2.98×10^8	3.16×10^{40}	L210 without recombinations
L210_AUG_NR	2.12×10^7	3.16×10^{40}	L210_AUG without recombinations

Table 3.2: The names, halo mass resolution, X-ray luminosity and a brief description of all the simulations used in this work.

their values, which have been used for calibration in Table 3.1.

The second set of calibrations is for the reionisation calculations. The photon budget is influenced by the escape fraction ($f_{\rm esc}$) of the galaxies, which sets the fraction of photons that are able to survive the absorption by dust and neutral gas in and around the galaxies and escape into the IGM. The high-z escape fraction is one of the least constrained parameters in the literature. In this work, we use a prescription that is skewed towards the high redshifts as the shallower potential of the small galaxies at high-z results in more photons escaping their hosts. Additionally, the impact of the Monte Carlo haloes is more relevant at high-z, and this implementation helps to bring out the importance of these galaxies. In this work use an $f_{\rm esc}$ that evolves with redshift z as:

$$f_{\rm esc} = f_{\rm esc,0}\left(\frac{1+z}{6}\right)^{\alpha_{\rm esc}}, \tag{3.14}$$

where $f_{\rm esc}^{o}$ is the escape fraction normalisation and $\alpha_{\rm esc}$ sets the escape fraction redshift scaling. We tune these parameters such that our reionisation history matches the measured constraints on the IGM neutral fractions (Figure 3.3) and the integrated optical depth of CMB photons (τ_e) from scattering off free electrons (Figure 3.4; Planck Collaboration et al. 2020).

3.4. Reionisation Predictions

In this section, we demonstrate the full reionisation model from our L210_AUG simulation. In particular, we focus on comparing the impact of the missing low-mass haloes from the L210 simulation to illustrate the importance of mass resolution.*

Since we are missing smaller mass haloes below the mass resolution in L210 we delay the build-up of cosmic stellar mass within galaxies. Thus, there is also a delay in any physical property that is dependent on the total stellar mass (such as ionisations and radiation backgrounds). This will result in the X-ray background forming too late, and

*The reader is cautioned that in the current implementation of Meraxes we are only forming and evolving atomically cooled galaxies and are thus missing the possible contribution from smaller galaxies in molecularly cooled haloes (so-called mini-haloes) which are likely to contain PopIII stars. These mini-haloes can contribute to the build-up of the background radiation fields and will have an impact on reionisation (see, for example, Qin et al. 2020, 2021a; Ventura et al. 2023, 2024). Thus, discussions on the appearance of features in this work will also be delayed relative to simulations, which also include mini-haloes.

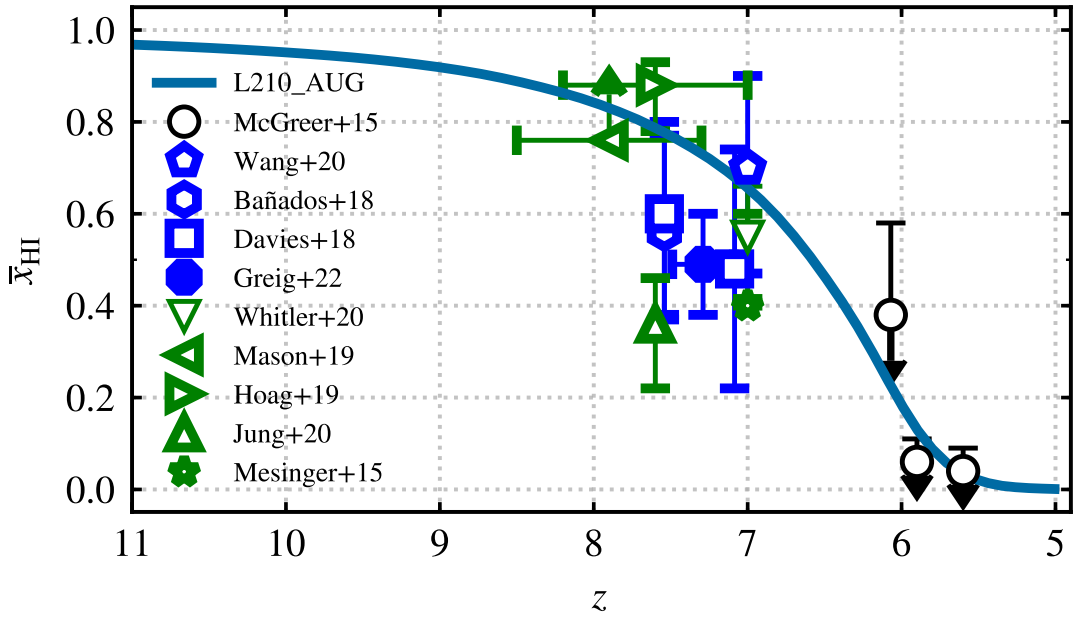

Figure 3.3: Constraints on the reionisation history of the L210_AUG simulation (blue curve). We use an evolving redshift-dependent escape fraction prescription for constraining the EoR history. The observational data are from analyses of dark pixels of Lyα & Lyβ forests (McGreer et al. 2015), and Lyα damping wing absorptions (Mesinger et al. 2015 Bañados et al. 2018, Davies et al. 2018, Mason et al. 2019, Hoag et al. 2019, Whitler et al. 2020, Jung et al. 2020, Wang et al. 2020, Wold et al. 2022, Greig et al. 2022a).

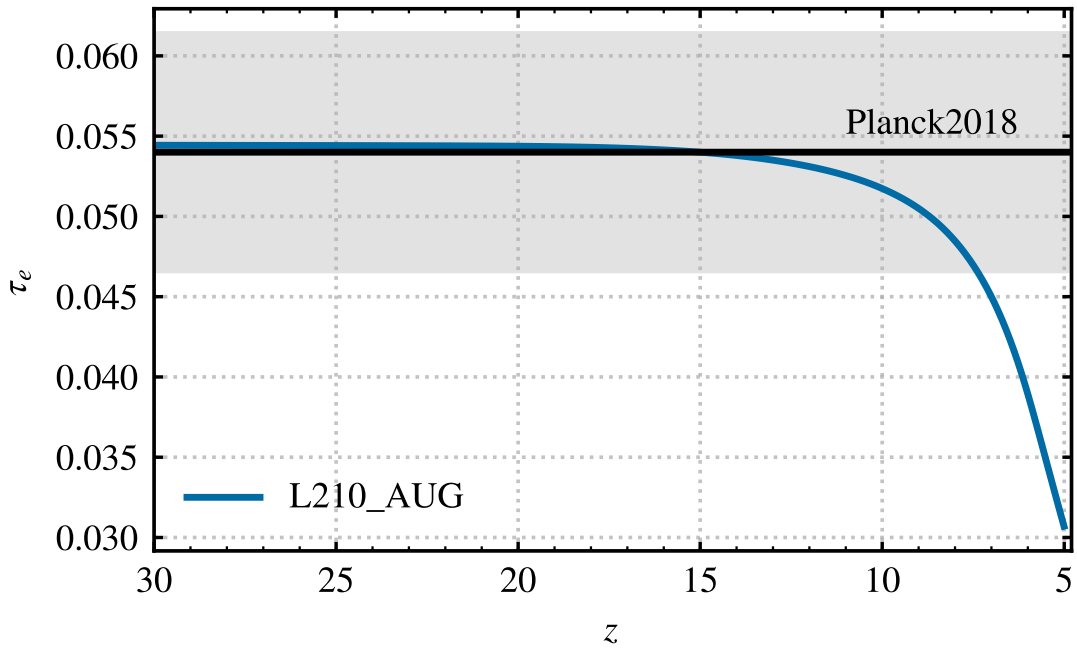

Figure 3.4: Figure shows the optical depth to CMB photons due to the free electrons. The blue curve is the integrated optical depth (τ_e) computed from the fiducial L210_AUG simulation. The black curve and the shaded region show the most recent measurement of τ_e from the *Planck 2018* collaboration (Planck Collaboration et al. 2020).

correspondingly, the IGM cools for longer before heating. However, the L210_AUG simulation includes a realistic galaxy population capable of producing the whole X-ray background. We, therefore, present here the first large-scale ($> 200h^{-1}$ Mpc) simulation of the thermal and ionisation history of the cosmic HI incorporating realistic galaxy formation and evolution physics at masses down to the atomic cooling limit.

As the first application of this simulation, we explore the evolution of the volume-averaged neutral fraction $x_{\rm HI}$, the 21-cm GS and the 21-cm PS in this section.

3.4.1. EoR history and ionisation morphology

Figure 3.5 shows the resultant EoR histories. As shown in Figure 3.3, by construction, these resultant reionisation histories are consistent with all existing limits and constraints on the IGM neutral fraction during the reionisation epoch. The L210_AUG box starts to reionise much earlier than L210 owing to the introduction of low-mass galaxies found only in Monte-Carlo haloes. However, the fiducial L210_AUG and L210 simulations both finish reionisation at approximately the same redshift. There are two main reasons

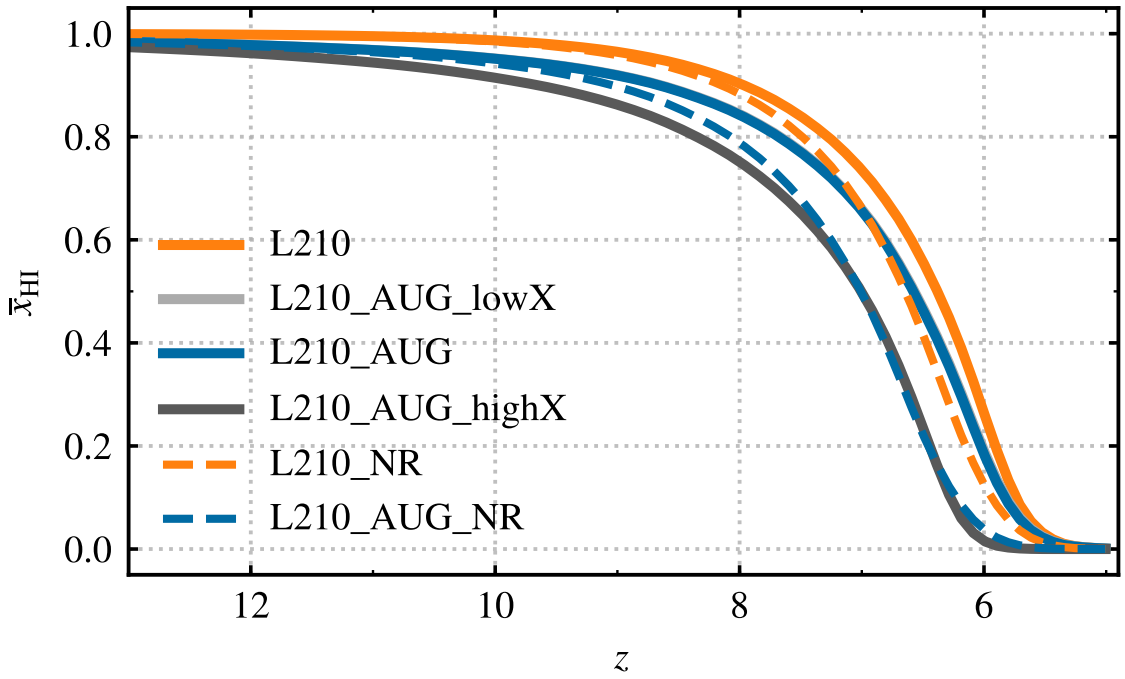

Figure 3.5: Global reionisation history of the L210_AUG (solid blue) and L210 (solid orange) simulations. L210_AUG starts reionising earlier and also has a much more extended reionisation phase. We also vary the galaxy X-ray luminosity in our model (L210_AUG_HIGHX & L210_AUG_LOWX in dark and light grey, respectively). The L210_AUG_LOWX is almost identical to L210_AUG, and thus, the curves overlap. The two dotted curves are the same as L210_AUG & L210, except that we do not include inhomogeneous recombinations.

for this. Firstly, towards the end of the EoR, reionisation is primarily maintained by larger mass haloes (which are accurately simulated across both simulations) while the lower mass galaxies are more relevant at earlier times. Secondly, the impact of inhomogeneous recombination in the two simulations is different. Since small galaxies, which initiate reionisation in L210_AUG are short-lived, the cosmic gas recombines until sufficiently big galaxies have had time to form and produce enough ionising photons to complete reionisation. In order to check the role of recombination, we, therefore, ran two additional simulations L210_AUG_NR and L210_NR (shown with dotted lines) where we have turned off inhomogeneous recombinations (i.e. setting $\overline{n}_{\rm rec} = 0$ in equation 2.16). Turning off inhomogeneous recombination results in L210_AUG_NR reionising much earlier than L210_NR as expected.

Figure 3.6 shows slices from the ionisation fields of L210_AUG (top row) and L210 (middle row) for $x_{\rm HI} \sim 0.75$, 0.50, & 0.25. Each slice is $210h^{-1}$ Mpc on a side and $\sim 0.2h^{-1}$ Mpc thick. In order to emphasise the effects of the introduction of the small haloes, we compare the simulations at the same volume-averaged neutral fractions ($x_{\rm HI}$). At any given $x_{\rm HI}$, there will be more small haloes in L210_AUG compared to L210. One of the main impacts of the smaller haloes is to force reionisation to begin earlier, as discussed above. Thus, on average, the large ionised regions in L210_AUG are smaller as they are driven by lower stellar masses relative to the L210 simulation, as indicated by the earlier redshift for a fixed neutral fraction. The third row shows the difference between the two simulations with the colour gradient chosen to highlight the contribution from the extra small mass haloes. In the "Difference" (bottom row), we have subtracted the ionisation fields of L210_AUG from L210 to clearly bring out the impact of the augmentation.

3.4.2. 21-cm statistics

Figure 3.8 shows the 21-cm GS for L210_AUG and L210. We find that the L210_AUG simulation has a similar (but broader) absorption feature, though occurring earlier in redshift, relative to L210. This highlights the importance of introducing the low-mass haloes beyond the resolution limit of L210. By including these in L210_AUG, the Lyα and X-ray backgrounds build up at earlier times due to the additional low-mass haloes. The Lyα background couples the spin temperature to the gas kinetic temperature $T_{\rm K}$, which is much lower than the CMB temperature $T_{\rm CMB}$, resulting in the broader and

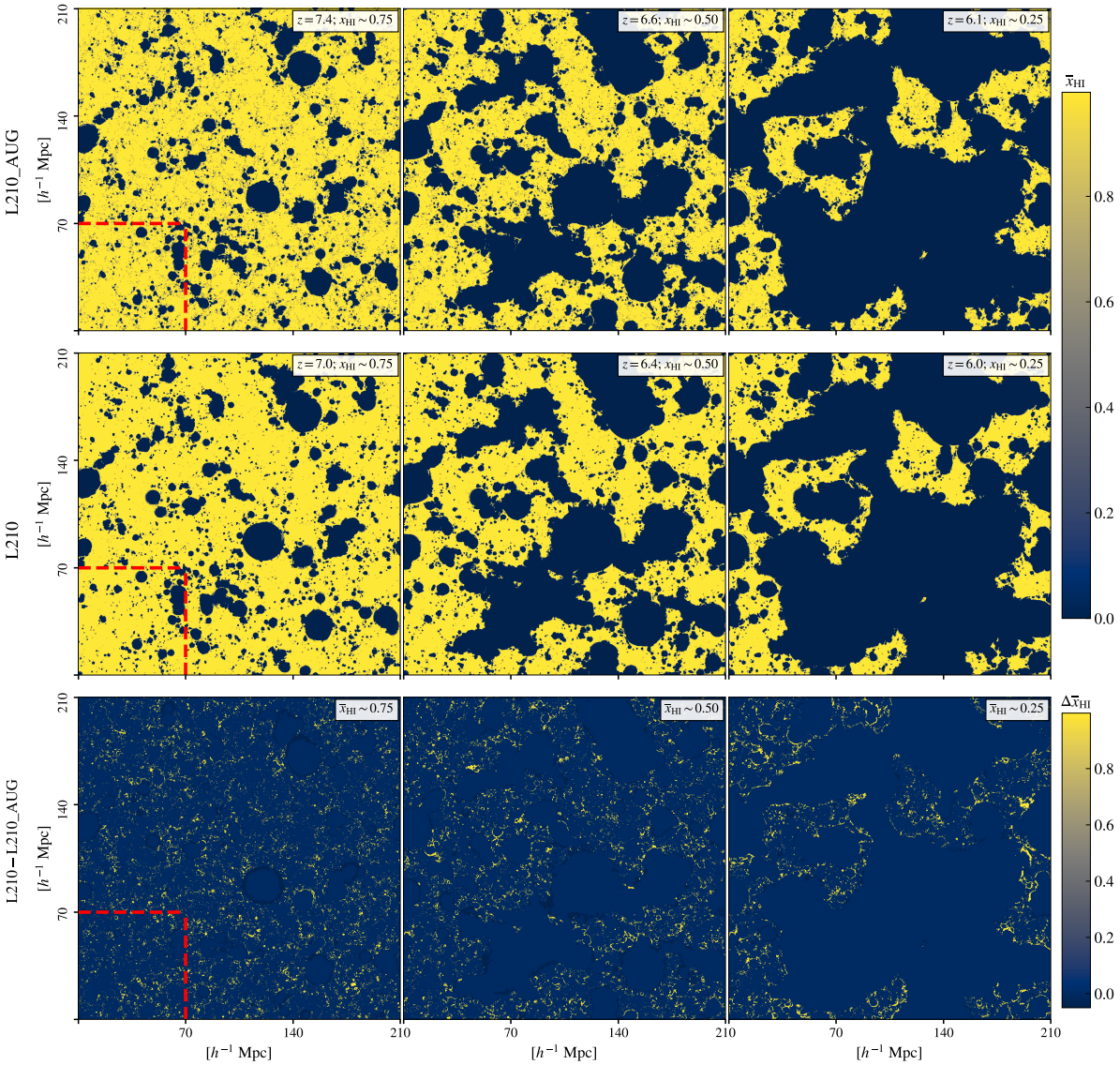

Figure 3.6: 2D slices of the $x_{\rm HI}$ grid from the L210_AUG (upper row) and L210 (bottom row) simulations. From left to right, the columns correspond to fixed neutral fractions, $x_{\rm HI} \sim$ 0.75, 0.5, 0.25. For a particular $x_{\rm HI}$, L210_AUG has a higher redshift compared to L210 because of the higher number of ionising photons. Yellow represents neutral hydrogen (HI), and blue regions are ionised hydrogen (HII) bubbles. Each slice is 210 h^{-1} Mpc on a side and ~ 0.2 h^{-1} Mpc thick. The last row shows the difference between the ionisation fields of the two simulations ($\Delta x_{\rm HI} = \overline{x}_{\rm HI,L210} - \overline{x}_{\rm HI,L210\_AUG}$). We have used a colour gradient that is weighted towards the small-scale structures to highlight the small HII regions that are due to the Monte-Carlo haloes. The red dashed regions in the first column show the size of our 70 h^{-1} Mpc side-length subvolumes (see section 3.5 for further details).

earlier absorption. We also point out that the gradient of the absorption feature in the 21-cm GS is larger in L210 as compared to L210_AUG. The delayed but sudden formation of sources in L210, relative to L210_AUG, results in a comparatively rapid build-up of the stellar mass and, consequently, the radiation backgrounds.

Figure 3.9 compares the evolution of the spherically-averaged 3-D 21-cm PS at fixed spatial scales ($k \sim 0.1$ Mpc^{-1} and $k \sim 1$ Mpc^{-1}) for the L210 and L210_AUG simulations. For both of these scales, the evolution of the power for L210 is qualitatively similar to L210_AUG. However, the delayed formation of stellar mass in L210 results in there being considerable differences between the timing of the peaks. The Lyα-coupling peak in L210 is delayed relative to L210_AUG by $\Delta z \sim 3$. Below $z \sim 7$, when the X-rays have already initiated the EoH and EoR is well on its way, the power in both L210_AUG and L210 becomes similar.

Even though the large-scale ($k \sim 0.1$ Mpc^{-1}) 21-cm power is expected to have three peaks corresponding to the Lyα-pumping, X-ray heating, and reionisation epochs (see Pritchard & Furlanetto 2007, Mesinger et al. 2013) we observe only two peaks in our simulations. The EoH peak, expected at $z \sim 12$ for L210_AUG (corresponding to the minima of the global signal; see Figure 3.8) is masked by the Lyα peak. They have merged together into one broad peak owing to the timing and build-up of the backgrounds during these two epochs*.

On the other hand, the redshift evolution of the 21-cm power on small-scales ($k \sim 1$ Mpc^{-1}) is characterised by two peaks corresponding to a combined Lyα-pumping and EoH, and an EoR peak (Qin et al. 2020). Typically, on small scales, the impact of the EoH is harder to disentangle as it primarily impacts larger scales due to the larger mean free path of X-ray photons.

3.4.3. *Effects of varying the Galaxy X-Ray Luminosity in the early Universe*

The large volume of our simulations enables the exploration of the effects of X-rays on the EoR morphology with a full source population. Here, we only vary $L_{X<2\,\mathrm{keV}}$/SFR, keeping E_0 and α_X fixed. In addition to our fiducial value of 3.16×10^{40} erg s^{-1}M$_\odot^{-1}$yr (Madau & Fragos, 2017), we also consider 3.16×10^{38} and 3.16×10^{42} erg s^{-1}M$_\odot^{-1}$yr

*Another contributing factor is the slow build-up of the backgrounds owing to the time-scales of ionising photons (as evidenced by the different gradients of the global signals).

which we call L210_AUG_LOWX and L210_AUG_HIGHX simulations respectively (see table 3.2). This enables us to encompass the plausible range of contribution of the X-rays in the early Universe (Fialkov et al. 2017, Greig & Mesinger 2017). Figure 3.7 shows the light cone evolution (of δT_b) for these simulations. We note that The HERA Collaboration et al. 2022 has recently ruled out a number of "cold reionisation" models corresponding to low X-ray luminosity. In light of this, our L210_AUG_LOWX model is a very unlikely scenario. However, our aim in this work is to develop an intuition for the impact of X-rays in the early Universe from a galaxy SAM and reionisation simulation.

In the next subsections, we compare the impact of the X-ray photons in the early Universe relative to our fiducial model.

EoR history

The ionisation photon budget is dominated by UV photons, with the X-rays contributing at most $10-15$ per cent in the most extreme models (Mesinger et al. 2013). Figure 3.5 shows the reionisation histories from all three of the augmented simulations. We find that, in agreement with studies in the literature (Mesinger et al. 2013), though the role of X-rays in ionising the HI is much less than that of the UVB, they can hasten reionisation (see L210_AUG vs L210_AUG_HIGHX in Figure 3.5).

21-cm Light Cone

In Figure 3.7 we show the 21-cm brightness temperature (δT_b) light cone slices from our simulations. The 21-cm signal, being a line transition, evolves along the line-of-sight and light cones provide a realistic representation of the evolution of such cosmic signals. We stitch together the δT_b grids from our coeval simulation boxes to generate the light cone by linearly interpolating them in cosmic time between snapshots. The delayed but rapid evolution in the case of L210, compared to the rest of the simulations, underscores the impact of the mass resolution on the 21-cm signal. For L210_AUG_LOWX, the signal remains in absorption across our full redshift range whereas for L210_AUG_HIGHX it is mostly in emission.

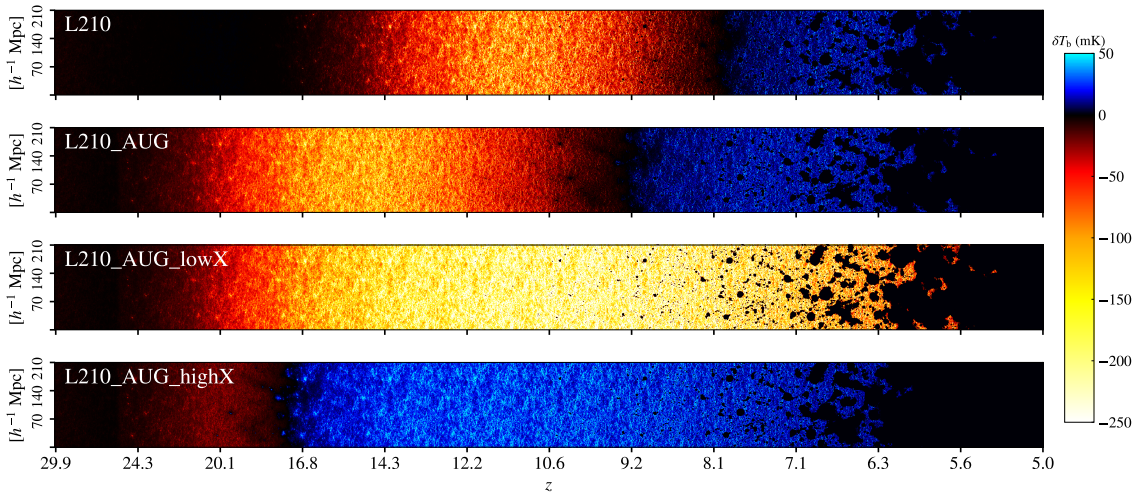

Figure 3.7: The light cone evolution of the 21-cm brightness temperature (δT_b) from our simulations. L210 is characterised by the delayed but rapid evolution of δT_b because of its lower resolution. We also point out L210_AUG_LOWX, characterised by low galaxy X-ray luminosity, in which the cosmic HI remains cold and never goes into emission.

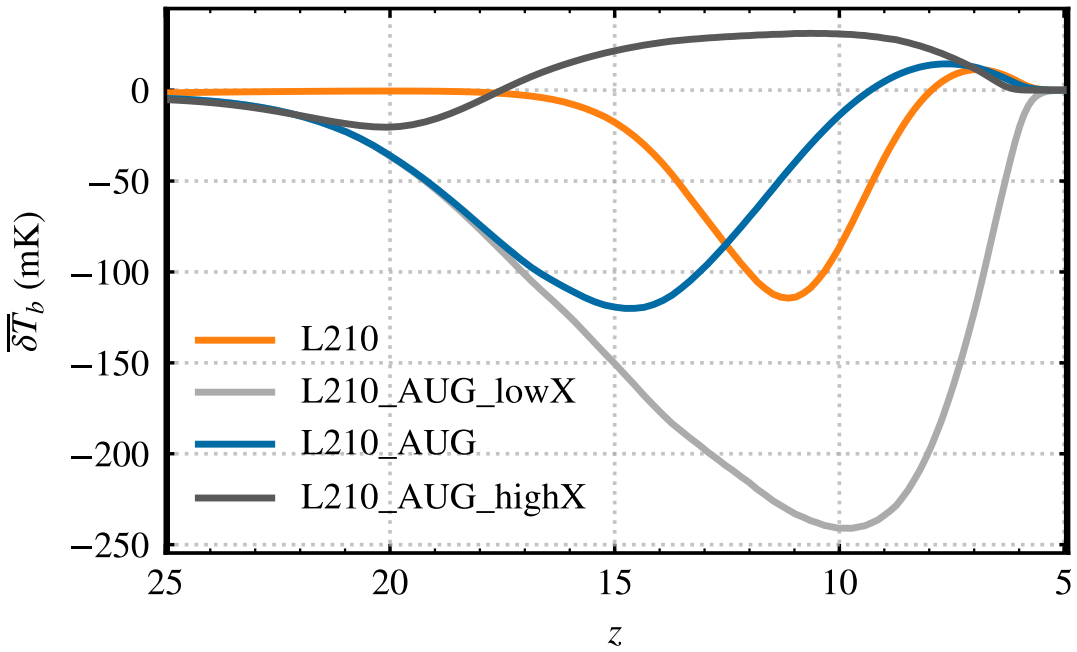

Figure 3.8: Figure shows the effect of X-rays on the 21-cm GS from the cosmic dawn and EoR. As expected, more X-rays (dark grey) cause the signal to be observed in emission earlier whereas a lack of X-rays (light grey) causes a deeper absorption feature.

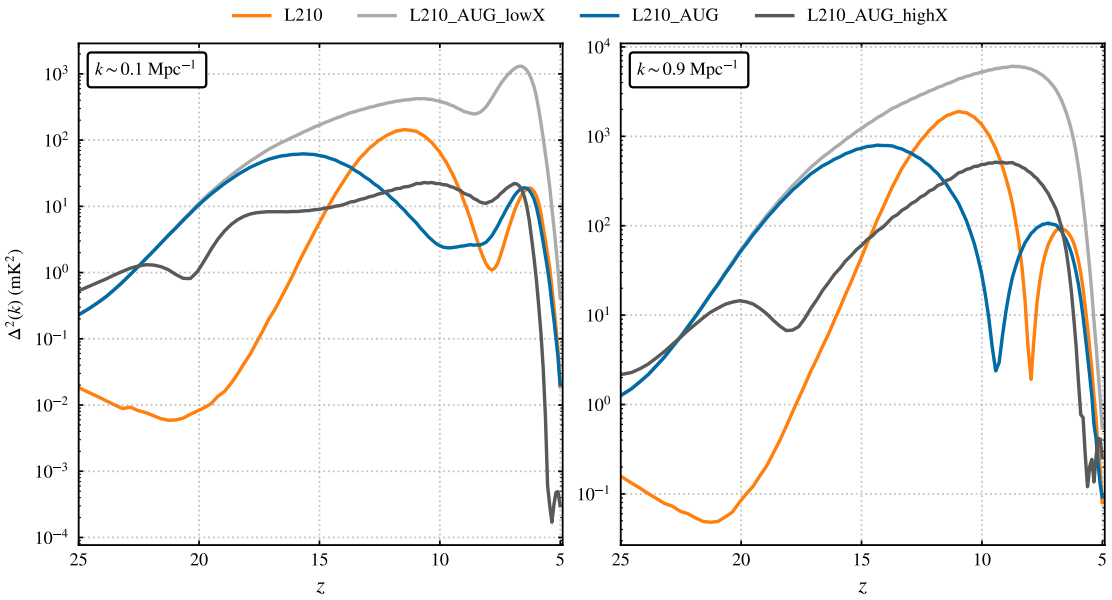

Figure 3.9: We show the 21-cm power spectra for the simulations at two scales ($k \sim 0.1$ Mpc^{-1} on left and $k \sim 1$ Mpc^{-1} on the right).

21-cm Global Signal

Figure 3.8 shows the 21-cm GS from all four simulations. As shown in section 3.4.3, the main physical impact of X-ray photons is to heat the cosmic gas. With respect to our fiducial L210_AUG simulation, L210_AUG_LOWX simulation has less X-ray photons resulting in the cosmic gas being colder, and hence a stronger absorption dip. All of our simulations except L210_AUG_LOWX go into emission (also evident from Figure 3.7). L210_AUG_HIGHX, being characterised by significantly more X-ray photons than the other simulations, has a hot IGM resulting in the signal going into emission at $z \lesssim 17$ as well as the merging of the Lyα-pumping epoch with the EoH.

21-cm Power Spectra

In Figure 3.9 we compare the evolution of the 21-cm PS at $k \sim 0.1$ Mpc^{-1} and $k \sim 1$ Mpc^{-1}. Like the 21-cm GS, the shape and amplitude of the 21-cm PS are strongly affected by the X-ray luminosity. An accurate measurement of the 21-cm PS will thus have great constraining power on the properties of the X-ray sources in the early Universe (see, for example, the recent constraints from The HERA Collaboration et al. 2022).

At large scales (left panel of Figure 3.9) we observe the expected features, though there is considerable variation in the timing and duration among the simulations. We

note that the L210_AUG_lowX simulation has the highest power for most epochs (with peak power during the EoR at $z \sim 7$) owing to the large temperature contrasts due to the cold IGM. The inefficient heating because of the low X-ray luminosity has also resulted in the EoH and EoR peaks merging together. L210_AUG simulation shows the 3 expected peaks with the power peaking at $z \sim 16$ corresponding to the absorption in the 21-cm GS; there is thus more power during the Lyα-pumping epoch. L210_AUG_highX is characterised by less power during all epochs. Though the general features are similar to the L210_AUG simulation, the amplitude and timing are different due to the reduction in the amplitude of the IGM temperature contrast. The L210_AUG_highX simulation is characterised by much smaller temperature fluctuations than the other two simulations. Thus, during the EoH, this simulation has the smallest amplitude.

The right panel shows the power on small scales. The power on this scale exhibits the expected behaviour. Interestingly, L210_AUG_highX has clearly differentiated Lyα-pumping and EoH peaks. The EoH peak, in this case, has merged with the EoR peak, owing to the extended EoH because of the large X-ray luminosity.

3.5. Cosmic Variance in EoR Statistics

Measurement of any statistical signal from a finite volume of the Universe introduces an inherent uncertainty in its variance since we are only sampling one realisation of the underlying statistical ensemble. This is termed the *cosmic variance*[*]. In this section, as an application of our large volume simulations, we explore the cosmic variance of the 21-cm signal.

To explore the cosmic variance, we divide each of the augmented simulations into 27 equal sub-volumes, each of side 70 h^{-1} Mpc. Each sub-volume is larger than the typical largest ionised regions, even during the late stages of reionisation. The 70 h^{-1} Mpc sub-volumes are also comparable to most state-of-the-art radiation hydrodynamical simulations in the literature (Kaurov & Gnedin 2015, Feng et al. 2016, Springel et al. 2018, Ocvirk et al. 2020, Kannan et al. 2022) and also to the largest simulation volume on which Meraxes has been run (Qiu et al. 2019) as part of the DRAGONS project. Our ensemble,

[*]Strictly speaking this is the Poisson noise from the finite size of the simulation boxes. Cosmic variance is when this is extended to the case of the whole Universe. The distinction is subtle but the terms are used interchangeably in the literature (as is in the present work).

therefore, provides an estimate of the cosmic variance in these simulations.

3.5.1. EoR History and 21-cm Global Signal

Figure 3.10 shows the spread in the reionisation histories for the different subvolumes (in light-blue) relative to L210_AUG (in blue). We find that the range in redshift for reionisation histories among sub-volumes at $x_{\rm HI} \sim 0.5$ is $\Delta z \sim 0.8$. The bottom panel shows the standard deviation of $x_{\rm HI}$ ($\sigma(x_{\rm HI})$) among the subvolumes. We see the same trend among the simulations except that features in the L210_AUG_HIGHX simulation occur earlier relative to the L210_AUG and L210_AUG_LOWX simulations (which are almost identical).

In Figure 3.11, we do a similar analysis for the 21-cm GS, with the top panel showing the signal from the subvolumes and the bottom panel showing the standard deviation. Comparing the bottom panels of Figure 3.10 & Figure 3.11, we see that the fractional change in $\overline{\delta T}_{\rm b}$ is higher than in $x_{\rm HI}$. During the EoH, the scatter in $\delta T_{\rm b}$ is driven by variations in $T_{\rm S}$ while during reionisation the scatter in $x_{\rm HI}$ dominates.

3.5.2. 21-cm Power Spectra

Figure 3.12 shows the 21-cm PS from the L210_AUG simulation (in blue) at $x_{\rm HI} \sim 0.95$, 0.8, 0.70, 0.60, 0.30, and 0.10. The 21-cm PS from the 27 subvolumes (in light blue) are also shown. The scatter in the 21-cm PS increases for decreasing k-value (towards large scales) and decreasing redshift (as reionisation progresses). The spread in power for large k-values is larger than the spread at small k-values for all redshifts. This is due to sample variance since there are fewer modes at these large scales in the volumes to average over. At low redshifts, most of the 21-cm emission comes from sparse, isolated neutral patches, leading to considerable scatter in the 21-cm power.

The power spectrum quantifies the variance in amplitudes of a random field on different scales. A purely Gaussian-random field is fully specified by its power spectrum (Peebles 1980). The cosmic variance of the power spectrum, in this case, should simply be the Poisson sampling error, which depends only on the number of modes in each spherical shell in k-space. However, higher-order statistics are required to capture the information for non-Gaussian fields.

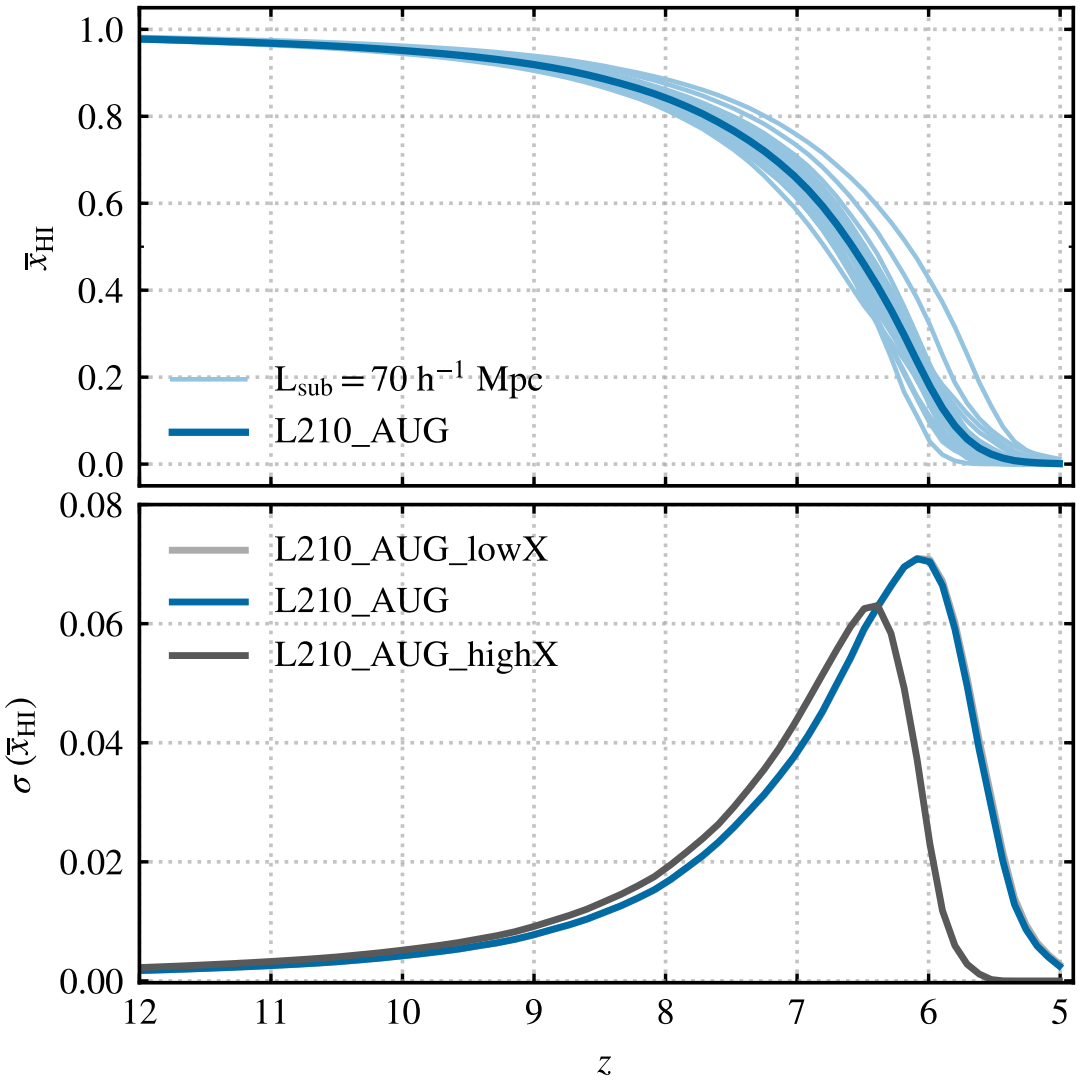

Figure 3.10: The blue curve shows the EoR history of the L210_AUG simulation, and the lighter shades show the EoR history in the 27 subvolumes which the L210_AUG has been divided into. We find a spread in the EoR history with almost $\Delta z \sim 0.8$ around $x_{\rm HI} \sim 0.5$. The bottom panel shows the standard deviation of the $x_{\rm HI}$ [$\sigma(x_{\rm HI})$] among the subvolumes, and we show this for all the augmented simulations. The L210_AUG_LOWX (light grey) curve is identical to the L210_AUG curve and lies behind it.

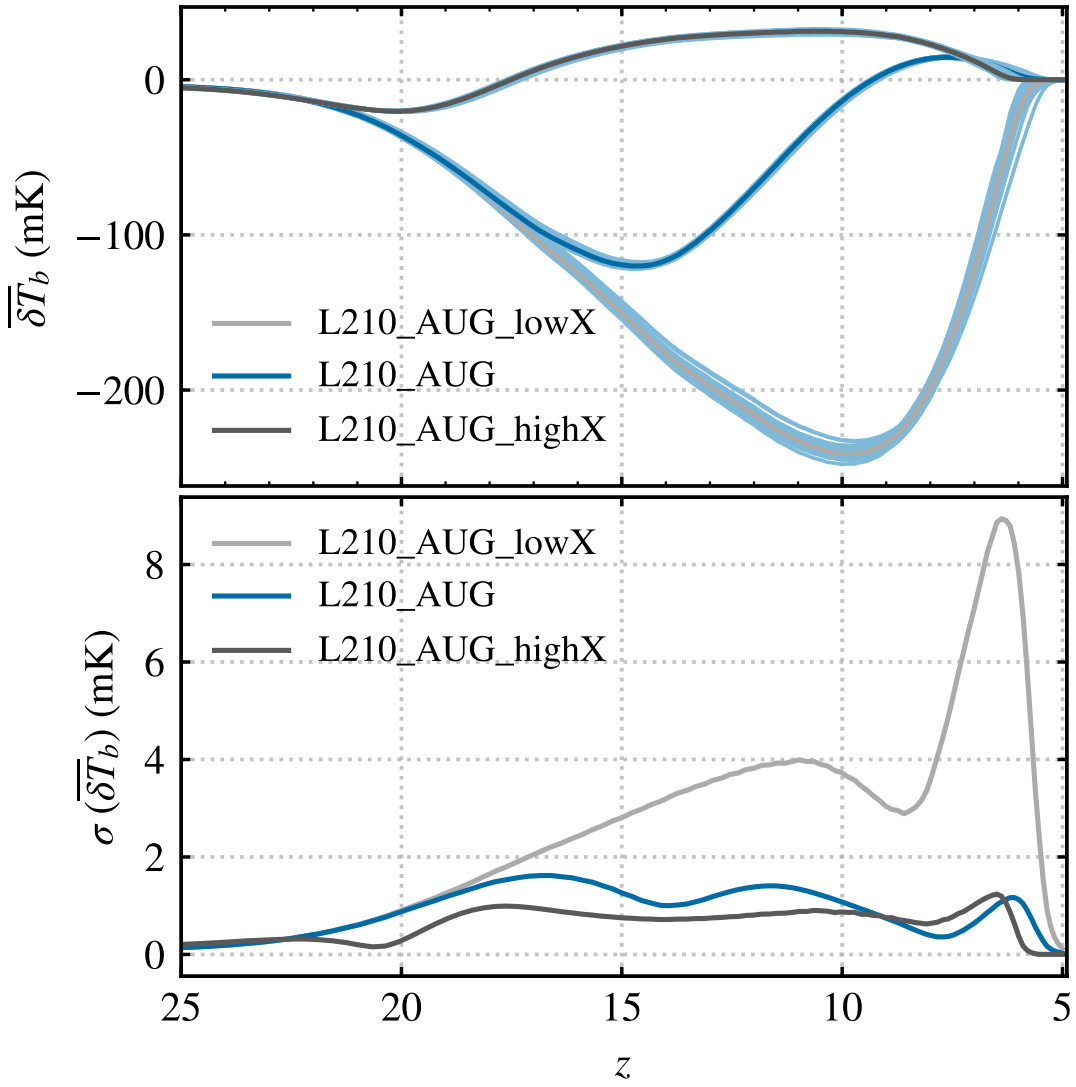

Figure 3.11: The evolution of the 21-cm GS among the subvolumes (light blue) for all three of the augmented simulations. The bottom panel shows the standard deviation of the global signal among the 27 subvolumes.

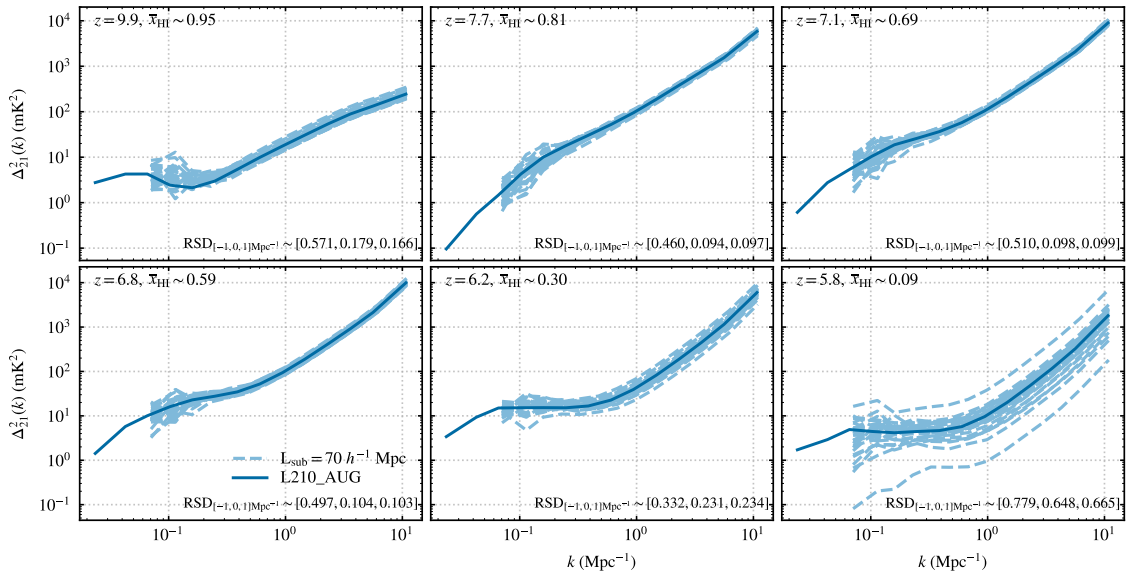

Figure 3.12: The fiducial augmented simulation L210_AUG has been subdivided into 27 equal subvolumes. Shown (dashed light blue) here are the 21-cm power spectra from these subvolumes. The power spectra from the whole volume are also shown (solid blue). The subplots correspond to $x_{\rm HI}$ ~0.95, 0.8, 0.70, 0.50, 0.30, 0.10. We also show the relative standard deviation (RSD = $\frac{\text{standard deviation}}{\text{mean}}$) of the powers at $k = 10^{-1}, 10^{0}$ & 10^{1} Mpc^{-1} spatial scales from the subvolumes.

The 21-cm field is non-Gaussian, especially on small scales and during the final stages of the EoR. Initially, the 21-cm emission traces the underlying matter-density field, which is Gaussian on large scales where the evolution is governed by linear theory. However, once the complex 3D morphology of the radiation fields (e.g. ionisation, X-ray or Lyman-alpha) begins to impact the 21-cm signal, the statistics will deviate from Gaussianity (Morales & Wyithe 2010). Hence, the cosmic variance of the 21-cm power spectrum will be larger than the Poisson sampling error. Here, we explore the impact of non-Gaussianity on the cosmic variance uncertainty of the 21-cm PS.

Mondal et al. (2016) showed that non-Gaussianity has an appreciable impact on the error-covariance of the power spectra. The full error-covariance matrix of the 21-cm PS is given by

$$\boldsymbol{C}_{ij} = \frac{1}{V}\left[\left(\frac{(2\pi)^2[\overline{P}(k_i)]^2}{k_i^2\Delta k_i}\right)\delta_{ij} + \overline{T}(k_i,k_j)\right], \tag{3.15}$$

where V is the simulation volume, k_i is the average spatial frequency in the ith bin, Δk_i is the bin-width of the ith bin, $\overline{P}(k_i)$ is the power spectrum averaged over the ith bin, and $\overline{T}(k_i,k_j)$ is the average trispectrum. This trispectrum component arises from the

non-Gaussianity of the 21-cm signal.

Generally, studies in the literature make the simplifying assumption that the 21-cm field is Gaussian and ignore the second term in equation (3.15) giving

$$\begin{aligned}\delta P(k_i) &= \sqrt{\mathbf{C}_{ii}} \\ &= \sqrt{\frac{(2\pi)^2[\overline{P}(k_i)]^2}{V k_i^2 \Delta k_i}}.\end{aligned} \tag{3.16}$$

Hence, any deviation from equation (3.16) measured from our 27 sub-volumes must occur as a result of the non-Gaussianity of the 21-cm signal.

From each of the 27 equal sub-volumes, we compute the spherically averaged power spectrum and show in Figure 3.13 the ratio of the measured cosmic variance from the sub-volumes ($\sigma_{\rm measured}$) to that expected theoretically from equation (3.16) ($\sigma_{\rm theoretical}$). Specifically, $\sigma_{\rm measured}$ is computed as the standard deviation of the power among the subvolumes as a function of k, while $\sigma_{\rm theoretical}$ is computed using equation (3.16) where the $\overline{\rm P}(k_i)$ is the mean 21-cm PS from the subvolumes. For a Gaussian field, we expect the ratio $\sigma_{\rm measured}/\sigma_{\rm theoretical}$ to be unity. We provide this ratio for $x_{\rm HI} \sim 0.95$, 0.75, 0.50, and 0.25 in each of our three augmented simulations.

We find similar features and trends among all of our simulations. The ratio* increases from small to large k-values implying that the large-scales are more Gaussian in nature compared to the small scales.

Our results agree qualitatively with Mondal et al. (2015) and Mondal et al. (2016) who show that the non-Gaussianity of the 21-cm field grows with the progress of reionisation. During the early stages of the EoR ($x_{\rm HI} \sim 0.80$ case), we find that the contribution of non-Gaussianity to the variance of the 21-cm PS is comparable to the Gaussian term for $0.1 \lesssim k \lesssim 0.4\,{\rm Mpc}^{-1}$ (where the ratio $\sigma_{\rm measured}/\sigma_{\rm theoretical} \sim \mathcal{O}(1)$) while for $k \gtrsim 2\,{\rm Mpc}^{-1}$ the ratio is > 10. As EoR progresses (see $x_{\rm HI} \sim 0.30$ subplot), this ratio becomes $2-4$ for $0.1 \lesssim k \lesssim 0.4\,{\rm Mpc}^{-1}$ and up to ~ 100 for $k \gtrsim 2\,{\rm Mpc}^{-1}$. At the same time, we find that the transition to non-Gaussianity in our model appears to occur earlier than in Mondal

*At the same time, we caution that it is possible we over-predict the cosmic variance due to the smaller size of our sub-volumes. Our sub-volumes are 70 h^{-1} Mpc, which are slightly smaller than expected for convergence of the statistics. This may also explain why our ratio $\sigma_{\rm measured}/\sigma_{\rm theoretical}$ sits above unity for the largest scales (i.e. where it is expected to be Gaussian).

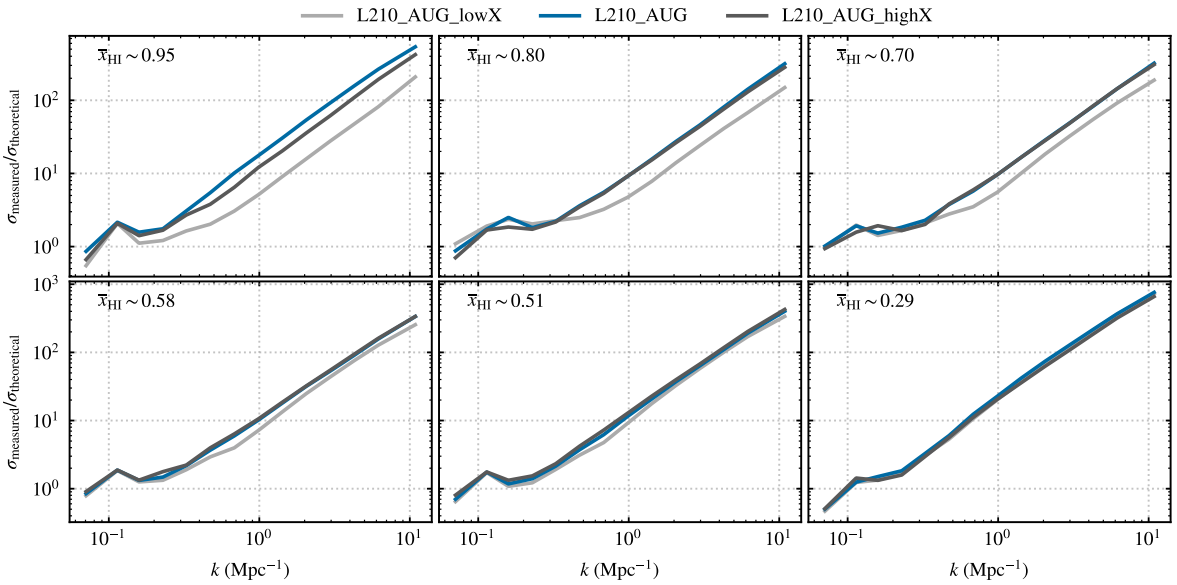

Figure 3.13: Figure shows the ratio of measured to theoretical errors. We compute the standard deviation of the power among the 27 subvolumes as a function of k, $\sigma_{\rm measured}$. We compare this with the $\sigma_{\rm theoretical}$ where we use the mean of the powers among the subvolumes as the $\overline{P}(k_i)$ in equation (3.16).

et al. (2016). Likely, this is a result of the detailed physics prescription of our model, making a direct comparison hard, particularly at high redshifts since our simulations include spin temperature fluctuations, which likely add to the non-Gaussianity in the 21-cm signal.

Our results show that when estimated by assuming that the 21-cm field is Gaussian (i.e. using equation 3.16), towards the end of EoR, the cosmic variance within ~ 100 Mpc boxes is underestimated by a factor of ~ 2 within $k \sim 0.1 - 0.5$ Mpc^{-1} scales which are the main focus of the current and upcoming telescopes observing 21-cm fluctuations.

3.6. Conclusion

In this chapter, we have introduced an updated version of the Meraxes semi-analytic model, which for the first time includes heating from X-rays and thermal evolution of gas in the IGM. In order to have sufficient volume for calculating the effect of X-rays during reionisation, we utilise a new, large-volume N-body simulation with side-length $L = 210\ h^{-1}$ Mpc and 4320^3 particles (L210). To resolve all atomically cooled haloes out to $z = 20$ necessary for studying galaxy formation (of $\sim 2 \times 10^7\ h^{-1}\ {\rm M}_\odot$) we performed Monte-Carlo augmentation of this simulation using DarkForest (L210_AUG).

This achieves an effective N-body particle number of $\sim 10,000^3$. L210_AUG is a unique dataset for exploring galaxy formation physics and its impact on the timing and morphology of the EoR. Coupling MERAXES to this augmented simulation enables the exploration of the different galaxy formation parameters on the 21-cm signal. We found that the inclusion of these Monte-Carlo haloes has a significant impact on the build-up of stellar mass in our simulations and, consequently, on reionisation, which commences earlier and is more gradual. We also find that Lyα-coupling and X-ray heating, and hence the end of the 21-cm global minima, occur earlier in the higher resolution simulation. In addition, we find that the timing and duration of the peaks of the 21-cm power spectrum (PS) are different in the augmented higher-resolution simulation. These results underscore the need for both large volume and sufficient mass resolution for simulations exploring the EoR.

The large volume of our simulation and the implementation of thermal and spin temperature evolution in MERAXES enables exploration of the impact of X-ray luminosity on heating the HIgas. In agreement with semi-numerical studies (Mesinger et al. 2013, Greig & Mesinger 2017), we show that while their impact on the reionisation history is minimal, X-rays can have an appreciable impact on both the 21-cm GS and on the 21 PS. Observations of the 21-cm PS will thus provide constraints on the X-ray properties of the sources in the early Universe.

Taking advantage of the large volume of our simulation, we explore the scatter in the reionisation history and the 21-cm global signal within 27 sub-volumes of side 70 h^{-1} Mpc, which are each comparable to our previous simulations and state-of-the-art radiation-hydrodynamical simulations in the literature. We compare the standard deviation in the 21-cm PS amongst these sub-volumes to the Gaussian expectation for the variance of a random field. As previously described in Mondal et al. (2016), we find that the non-Gaussianity of the signal contributes significantly to the variance of the 21-cm PS on all scales and increases towards the small scales. However, this work is the first study of the error-variance of the 21-cm PS at high redshifts in a model that also includes both a model of galaxy formation and spin temperature fluctuations. We find that the assumption of Gaussianity for the 21-cm field results in underestimating the cosmic variance of the 21-cm PS by a factor of $\gtrsim 2$ for the scales relevant for the SKA ($k \sim 0.1 - 0.5$ Mpc^{-1}).

"Astronomy? Impossible to understand and madness to investigate."

Sophocles, c. 420 BCE

"No."

William Shakespeare, *Hamlet, Act 3, Scene 3, line 87*

CHAPTER 4

Fisher matrix forecasts on the astrophysics of galaxies during the epoch of reionisation from the 21-cm power spectra

This chapter is based on

S. Balu, B. Greig, J. S. B. Wyithe,

Monthly Notices of the Royal Astronomical Society, 525, 2, 2023,3032

reformatted with the following changes only:

- **The text is styled to match the rest of this thesis.**
- **Figure 4.3 showing the distributions of the stations comprising the SKA-low is a new addition.**
- **Where necessary, bibliographic records are updated.**

Abstract

The hyperfine 21-cm transition of neutral hydrogen from the early Universe ($z > 5$) is a sensitive probe of the formation and evolution of the first luminous sources. Using the Fisher matrix formalism we explore the complex and degenerate high-dimensional

parameter space associated with the high-z sources of this era and forecast quantitative constraints from a future 21-cm power spectrum (21-cm PS) detection. This is achieved using MERAXES, a coupled semi-analytic galaxy formation model and reionisation simulation, applied to an N-body halo merger tree with a statistically complete population of all atomically cooled galaxies out to $z \sim 20$. Our mock observation assumes a 21-cm detection spanning $z \in [5, 24]$ from a 1000 h mock observation with the forthcoming Square Kilometre Array and is calibrated with respect to ultraviolet luminosity functions (UV LFs) at $z \in [5, 10]$, the optical depth of CMB photons to Thompson scattering from *Planck*, and various constraints on the IGM neutral fraction at $z > 5$. In this work, we focus on the X-ray luminosity, ionising UV photon escape fraction, star formation and supernova feedback of the first galaxies. We demonstrate that it is possible to recover 5 of the 8 parameters describing these properties with better than 50 per cent precision using just the 21-cm PS. By combining with UV LFs, we are able to improve our forecast, with 5 of the 8 parameters constrained to better than 10 per cent (and all below 50 per cent).

4.1. INTRODUCTION

Following recombination, the early Universe entered the cosmic Dark Ages, characterised by a neutral intergalactic medium (IGM) and the absence of luminous sources. The formation of the first stars and galaxies ushered in the era of the Cosmic Dawn. The intense ionising ultraviolet (UV) radiation, characteristic of the young and massive stars, as well as X-rays (possibly from high-mass X-ray binaries, e.g. Mesinger et al. 2013) impacted the thermal and ionisation state of the IGM. This period was brought to its conclusion in the Epoch of Reionisation (EoR) when the UV photons ionised the neutral hydrogen (HI) rendering it transparent to UV photons.

Considerable effort has been expended in the last few decades to unravel the complex physics of this period (see Mesinger (2019) and references therein). The forbidden 21-cm hyperfine transition signal is well-suited for this purpose because of its extreme sensitivity to the formation and evolution of the first stars and galaxies ($z \lesssim 30$) as well as the various feedback mechanisms in the early Universe. Though the ultimate goal of 21-cm interferometric experiments is to map the tomography of the IGM as a function of

frequency (redshift) the first set of observations will be statistical in nature. The 21-cm signal is observed as a brightness temperature against a background source (which is almost always assumed to be the cosmic microwave background (CMB) radiation; e.g. Furlanetto et al. 2006b; Morales & Wyithe 2010; Pritchard & Loeb 2012). The 21-cm power spectrum (21-cm PS) quantifies the fluctuations in the brightness temperature of the 21-cm signal across the sky. Though a detection has yet to be made, current experiments such as the Murchison Widefield Array (MWA*; Tingay et al., 2013), LOw Frequency ARray (LOFAR†; Van Haarlem et al., 2013), Hydrogen Epoch of Reionization Array (HERA‡; Deboer et al., 2017) have already begun setting upper limits on the 21-cm PS (see Mertens et al., 2020; Trott et al., 2020; The HERA Collaboration et al., 2022). The upcoming Square Kilometre Array (SKA§; Koopmans et al., 2014) will revolutionise 21-cm EoR cosmology with its unprecedented sensitivity.

The amplitude and shape of the 21-cm PS are extremely sensitive to the thermal and ionisation state of the IGM (Barkana & Loeb, 2001) and hence the astrophysics of early galaxy formation and evolution (Geil et al., 2016; Mesinger, 2019). It is therefore imperative that realistic and efficient models (see Gnedin & Madau, 2022, for a recent review) are available to interpret current and upcoming observations. Despite significant progress in this direction, there is considerable uncertainty about the properties of the underlying source populations of these models.

In this chapter, we ask: *What can we learn about the underlying physical processes driving reionisation from a successful detection of the 21-cm PS?* Simulating reionisation requires large volumes ($\gtrsim 200\ h^{-1}$ Mpc according to Iliev et al. (2014); Kaur et al. (2020) for convergent reionisation and 21-cm statistics). As a result, a number of models have been developed to make large-scale but computationally efficient realisations of the early Universe (e.g. Santos et al., 2010; Battaglia et al., 2013; Ghara et al., 2015; Hassan et al., 2016; Choudhury & Paranjape, 2018; Murray et al., 2020). To investigate the utility of the 21-cm PS as a probe of galaxy formation, MCMC methods have been utilised to place constraints on the parameterised properties (e.g. UV and X-ray) of population II & III star-forming galaxies (e.g. Greig & Mesinger, 2015, 2017; Park et al., 2019; Qin et al., 2020,

*www.mwatelescope.org
†www.astron.nl/telescopes/lofar
‡reionization.org/
§www.skao.int

2021a; Maity & Choudhury, 2022b; Bevins et al., 2023).

In this work, we use MERAXES (Mutch et al., 2016) - a semi-analytic model (SAM) of galaxy formation (see Somerville & Davé, 2015, for a recent review) and evolution self-consistently coupled to a reionisation model* - to forecast constraints on astrophysical properties of the early galaxies.

Unlike semi-numerical models (for example, Mesinger et al. 2011, Choudhury & Paranjape 2018), which focus on population-averaged quantities (i.e. these models generally have no galaxies), MERAXES provides a realistic population of galaxies as sources of photons. MERAXES incorporates a detailed, physically motivated galaxy formation and evolution model that includes baryonic infall, gas cooling, star formation, supernova feedback, active galactic nuclei feedback, galaxy mergers, etc. Another essential feature of MERAXES , important for its application to reionisation, is the simultaneous processing of all the galaxies in the simulation volume, thereby enabling the spatial coupling of reionisation to galaxy evolution.

By efficiently coupling the reionisation of the IGM via a modified version of 21CMFAST (Mesinger et al., 2011; Murray et al., 2020) and an underlying galaxy population sourced from a dark matter only N-body simulation, MERAXES is thus well-suited to explore the underlying parameter space of the complex astrophysics of this era. We deploy MERAXES on a 210 h^{-1} Mpc cosmological simulation, which resolves all the atomically cooled galaxies down from $z \sim 20$. The large volume and high mass resolution of our simulation make an MCMC analysis using MERAXES prohibitively expensive computationally†.

Using a Fisher matrix analysis, we forecast the constraints on a total of 8 astrophysical parameters in our model that directly control the X-ray luminosity, UV escape fraction, star formation rates and supernova feedback of the galaxies of the high-z Universe. Focusing on the upcoming SKA1-low, we forecast constraints from the 21-cm PS before exploring the improvements available when combining information from the UV LFs.

The chapter is organised as follows: In section 4.2 we introduce our N-body simulation

*For alternate approaches see (e.g. Seiler et al., 2019; Visbal et al., 2020; Hutter et al., 2021; Ma et al., 2023)

†Recently, Mutch et al. (2023) utilised MERAXES to place constraints on the UV escape fraction of the early galaxies using existing high-z observations within an MCMC framework. This was only feasible owing to (i) not considering the thermal state of the IGM and (ii) using a simulation volume 30 times smaller with a 6 times lower mass resolution for the galaxies.

(section 4.2.1), galaxy SAM (section 4.2.2) and the reionisation model (section 4.2.3). We also introduce our set of eight astrophysical parameters in this section. In section 4.3 we introduce our mock observation, in section 4 we describe the Fisher Matrix formalism and we analyse our results in section 5. All quantities quoted are in comoving units unless otherwise stated.

4.2. Simulating the 21-cm signal

We give a brief review of our underlying dark matter-only N-body simulation and the Meraxes SAM in this section. We focus on a subset of the free parameters in our model that directly impact the star formation, supernova feedback, UV escape fraction, and X-ray luminosity of the galaxies.

4.2.1. N-body merger trees

For this work, we use L210_AUG, introduced in chapter 3, as the fiducial simulation. As pointed out, L210_AUG is capable of resolving all the atomically cooled haloes down from $z = 20$ in a 210 h^{-1} Mpc simulation volume.

4.2.2. A realistic galaxy population from Meraxes

In the following sections, I summarise the relevant implementations in Meraxes. See chapters 2 and 3 for more details. Table 4.1 lists the free parameters of our model along with their adopted fiducial values.

Star formation prescription

The star formation prescription of Meraxes is detailed in section 2.2.3. In this work, I predict the fractional uncertainty on the free parameters $\Sigma_{\rm SF}$ and $\alpha_{\rm SF}$ in equation 2.5 in section 4.5.

Supernova feedback prescription

The star formation rates and the stellar mass of galaxies are regulated by a number of feedback mechanisms, including supernovae (SNe) feedback from evolved stars, AGN

Parameter	Description	Equation	Fiducial value	21-cm PS alone 1-σ values (% uncertainty)	UV LF alone 1-σ values (% uncertainty)	21-cm PS & UV LF 1-σ values (% uncertainty)
$\log_{10}\left(\frac{L_X}{\mathrm{SFR}}\right)$	X-ray luminosity per SFR	3.12	40.50	0.0043		0.0027
				(1.1×10^{-2})	–	(7.0×10^{-3})
E_0	Minimum X-ray photon energy	3.12	500.00	54.5058		27.708
				(10.9)	–	(5.54)
$f_{\mathrm{esc},0}$	Escape fraction normalisation	3.14	0.14	0.0092	0.1172	0.0069
				(6.55)	(83.71)	(4.91)
α_{esc}	Escape fraction redshift scaling	3.14	0.20	0.1339	0.2552	0.0965
				(66.93)	(127.6)	(48.25)
$\log_{10}(\Sigma_{\mathrm{SF}})$	Critical mass normalisation	2.4	-1.86	4.3496	0.9511	0.7401
				(233.9)	(51.13)	(39.8)
$\log_{10}(\alpha_{\mathrm{SF}})$	Star formation efficiency	2.5	-1.00	0.0883	0.0661	0.0436
				(8.82)	(6.61)	(4.35)
$\log_{10}(\epsilon_0)$	Supernova ejection efficiency	2.11	0.19	0.0470	0.0668	0.0309
				(25.1)	(35.16)	(16.56)
$\log_{10}(\eta_0)$	Supernova reheat efficiency	2.12	0.84	0.9876	0.0991	0.0658
				(117.03)	(11.8)	(7.803)

Table 4.1: The first column lists the free astrophysical model parameters for which we forecast the uncertainties with a Fisher matrix formalism. The next three columns give a short description, the corresponding equation in the text, and their adopted fiducial values in this work, respectively. The fifth and sixth columns list the forecasted 1-σ constraints using just the 21-cm PS and just the UV LFs, respectively, and the final column gives the same for a joint analysis of both the 21-cm and the UV LFs. The fractional uncertainties are given in brackets as a percentage. We do not vary the X-ray parameters, L_X/SFR and E_0, for the UV LFs analysis as they do not have an impact on the UV LFs.

feedback, and the ionisation state of the IGM regulated by the progress of reionisation. Mutch et al. (2016) showed that SNe feedback dominates over self-regulation by the EoR.

Section 2.2.4 details the supernova implementation within MERAXES. In this work, I forecast the fractional uncertainty on both ϵ_0 and η_0 (see equations 2.11 & 2.12) and summarise their values in Table 4.1.

Escape fraction of the UV ionising photons

The fraction of ionising photons escaping into the IGM from galaxies plays an important role in regulating the ionisation fraction and morphology. The model parameters that directly impact the ionisation state have a pronounced effect on the 21-cm PS. Following Mutch et al. (2016), we employ an escape fraction ($f_{\rm esc}$) prescription for galaxies that is solely redshift dependent* (though see, e.g. Kimm et al., 2017; Yeh et al., 2023, for $f_{\rm esc}$ implementations that vary with other galaxy properties). This results in an $f_{\rm esc}$ that is skewed towards higher z, motivated by two factors. First, it is easier for the photons to climb out of the shallower potential well of low-mass high-z galaxies. Second, early galaxies are also characterised by less dust attenuation compared to their low-redshift counterparts. I allow both the escape fraction normalisation $f_{\rm esc,0}$ and redshift scaling $\alpha_{\rm esc}$, from equation 3.14, to be free parameters in this work.

4.2.3. *Evolution of the IGM*

The thermal evolution and ionisation state of the IGM follows 21CMFAST (Mesinger et al., 2011; Murray et al., 2020), modified to take advantage of our realistic galaxy population. MERAXES sources the dark matter density and velocity fields from the underlying N-body simulation.

Thermal state of the IGM

The thermal state of a HIcloud, characterised by the spin temperature $T_{\rm S}$, depends upon the radiation that is impinging upon it. Even though $T_{\rm S}$ is influenced by both UV and X-ray photons, the latter has a considerably more pronounced impact (e.g. Mesinger et al., 2013).In this work, I forecast constraints on both E_0 and $L_{\rm X}$/SFR † (see equations 3.10 -

*See equation 3.14 and the associated discussion.

†Note that $L_{\rm X}$/SFR is the same as $L_{\rm X<2keV}$/SFR of chapter 3.

3.12) in this work and their fiducial values are summarised in Table 4.1.

4.3. Fiducial simulation and mock observations

In this section, we describe our fiducial model as well as the mock observations that are used to forecast constraints on our astrophysical model parameters.

4.3.1. Fiducial model parameters

The L210_AUG simulation from Balu et al. (2023) was calibrated against existing observables, including UV LFs from Bouwens et al. (2015) and Bouwens et al. (2021), and stellar mass functions from Song et al. (2016) and Stefanon et al. (2021). The reionisation history was calibrated using the Thomson scattering optical depth of free electrons to CMB photons from the Planck Collaboration et al. (2020) (see figures 3.3 & 3.4 of chapter 3).

The fiducial values of our 8 astrophysical model parameters are given in the fourth column of Table 4.1. These parameters can be divided into four groups of two in the order they are shown in the table, having direct control on the X-ray properties (L_X/SFR & E_0), the ionising UV photon escape fraction ($f_{esc,0}$ & α_{esc}), star formation (Σ_{SF} & α_{SF}), and the supernova feedback (ε_0 & η_0) of the galaxies.

4.3.2. 21-cm power spectra

The frequency dependence of the 21-cm signal imparts a line-of-sight evolution to the 21-signal. To account for this, Meraxes produces a 21-cm light-cone by stitching together (linearly) interpolated δT_b co-evolution grids. Following Greig & Mesinger (2018), we subdivide our 21-cm light-cone into cubical grids of a size equivalent to our simulation volume (1024^3) and calculate the spherically averaged dimensionless 21-cm PS in each (shown as the solid blue curve in Figure 4.1). We obtain 11 redshift chunks spanning $z \in [5, 24]$, and assign redshift values to them corresponding to the mid-point of each chunk.

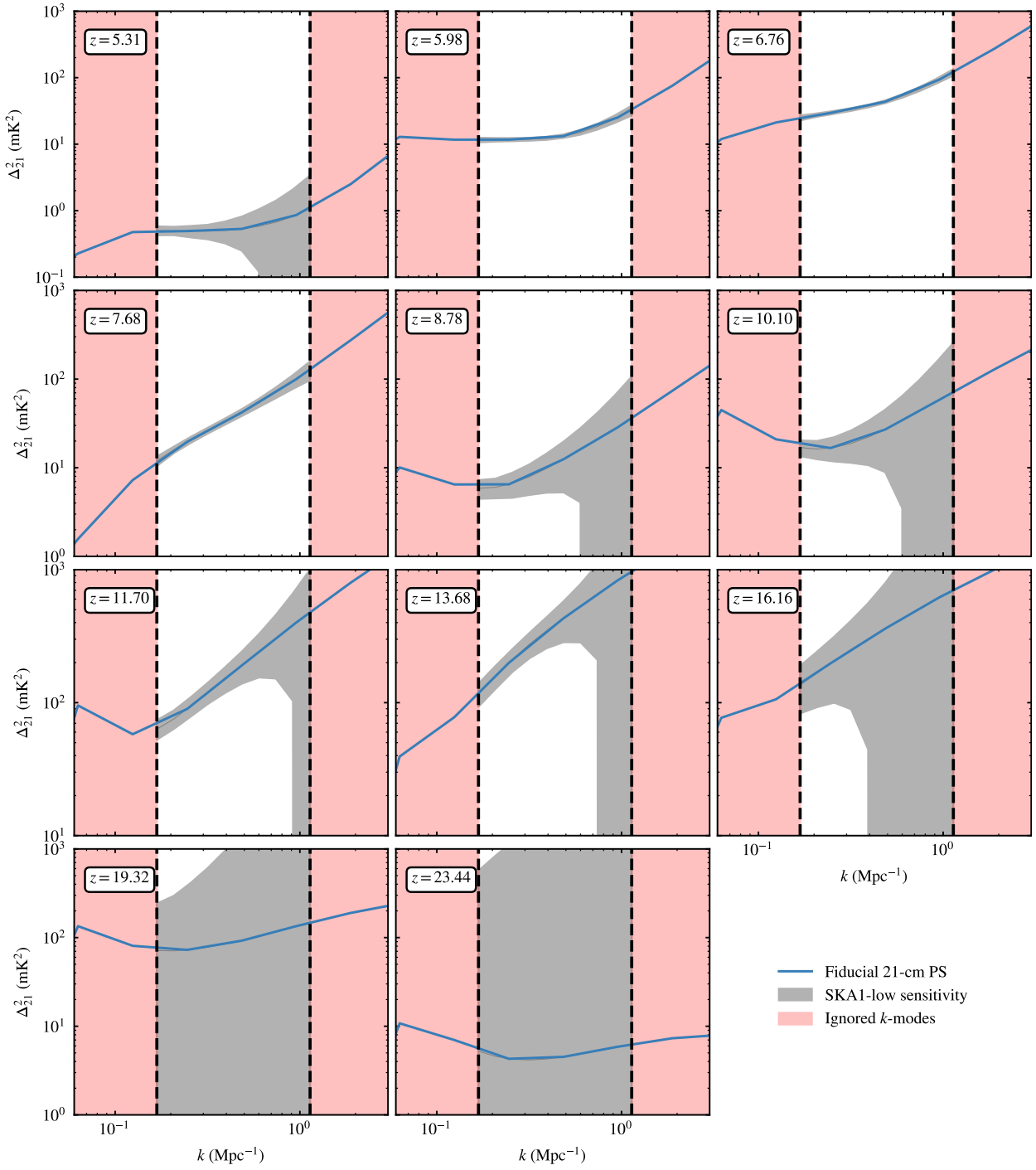

Figure 4.1: The blue curve shows the dimensionless 21-cm PS Δ^2_{21} (mK^2) from our fiducial model. The grey-shaded region represents the sensitivity (including both thermal and cosmic variance noise) to the 21-cm PS for a 1000 h observation with the upcoming SKA1-low. We use the 'moderate' foreground removal case from Pober et al. 2014 effectively ignoring (pink-shaded region) all the k-modes falling within the 21-cm foreground wedge (setting a $k_{\rm min} = 0.16$ Mpc^{-1}). The $k_{\rm max} = 1.4$ Mpc^{-1} is set by a combination of the spatial scales resolved by SKA1-low (set by the longest baseline in our model for the core) as well as scales we trust not to be dominated by the Poisson noise of the sources.

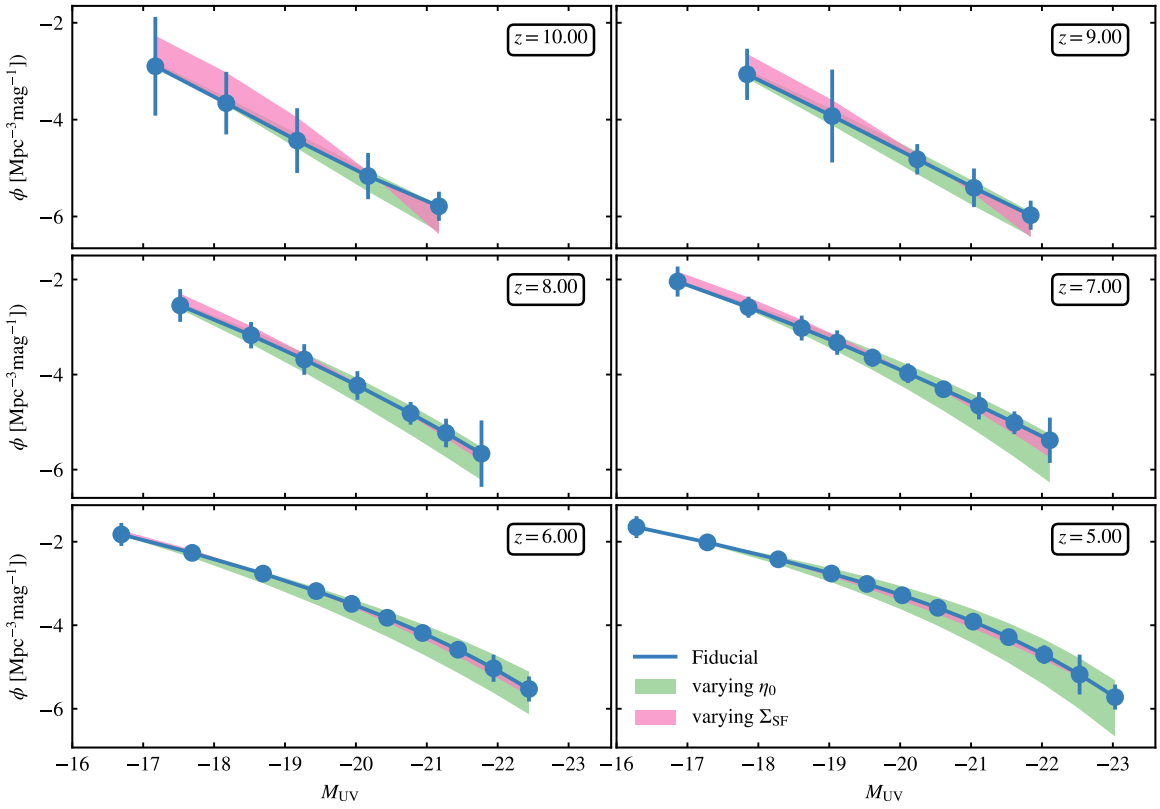

Figure 4.2: The solid blue curve shows the fiducial UV LFs from our calibrated simulation. The uncertainties on our UV LFs are equivalent to the fractional uncertainties from Bouwens et al. 2021. The shaded regions show the variation in the UV LFs as we vary the critical mass normalisation (Σ_{SF}; pink) and the efficiency of gas reheating due to SNe feedback (η_0; green) respectively. These parameters are varied by an order of magnitude about their fiducial values to explore their impact on the UV LFs and clearly demonstrate that including UV LFs in our Fisher matrix will improve our forecasted constraints as the range exceeds the 1-σ observational uncertainty.

4.3.3. *Modelling observational noise*

The sensitivity of a radio interferometer to the 21-cm PS can be divided into two components: (1) thermal noise, which is important on the small scales and (2) sample (cosmic) variance, prominent on the large scales. The total noise power $\sigma[\Delta^2_{21}(k)]$ is given by adding these two in quadrature:

$$\left[\frac{1}{\sigma[\Delta^2_{21}(k)]}\right]^2 = \sum_i \left(\frac{1}{\Delta^2_{\mathrm{N,i}} + \Delta^2_{21}}\right)^2, \tag{4.1}$$

where $\Delta^2_{\mathrm{N}}(k)$ is the thermal noise given by (Morales, 2005; McQuinn et al., 2006; Parsons et al., 2012):

$$\Delta^2_{\mathrm{N}}(k) \approx X^2 Y \frac{k^3}{2\pi^2}\frac{\Omega'}{2t} T^2_{\mathrm{sys}}, \tag{4.2}$$

where X and Y relate the bandwidths and solid angles to the comoving distance to the source, $\Omega^{'}$ is a factor dependant on the beam of the telescope (see Parsons et al., 2014), t is the integration time for the mode k, and T_{sys} is the system temperature given by $T_{\mathrm{sys}} = T_{\mathrm{sky}} + T_{\mathrm{rec}}$ with T_{sky} and T_{rec} being the sky and receiver temperature respectively. The quadrature addition and the form of the cosmic variance noise in equation 4.1 assume that the errors are Gaussian distributed which is reasonable for the relevant scales in this work (see Qin et al., 2021a,b; Prelogović & Mesinger, 2023).

In this work, we focus on a future observation of the 21-cm PS by the SKA. We limit our attention to the upcoming first phase of SKA – the so-called SKA1-low*. Figure 4.3 shows the distribution of the SKA stations. We only include the stations in the 'Central Core' of the SKA1-low, resulting in 224 stations of diameter 35 m distributed across a circular area with 1 km diameter, we calculate the sensitivity of the interferometer to the 21-cm PS using the 21CMSENSE † PYTHON package (Pober et al., 2013, 2014). We assume an observational campaign corresponding to 6 hours per night for 180 days, i.e. a total of 1080 hours and sky temperature T_{sky} to be dominated by galactic synchrotron emission (Thompson et al., 2017) scaling with frequency ν as $T_{\mathrm{sky}} = 60(\nu/300\mathrm{MHz})^{-2.55}\mathrm{K}$. Additionally, assuming that the telescope reflects 10 per cent of the response to the sky (Pober et al., 2014), we set $T_{\mathrm{rec}} = 40\mathrm{mK} + 0.1\mathrm{T}_{\mathrm{sky}}$. We combine partially coherent

*See the official SKA1 System Baseline Design document for further details.
†www.github.com/steven-murray/21cmSense

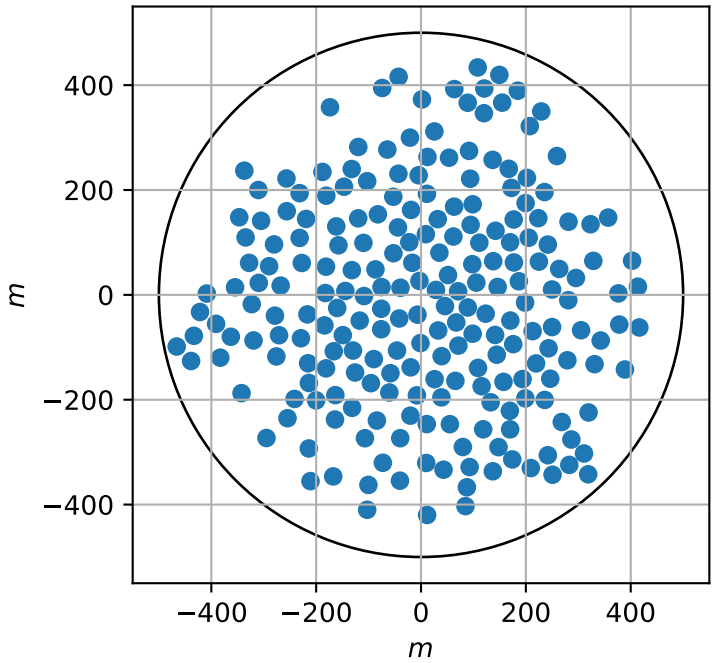

Figure 4.3: The stations that comprises the 'Central Core' of the upcoming SKA-low. There are 224 stations of diameter spread over a circle 1 km in diameter.

baselines to improve power spectrum sensitivity and use the 'moderate' foreground removal scenario of 21CMSENSE wherein we avoid the modes that are contained within the so-called foreground wedge (Datta et al., 2010) extending $k_{\parallel} = 0.1h$ Mpc^{-1} beyond the horizon limit. As we ignore the modes that fall within the foreground wedge, we are limited to $k_{\rm min} = 0.16$ Mpc^{-1} for our analysis. We also have an upper limit of $k_{\rm max} = 1.4$ Mpc^{-1}, which arises from a combination of both the spatial scales that are probed by the SKA1-low (set by the longest baseline we consider in our modelling of the array), as well as Poisson, shot noise from our simulation. Figure 4.1 shows the PS noise (grey shaded region) that is generated for our fiducial 21-cm PS. The red-shaded region shows k-modes that are ignored in this work.

4.3.4. Luminosity Functions

In addition to exploring the possible constraints on our fiducial model from the 21-cm PS we also consider the improvements that are achievable with a joint analysis of both the 21-cm PS and the UV LFs.

In addition to updates on the SNe feedback (see section 2.2.4), Qiu et al. (2019) explored

different implementations of dust attenuation in MERAXES . Here, we adopt the parameterisation for the UV optical depth that depends on the dust-to-gas ratio (Charlot & Fall, 2000) of the galaxies. The model differentiates between a short-lived birth cloud of the stars as well as the interstellar medium (ISM) of the galaxy. Photons are absorbed by both the birth cloud (albeit for a short while until it gets depleted by star formation) and the ISM. The free parameters of MERAXES – particularly the ones impacting the galaxy physics – have been calibrated (see Qiu et al., 2019; Balu et al., 2023) to their fiducial values against infrared excess (IRX)-β, UV LFs, and stellar mass functions at $z > 5$.

The blue curves in Figure 4.2 are the UV LFs from our calibrated simulation. The error bars are determined by multiplying our simulated UV LF data by the corresponding fractional uncertainty on each data point from the Bouwens et al. (2021) observational data. In addition to the 21-cm PS, we thus have 6 UV LFs from $z \in [5, 10]$.

4.4. FORECASTS USING FISHER MATRICES

To place quantitative constraints on the model parameters, we use the Fisher information matrix ($\mathbf{F}_{ij}$; Tegmark et al., 1997; Albrecht et al., 2009). For any set of observations, the Fisher matrix provides the best possible constraints on the parameters of an assumed model. An implicit assumption is that the errors on these parameters are Gaussian and that the observational data points are statistically independent. In this limit, by the Cramer-Rao theorem, the covariance matrix ($\mathbf{C}_{ij}$) of the parameters is given by the inverse of the Fisher matrix

$$\mathbf{C}_{ij} = \mathbf{F}_{ij}^{-1}. \tag{4.3}$$

The 1σ error on the ith parameter is given by the square root of the corresponding term on the diagonal of the covariance matrix, i.e. $\sqrt{\mathbf{C}_{ii}}$. Another relevant property of the Fisher matrices is their additive nature, enabling one to do joint analyses of a different set of observations. This is achieved by adding the corresponding Fisher matrices together before calculating the joint $\mathbf{C}_{ij}$.

Given the likelihood function $\mathcal{L}$ (the probability of the data given the model parameters θ) $\mathbf{F}_{ij}$ is given by

$$\mathbf{F}_{ij} \equiv -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle. \tag{4.4}$$

For the present work, we compute $\mathbf{F}_{ij}$ as

$$\mathbf{F}_{ij} = \sum_{k,z} \frac{1}{\sigma^2(k,z)} \frac{\partial \Delta^2(k,z)}{\partial \theta_i} \frac{\partial \Delta^2(k,z)}{\partial \theta_j}, \tag{4.5}$$

where $\Delta^2(k,z)$ is the dimensionless 21-cm PS, $\sigma(k,z)$ is the observational uncertainty on a measured $\Delta^2(k,z)$ (see equation 4.1), and the summation is over all the k-modes and z-bins. Fisher matrices are thus sensitive to the derivatives of the 21-cm PS, with a larger value indicating increased sensitivity of the statistic to the model parameter. This translates into tighter constraining power on the model parameter. Parameters with a similar structure for the derivatives, as a function of k, will be degenerate (Pober et al., 2014). In Figure 4.4 of appendix 4.5.1, we show the noise $[\sigma(k,z)]$ weighted derivatives of the 21-cm PS with respect to each parameter we study in this work. Figure 4.1 makes it evident that the noise levels corresponding to a ~ 1000 h observation with SKA1-low is lowest for the redshifts $z \lesssim 8$. Thus, from equation 4.5, the constraining power will be dominated by these redshifts. Since the 21-cm PS is sensitive to different astrophysical parameters at different epochs (Furlanetto et al., 2006b), this will be reflected in our forecasted constraints.

Except for E_0, $f_{\rm esc,0}$ and $\alpha_{\rm esc}$, all other parameters are varied in log space as they can vary by at least an order of magnitude. For calculating the Fisher matrix, we vary each parameter by ± 4 per cent around its fiducial value*. In order to further avoid numerical artefacts while computing the derivatives, we fitted the mock 21-cm PS with a $5^{\rm th}$-order polynomial in log-space and interpolated the values for k-bins evenly spaced in $\log_{10}(k)$†.

The $\mathbf{F}_{ij}$ corresponding to the UV LFs are computed by appropriately modifying equation 4.5. Figure 4.7 shows the derivatives of the UV LFs with respect to our model parameters after being weighted by the error $[\sigma(M_{\rm UV},z)]$. We note here that while constructing the Fisher matrix corresponding to the UV LFs, we do not consider the ones varying the X-ray properties (i.e. E_0 & $\log_{10}(\rm L_X/SFR)$) as they do not have any impact on the galaxy properties which affect its luminosity‡.

*We explored different choices of the step-size and (visually) ensured the convergence of the derivatives.

†Polynomial fitting with different orders as well as in linear scales were explored to ensure convergence.

‡The primary impact of the X-ray photons is to set the spin temperature of the IGM gas. For scenarios

The primary advantage of a Fisher matrix analysis is the associated computational efficiency, as they are quick and easy to calculate. This is particularly important when a single model evaluation is too computationally expensive for an MCMC, as is the case for MERAXES*. I would like to stress that the constraints one obtains using the Fisher matrix analysis can be influenced by the adopted fiducial model (see Prelogović & Mesinger, 2024). Nevertheless, even though the particular quantitative constraints might change, as long as the underlying distribution is not particularly complicated in the parameter hyperspace, the relative constraints will still follow the same trend. A possible avenue of exploration is to explore the constraints presented in the next section by adopting a different fiducial model parameter set.

4.5. RESULTS

In this section, we report the quantitative constraints available with a mock observation using just the 21-cm PS and also combining the 21-cm PS with the UV LFs. Before that, we show the derivatives computed as part of the Fisher matrices.

4.5.1. Derivatives of the 21-cm PS

In this section, we show the derivatives of our primary statistics, i.e. the 21-cm PS (Figure 4.4). These derivatives are computed by perturbing the L210_AUG simulation about the fiducial model, one parameter θ_i at a time. These are then used to compute the Fisher matrix (see equation 4.5). Parameters corresponding to similarly shaped derivatives will be degenerate. Figure 4.4 shows these derivatives $[\partial\Delta^2_{21}/\partial\theta_i]$ weighted by the noise powers $[\varepsilon(k,z)]$, i.e. $\frac{1}{\varepsilon}\frac{\partial\Delta^2_{21}}{\partial\theta_i}$. We have employed a 'symmetric log' scaling on the vertical axis in the figure, with the scale being linear $\in [-10, 10]$ and log-scale otherwise.

with extremely high values of the X-ray luminosity, they can cause $\sim 10-15$ per cent impact on the ionisations of the HI (Mesinger et al., 2013). In these extreme cases, the X-ray can provide some constraints from UV LFs via the baryon modifier $f_{\rm mod}$. We have visually checked the constraints and have found them to be negligible for our fiducial model.

*A single model evaluation takes ~ 16 h on two 48-core nodes.

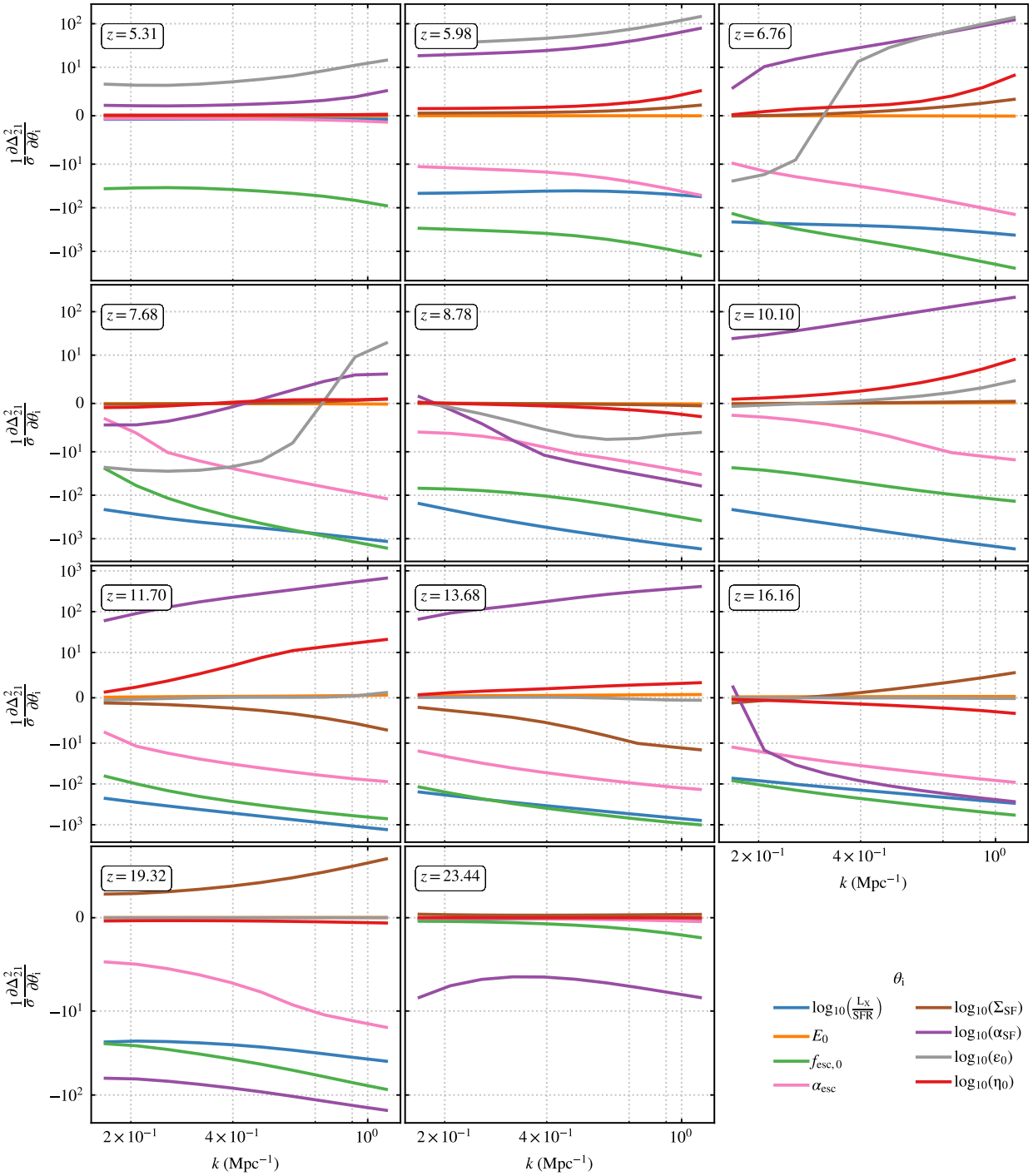

Figure 4.4: The derivatives of the 21-cm PS $[\partial\Delta^2_{21}/\partial\theta_i]$ with respect to the parameters varied in this work. The derivatives are weighted by the noise levels $[\sigma(k,z)]$. The vertical axes is linear $\in[-10,10]$ and log-scale otherwise.

4.5.2. Forecasts from the 21-cm PS alone

In Figure 4.5, we show the forecasts on our astrophysical parameter set from the 21-cm PS corresponding to a ~ 1000 h observation with SKA1-low. The dark (light) contours

correspond to the 1 (2)-σ 2-D marginalised confidence intervals for each astrophysical parameter, and the diagonal subplots show the 1-D normalised marginal probability distribution function. The fifth column of table 4.1 lists the 1-σ uncertainty on our parameters. The fractional uncertainties, $|\sigma(\theta_i)/\theta_i^{fid}|$, are given within brackets and are shown as blue squares in Figure 4.6.

We obtain tight constraints ($\lesssim$ 10 per cent) for the X-ray parameters (luminosity L_X/SFR & minimum X-ray photon energy escaping galaxies E_0), escape fraction normalisation ($f_{esc,0}$) and star formation efficiency (α_{SF}) while the SNe ejection efficiency (ϵ_0) is constrained to $\sim$ 25 per cent. On the other hand, the critical mass normalisation (Σ_{SF}) and the efficiency of SNe reheating (η_0) remain relatively unconstrained with $\sim$ 234 per cent and $\sim$ 117 per cent fractional 1σ uncertainties respectively. The relatively poor constraints on Σ_{SF} and η_0 are primarily because these parameters have negligible impact on the 21-cm PS (see below for more details).

As previously mentioned, our parameter set forms four groups of two with each group controlling a different aspect of galaxy properties.

1. *X-ray parameters*: The 21-cm signal is very sensitive to the amount and energy of the X-ray photons in the early Universe. The 21-cm PS, therefore, provides tight constraints on the X-ray parameters, E_0 ($\sim$ 11 per cent) and L_X/SFR ($\sim 10^{-2}$ per cent), in our model. These forecasts are consistent with similar works in the literature (e.g. Greig & Mesinger, 2017; Park et al., 2019; Mason et al., 2023b), which is unsurprising given they have the same X-ray implementation despite the differences in the source modelling (i.e. SAM).

2. *UV escape fraction*: f_{esc} is more sensitive to the normalisation $f_{esc,0}$ as opposed to the redshift power-law exponent (α_{esc}). This is also reflected in the derivative of the 21-cm PS with respect to these parameters (green and pink curves in Figure 4.4). We thus forecast tighter constraints on $f_{esc,0}$ (6.55 per cent) compared to α_{esc} (66.93 per cent).

3. *Star formation*: Among the star formation parameters, the efficiency of star formation (α_{SF}) is highly constrained relative to the critical mass normalisation (Σ_{SF}). The primary impact of Σ_{SF} is to establish a critical mass threshold needed for the stars to start forming. This parameter thus has only a secondary role in the stellar

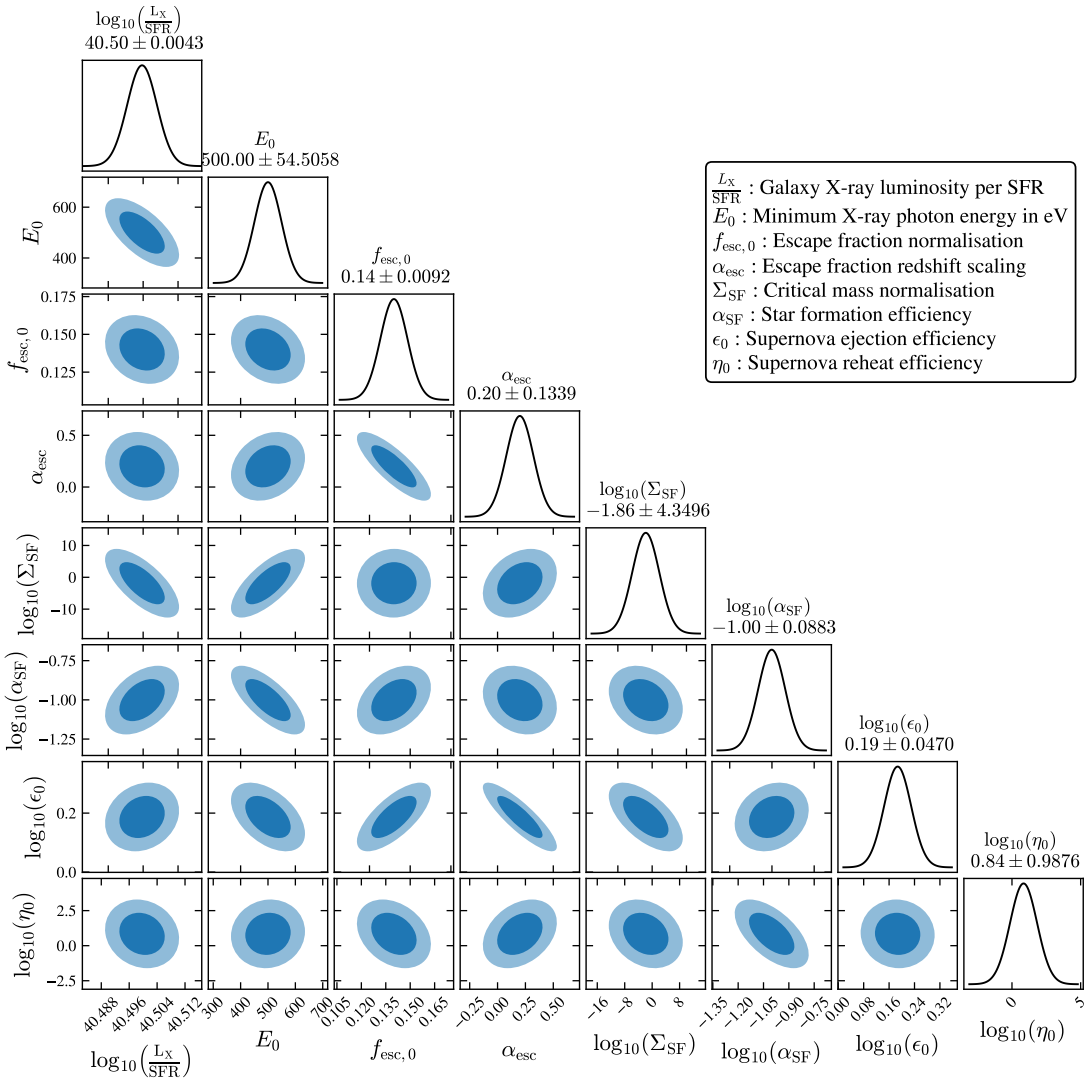

Figure 4.5: Constraints on our astrophysical model parameters from our Fisher matrix using a 1000 h mock observation with the SKA1-low. Dark (light) contours represent the 2-D marginalised confidence intervals, and the diagonal subplots show the 1-D marginalised probability distribution functions of our parameters.

mass content of galaxies as once a galaxy has accreted enough mass, the amount of stellar material is controlled by $\alpha_{\rm SF}$ (see section 4.2.2). Equation 2.5 also shows that these two parameters enter our model as a multiplicative term of the form $(1-\Sigma_{\rm SF})\alpha_{\rm SF}$ adding to their degeneracy.

4. *Supernovae feedback*: Compared to the star formation parameters, we find that the SNe parameters are less constrained, with ~ 25 per cent for the SNe ejection efficiency ϵ_0 while the efficiency of reheating by the SNe η_0 is only constrained to ~ 117 per cent. Similar to $\Sigma_{\rm SF}$, the low constraints on these parameters are due to the fiducial model values being such that the 21-cm signal is relatively insensitive to these processes.

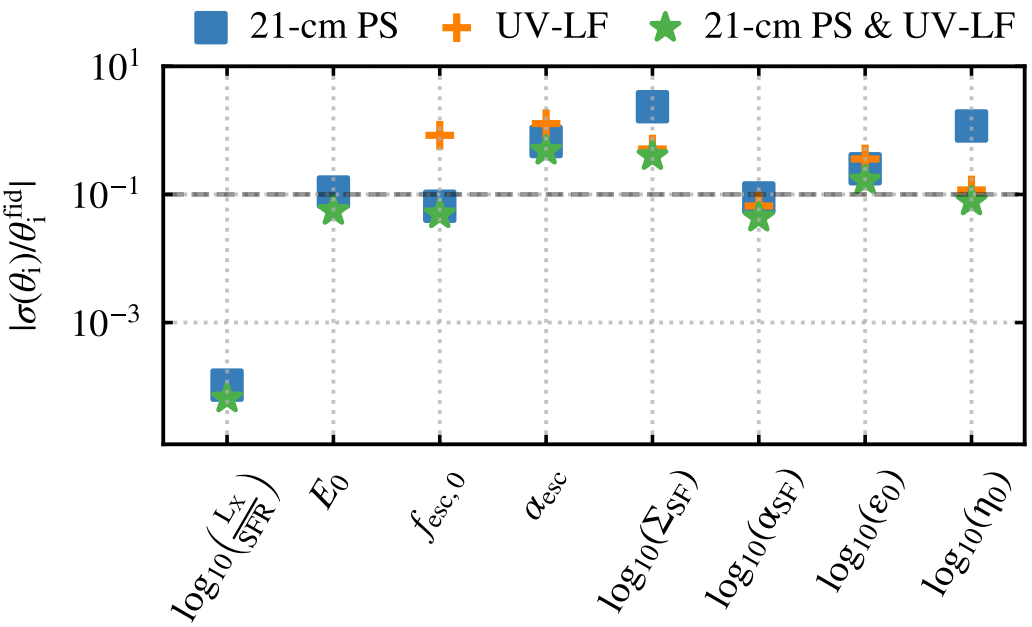

Figure 4.6: Fractional uncertainties on our eight free parameters. The vertical axis is the $1-\sigma$ uncertainties divided by their fiducial values. The blue squares are from just the 21-cm PS, the orange crosses from just the UV LFs, and the green stars are from the joint analysis. While constructing the UV LFs Fisher matrix, we do not consider the parameters controlling the X-ray luminosity of the galaxies, L_X/SFR & E_0. The grey dotted line shows the 10 per cent fractional uncertainty limit .

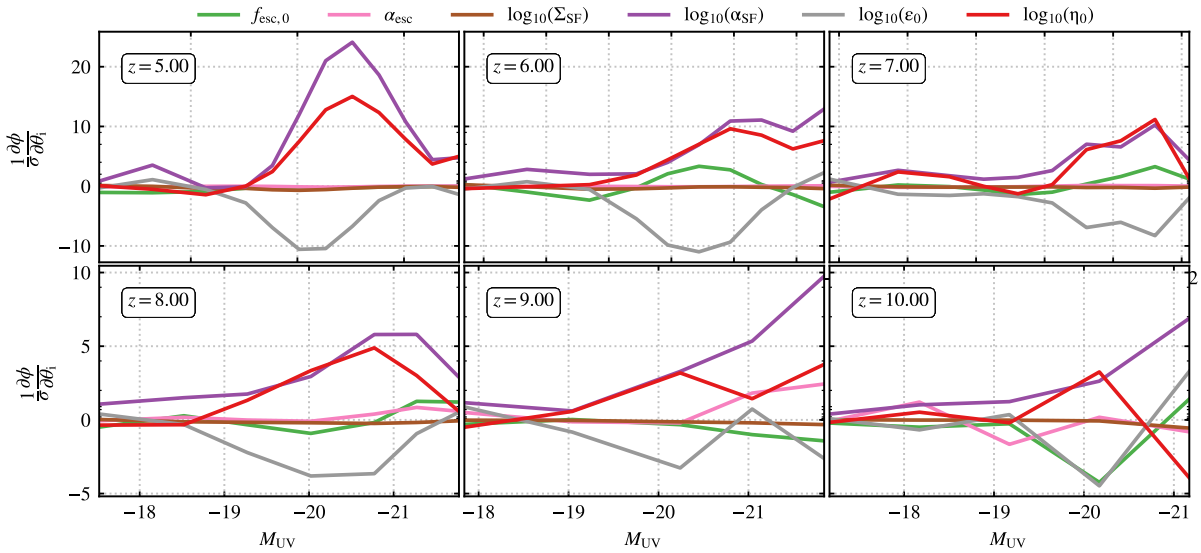

Figure 4.7: The derivatives of the UV LFs $[\partial\phi/\partial\theta_i]$ with respect to the parameters varied in this work. The derivatives are weighted by the noise levels $[\sigma(M_{\rm UV}, z)]$.

4.5.3. *Adding in the UV LFs*

This section shows the analysis using only the UV LFs. Figure 4.7 shows the derivatives of the UV LFs, computed by perturbing the L210_AUG simulation about the fiducial model, one parameter θ_i at a time. These are then used to compute the Fisher matrix. We do not vary the X-ray parameters, $L_{\rm X}$/SFR and E_0, for the UV LFs analysis as they do not have an impact on the UV LFs. Parameters corresponding to similarly shaped derivatives will be degenerate. Figure 4.7 shows these derivatives $[\partial\phi/\partial\theta_i]$ weighted by the noise powers $[\sigma(M_{\rm UV}, z)]$, i.e. $\frac{1}{\sigma}\frac{\partial\phi}{\partial\theta_i}$.

In Figure 4.8, we show the forecasted constraints from the UV LFs Fisher matrix. The dark (light) contours correspond to the 1 (2)-σ 2-D marginalised confidence intervals for each astrophysical parameter, and the diagonal subplots show the 1-D normalised marginal probability distribution function.

4.5.4. *Joint forecasts using the 21-cm PS and the UV LFs*

We next add in the Fisher matrix corresponding to the UV LFs from 6 redshifts $\in [5, 10]$. As the UV LFs functions are sensitive to parameters that do not have a pronounced impact on the ionisation morphology, the addition of the UV LFs into the analysis helps improve the overall constraints on the astrophysical model (Park et al., 2019). We show

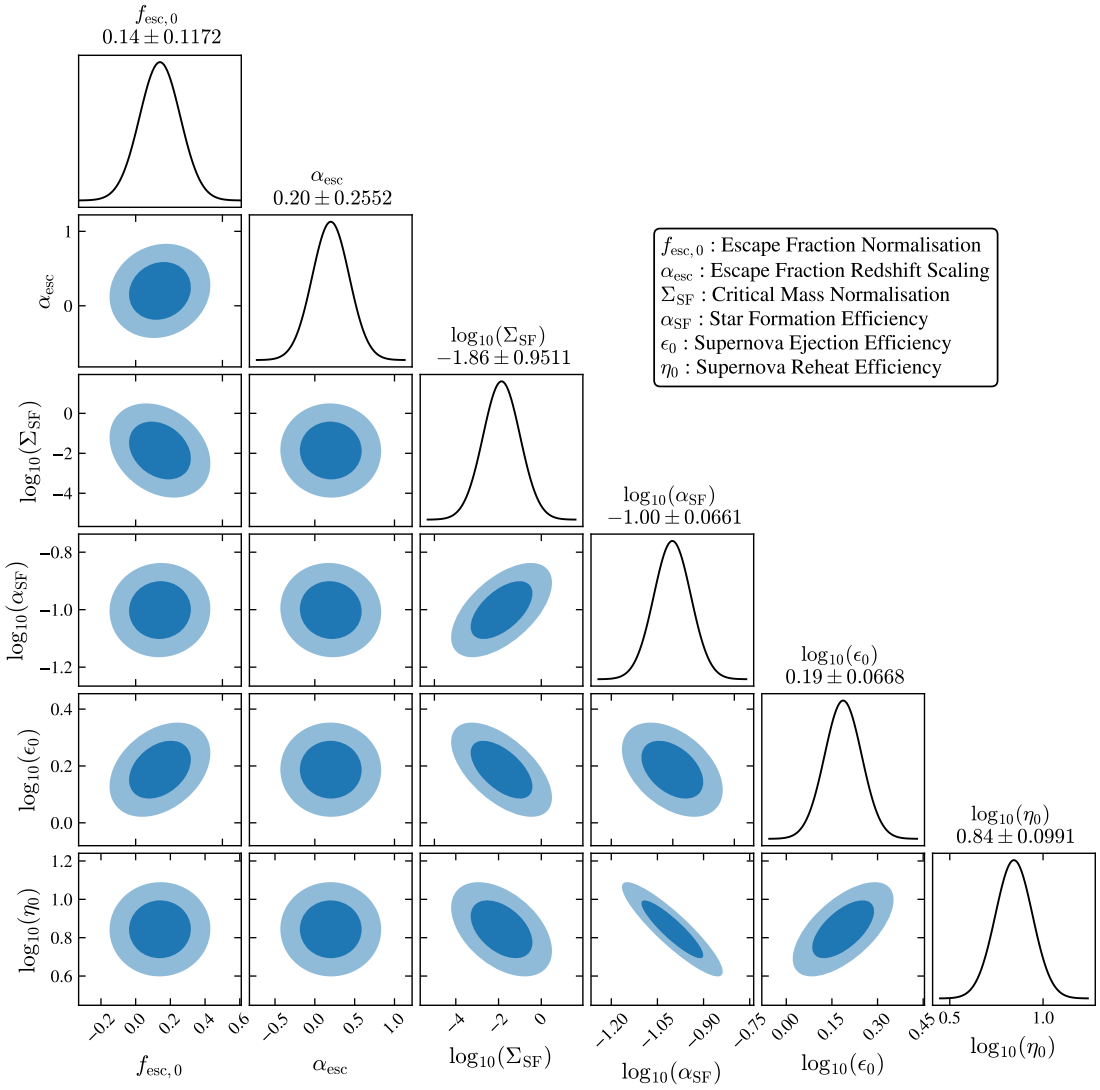

Figure 4.8: Constraints on our astrophysical model parameters from our Fisher matrix analysis using only the UV LFs. Dark (light) contours represent the 2-D marginalised confidence intervals and the diagonal subplots show the 1-D marginalised probability distribution functions of our parameters. Note that while computing the UV LFs Fisher matrices, we do not vary the X-ray parameters (see text for more details).

the impact of varying the two least constrained parameters when using just the 21-cm PS, $\Sigma_{\rm SF}$ and η_0, on the UV LFs in Figure 4.2. The green and pink shaded regions are from varying η_0 and $\Sigma_{\rm SF}$ respectively, by an order of magnitude about their fiducial values (we point out that these two parameters are positive by definition). The variation in the UV LFs is larger than the 1-σ errors, emphasising that including the UV LFs in our Fisher analysis will provide additional constraining power. As previously mentioned, we do not vary the X-ray parameters, $L_{\rm X}$/SFR & E_0, while computing the UV LF Fisher

matrix.

Figure 4.9 shows the constraints from the joint analysis of the 21-cm PS and UV LFs, and the last column of Table 4.1 lists the forecasted 1-σ uncertainty along with the fractional uncertainties for our model parameters. The green stars in Figure 4.6 are the fractional uncertainties for the joint analysis of both the 21-cm PS and the UV LFs, and the orange crosses represent the same from just the UV LFs. Figures 4.5 & 4.6 demonstrate that combining the UV LFs into the analyses helps to improve the constraints. This is most notable for parameters having a direct impact on the stellar mass of the galaxies, i.e. the ones related to star formation and supernova feedback, as UV LFs are more sensitive to these parameters compared to the 21-cm PS. This results in significant improvements in $\Sigma_{\rm SF}$ (from ~ 234 to ~ 40 per cent) and η_0 (from ~ 117 to ~ 8 per cent) for the joint case of both the 21-cm PS and the UV LFs as the degeneracies between stellar properties and escape fraction can be weakened. This then translates into improvements in other parameters.

The relatively larger improvement in $\Sigma_{\rm SF}$ compared to $\alpha_{\rm SF}$ following the inclusion of information from the UV LFs stems from the breaking of parameter degeneracies. In Figure 4.5 (21-cm PS only), $\Sigma_{\rm SF}$ is poorly constrained (highlighted by the large uncertainties), and $\Sigma_{\rm SF}$ and $\alpha_{\rm SF}$ are anti-correlated. In Figure 4.8 (UV LFs only), $\Sigma_{\rm SF}$ is now more tightly constrained and positively correlated with $\alpha_{\rm SF}$, indicating the majority of $\Sigma_{\rm SF}$'s constraining power comes from the UV LF. Thus, the inclusion of UV LFs provides considerable improvements to the constraining power of $\Sigma_{\rm SF}$. For $\alpha_{\rm SF}$ on the other hand, it is almost equally well constrained both by the 21-cm PS (Figure 4.5) and the UV LFs (Fig 4.8). Thus, combining the two observables provides modest improvements (dominated by the change in correlation).

4.5.5. Discussion of parameter degeneracies

Even though Fisher matrices implicitly make simplifying assumptions, such as Gaussian likelihoods and Gaussian errors, they nevertheless give useful insights. In this section, we explore the correlations and degeneracies among the model parameters. For brevity, we focus only on the results from the joint analysis of the UV LFs and the 21-cm PS (Fig 4.9), noting that the trends are similar for the 21-cm PS alone (Figure 4.5).

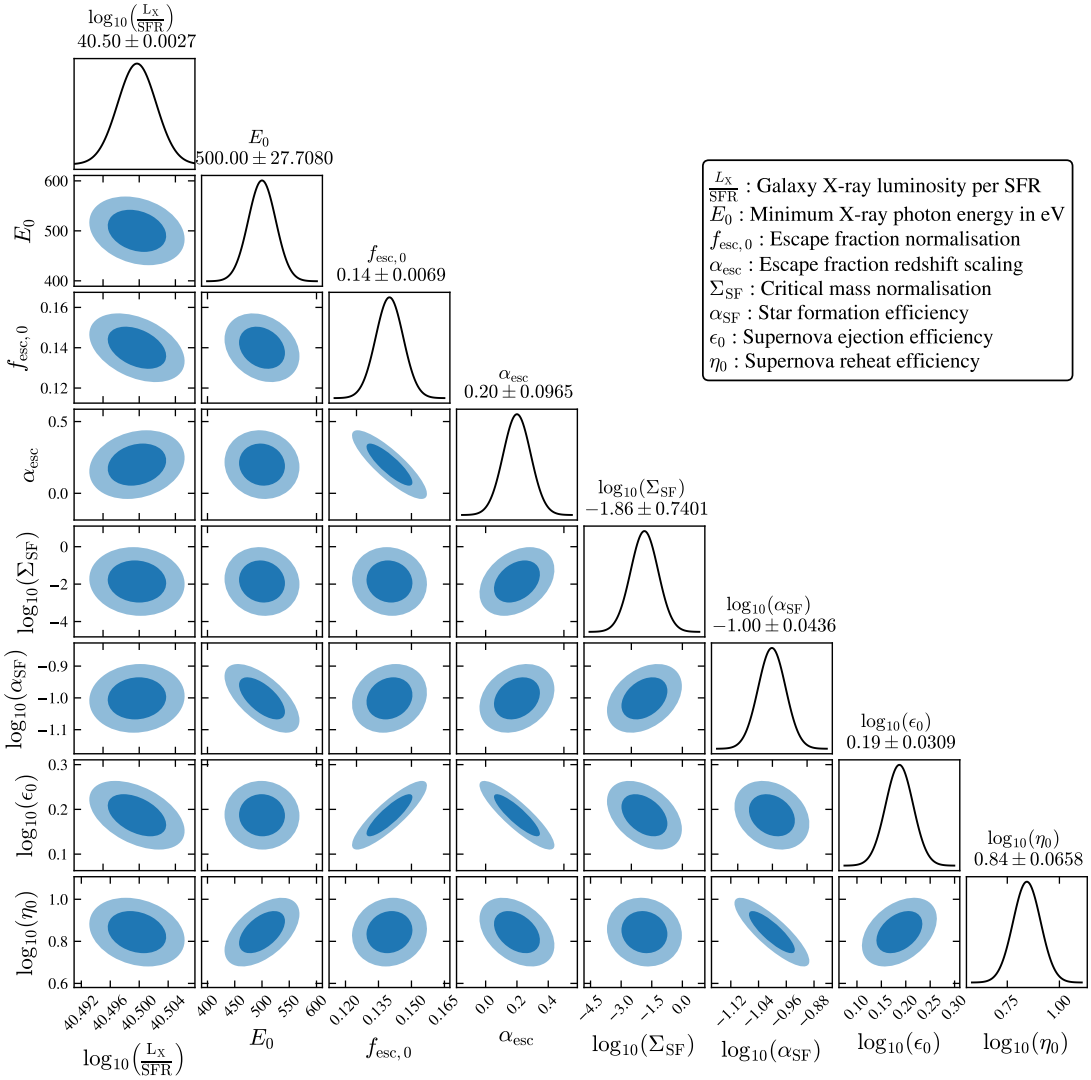

Figure 4.9: Same as Figure 4.5 except that we have also added in constraints from the UV LFs. This removes some of the degeneracies between our parameters and results in tighter constraints.

1. *X-ray parameters*: L_X/SFR anti-correlates with the UV escape fraction normalisation ($f_{esc,0}$) and the SNe ejection efficiency (ϵ_0). E_0 shows an anti-correlation with the star formation efficiency (α_{SF}) and positively correlates with the SNe reheat efficiency (η_0).

   These correlations can be understood from the impact of the X-rays on the EoR. X-rays can contribute, albeit a small fraction relative to the UV photons, to the ionisation of the local IGM of a galaxy, leading to photoionisation regulation of the amount of neutral gas available for accretion onto the galaxy and hence subsequent star formation. Thus, increasing L_X/SFR can be compensated by a decrease in the SNe feedback and/or the ionising UV escape fraction. E_0 sets the lowest energy for escaping X-rays from the galaxies. An increase in E_0, therefore, implies less IGM ionisation by the X-rays.

2. *UV escape fraction*: The correlation among α_{esc} and $f_{esc,0}$ is expected given the definition of the UV escape fraction (f_{esc}; see equation 3.14). The UV escape fraction, f_{esc}, directly influences the local IGM ionisation state and hence the amount of baryonic infall into the galaxy for star formation. The amount of cold gas available for star formation is influenced by the SNe ejection efficiency, which sets the amount of gas removed from the galaxy. An increase in f_{esc}, therefore, should be accompanied by a decrease in the SNe feedback.

   In light of this, the relatively strong positive correlation of $f_{esc,0}$ with the supernova ejection efficiency ϵ_0 is interesting. This can be understood from the behaviour of the UV escape fraction (i.e. the combination of $f_{esc,0}$ and α_{esc}) and the other physical processes. The positive correlation amongst the SNe parameters (η_0 & ϵ_0), as can be seen from equations (2.5 - 2.8) and Figure 4.8, also plays a contributing factor. α_{esc} is also anti-correlated with both η_0 and ϵ_0. It is, therefore, a combination of the very weak correlation of $f_{esc,0}$ with η_0 and the anti-correlation with α_{esc} that results in its positive correlation with ϵ_0.

   The comparatively tighter correlation of the escape fraction parameters with the SNe parameters compared to the X-ray parameters reflects the relative importance of the UV photons over the X-ray photons in the IGM ionisation state.

3. *Star formation*: As noted, we report correlations between the star formation effi-

ciency $\alpha_{\rm SF}$ and the minimum X-ray photon energy E_0.

The strong correlation between $\alpha_{\rm SF}$ and SNe reheat efficiency η_0 is expected as they are two of the parameters impacting the stellar mass in a galaxy.

On the other hand, the correlation between $\alpha_{\rm SF}$ and the critical mass normalisation $\Sigma_{\rm SF}$ is surprisingly weak given equation (2.5). This is because of the relatively small value of $\Sigma_{\rm SF}$. $\Sigma_{\rm SF}$ fixes the critical mass of the cold gas needed for star formation to commence in a galaxy. The small value of the parameter is justified on the physical grounds that a larger value will result in lower available gas for star formation.

4. *Supernovae feedback*: For completeness, we again point out the correlations between the SNe ejection (ϵ_0) and reheat (η_0) efficiencies, the strong correlations of ϵ_0 with $f_{\rm esc,0}$ & E_0, and that of η_0 with $\alpha_{\rm SF}$.

4.6. CONCLUSION

Using the semi-analytic galaxy formation model MERAXES we perform a Fisher matrix analysis to forecast the constraints on physical properties of galaxy formation and reionisation that will be available from future 21-cm PS observations. Specifically, we use the 210 h^{-1} Mpc cosmological simulation resolving all atomically cooled haloes (L210_AUG of Balu et al. 2023) down from $z = 20$.

SAMs are unique in enabling computationally efficient exploration of the underlying parameter space corresponding to the complex astrophysics of galaxy formation and evolution. We focused on 8 free parameters in our model that directly impact the X-ray luminosity, UV escape fraction, star formation rate and SNe feedback of the galaxies. We constructed a mock observation of the 21-cm PS, focusing on a 1000 h observation with the forthcoming SKA1-low. The observational uncertainty was modelled assuming foreground wedge avoidance using the PYTHON package 21CMSENSE. Using the Fisher matrix formalism, we find that 4 (5) out of the 8 parameters can be constrained to within $\lesssim 10$ ($\lesssim 50$) per cent using the 21-cm PS alone from the EoR. Specifically, we forecast constraints on our X-ray parameters (the luminosity $L_{\rm X}$/SFR and the minimum energy of the photon escaping the early galaxies E_0) that are comparable to similar works in the

literature. We also forecast tight constraints on parameters controlling the star formation efficiency ($\alpha_{\rm SF}$) as well as the normalisation of the UV escape fraction ($f_{\rm esc,0}$) of the early galaxies. On the other hand, SNe feedback parameters remain largely unconstrained, reflecting that the 21-cm PS is relatively insensitive to them. The complex astrophysics of the early galaxy formation and evolution is captured in the degeneracies and correlations among the model parameters. Of particular interest are the correlations among the UV escape fraction and SNe feedback and those between the star formation efficiency and X-ray parameters of the galaxies.

To improve the overall constraining power of our analysis, we added the Fisher matrix corresponding to the UV LFs from redshifts $z \in [5, 10]$. This results in an improvement in all of our parameter forecasts, most notably the critical mass normalisation $\Sigma_{\rm SF}$ and the SNe reheat efficiency η_0. This is not surprising as these parameters primarily control the stellar mass content of the galaxies to which the UV LFs are very sensitive. Incorporating the UV LFs into the analyses results in 5 of our parameters being constrained to $\lesssim 10$ per cent and all 8 of them being to within $\lesssim 50$ per cent. Our forecasts illustrate that detailed observations of reionisation with the SKA will be valuable in constraining the astrophysics of the early galaxies.

"Happy is he who gets to know the reasons for things."

Virgil (c. 70–19 BCE)

*"I understand *nothing*."*

Michael Scott, *The Office, S05E24*

CHAPTER 5

Conclusion

In this final chapter, I summarise the work that comprises this thesis and sketch some avenues of possible future directions.

5.1. Summary

The Epoch of Reionisation (EoR) is a milestone event in the history of the Universe. The first generation of the luminous sources in the Universe ionises their surroundings, and these HII bubbles grow with time and overlap, filling the entire Universe. The change in the global ionisation and thermal state of the intergalactic medium (IGM) changed its cooling properties. The EoR thus impacts all subsequent galaxy formation and evolution in the Universe. One of the most promising probes of this period is the 21-cm hyperfine transition of the neutral hydrogen atoms, which is very sensitive to the thermal and ionisation state of the IGM. This thesis is concerned with efficient large-scale cosmological simulations of the HI 21-cm signal from the high redshifts ($z \gtrsim 5$).

In chapter 1, after a brief historical overview of the development of our current understanding of the Universe, I introduce the basics of the ΛCDM model of cosmology. Thereafter, I elaborated on the formation of the large-scale structures in the Universe. I finally concluded the chapter with a brief introduction to the EoR and the various

probes that are used to constrain this period, with a particular focus on the 21-cm line. The physics of the 21-cm line, as well as the various mechanisms that can influence the cosmological 21-cm signal, were discussed.

In chapter 2, after a brief, non-exhaustive summary of the various categories of cosmological simulations in the literature, I introduced the MERAXES semi-analytical model (SAM) of galaxy formation and evolution. SAMs post-process the dark matter halo merger trees from the dark matter-only N-body simulations. Baryonic physics, such as gas cooling, star formation, supernova and active galactic nuclei feedback, galaxy mergers, metal pollution, etc., are implemented as sub-grid prescriptions. SAMs are thus computationally efficient relative to, say, hydrodynamical simulations and are, hence, ideal for exploring the rich and complex parameter space of the underlying physics. MERAXES is one of the few SAMs in the literature that is tailor-made to explore the interplay between the first generation of galaxies and the EoR. This is achieved by an implementation of the IGM reionisation that is based on the semi-numerical code 21CMFAST .

In chapter 3, the first published work of this thesis, I updated MERAXES to include the thermal state of the IGM. The thermal state of the IGM, sensitive to the heating by the X-ray photons, influences the spin temperature (T_S) of the cosmic HI gas. The 21-cm signal depends upon the T_S, and it is vital to be able to model it within simulations. The long mean-free paths of the X-ray photons in the early Universe also necessitate simulation volumes of large volumes. In this work, I deployed MERAXES on the largest simulation, in volume, to date of $210\ h^{-1}$ Mpc on a side. Using DARKFOREST , a Monte-Carlo algorithm-based code, I augmented the merger trees such that this large-volume simulation resolves all haloes of mass $\sim 2 \times 10^7\ h^{-1}\ \mathrm{M}_\odot$, the atomic-cooling limit at $z = 20$. I then explored the impact of the amount of X-ray photons emitted by the galaxies, parameterised as an X-ray luminosity per star formation rate, on the 21-cm signal. Taking advantage of the large volume of the fiducial simulation, I also explored the non-Gaussianity of the 21-cm signal by exploring the cosmic variance of the 21-cm statistics, the 21-cm global signal and the 21-cm power spectra (21-cm PS).

In chapter 4, the second published work of this thesis, I explore the constraints on the galaxy astrophysics that will be enabled by a future detection of the 21-cm PS. Employing the Fisher matrix formalism, I forecasted the constraints on eight of the free parameters of our model. These parameters control the star formation efficiency, su-

pernova feedback, escape faction of the UV ionising photons, and the X-ray luminosity of the galaxies in MERAXES. For a 1000 h observational campaign with the upcoming SKA-low, it should be possible to constrain these parameters to better than 50 per cent. Adding in the constraints from other observables, such as the UV luminosity functions, will help improve these constraints. It is possible to achieve 10 per cent or better with a joint analysis of both the 21-cm PS and the UV luminosity functions.

5.2. FUTURE DIRECTIONS

With innovations in foreground modelling and an improved understanding of systematics and the instruments, the 21-cm EoR cosmology is poised to make big strides with respect to observations. At the time of writing, construction of the SKA had begun in Western Australia. Simulations and models such as the ones in this thesis will be of paramount importance once measurements start coming in. Already, the upper limits of the 21-cm PS are starting to rule out and constrain some of the theoretical models (The HERA Collaboration et al., 2022). The future of 21-cm cosmology is looking brighter than ever.

At other frequencies on the electromagnetic spectrum, the recent launch of the *JWST* has culminated in the high-z Universe witnessing an explosion in observational data. The highest redshift for spectroscopically confirmed galaxies has been pushed to $z \sim 13$ (Curtis-Lake et al., 2023), and photometric candidates have been reported at $z \sim 16$ and above (e.g. Atek et al., 2022; Adams et al., 2022; Naidu et al., 2022; Labbé et al., 2023; Harikane et al., 2023). Some of these galaxy candidates, if confirmed, might challenge our models of galaxy formation and evolution (see for example, Boylan-Kolchin, 2023; Mason et al., 2023a). This is not unexpected as most of our understanding of these is informed to a large extent by observations in the local Universe. Semi-analytical models such as MERAXES are best suited to explore these exciting avenues of research. I was involved in one such work where we identified analogues for $z \gtrsim 12$ galaxies from within the fiducial MERAXES simulations (Qin et al., 2023).

It is evident from these works as well as the past works making use of MERAXES that the model, and SAMs in general, has a wider application to the research in the high-z Universe. As the computational resources at hand and our understanding of the different

astrophysical processes improve, models such as these undergo continued development. For the particular case of MERAXES , works are already underway to implement Population III stars as well as emission line (including Lyα, Hα, Hβ, OII, and OIII) modelling. These will facilitate more accurate modelling of the 21-cm signal as well as applications to galaxy surveys with JWST, Euclid, SPHEREx, ALMA, etc.

Many questions about the Universe await an answer. This thesis, hopefully, has made some contribution in this direction.

Software citations: It is a pleasure to acknowledge the numerous open source software used in this work, including but not limited to, the PYTHON (Van Rossum & Drake Jr, 1995) open source community, in particular, NUMPY (Harris et al., 2020), MATPLOTLIB Hunter (2007), SCIPY (Virtanen et al., 2020), CYTHON: Behnel et al. (2011), H5PY, JUPYTER (Granger & Pérez, 2021), and CORRFUNC: Sinha & Garrison (2020).

References

Adams N. J., et al., 2022, Monthly Notices of the Royal Astronomical Society, 518, 4755

Ahn K., Iliev I. T., Shapiro P. R., Srisawat C., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1486

Albrecht A., et al., 2009, Findings of the Joint Dark Energy Mission Figure of Merit Science Working Group, http://arxiv.org/abs/0901.0721

Almeida A., et al., 2023, The Astrophysical Journal Supplement Series, 267, 44

Alpher R. A., Bethe H., Gamow G., 1948, Physical Review, 73, 803

Andernach H., Zwicky F., 2017, English and Spanish Translation of Zwicky's (1933) The Redshift of Extragalactic Nebulae, doi:10.48550/arXiv.1711.01693, https://ui.adsabs.harvard.edu/abs/2017arXiv171101693A

Atek H., et al., 2022, Monthly Notices of the Royal Astronomical Society, 519, 1201

Balu S., Greig B., Qiu Y., Power C., Qin Y., Mutch S., Wyithe J. S. B., 2023, Monthly Notices of the Royal Astronomical Society, 520, 3368

Barkana R., 2018, Nature, 555, 71

Barkana R., Loeb A., 2001, Physics Report, 349, 125

Barkana R., Outmezguine N. J., Redigol D., Volansky T., 2018, Physical Review D, 98, 103005

Battaglia N., Trac H., Cen R., Loeb A., 2013, The Astrophysical Journal, 776, 81

Bañados E., et al., 2018, Nature, 553, 473

Becker R. H., et al., 2001, The Astronomical Journal, 122, 2850

Becker G. D., Aloisio A. D. ., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, MNRAS, 508, 1853

Behnel S., Bradshaw R., Citro C., Dalcin L., Seljebotn D. S., Smith K., 2011, Computing in Science \& Engineering, 13, 31

Benson A. J., 2012, New Astronomy, 17, 175

Benson A. J., Cannella C., Cole S., 2016, Computational Astrophysics and Cosmology, 3, 3

Bevins H. T. J., Heimersheim S., Abril-Cabezas I., Fialkov A., Acedo E. d. L., Handley W., Singh S., Barkana R., 2023, Joint analysis constraints on the physics of the first galaxies with low frequency radio astronomy data, http://arxiv.org/abs/2301.03298

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, The Astrophysical Journal, 379, 440

Bouwens R. J., et al., 2015, Astrophysical Journal, 803, 1

Bouwens R. J., et al., 2021, The Astronomical Journal, 162, 47

Bower R. G., 1991, Monthly Notices of the Royal Astronomical Society, 248, 332

Bowman J. D., Rogers E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature Publishing Group

Boylan-Kolchin M., 2023, Nature Astronomy, 7, 731

Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, The Astrophysical Journal, 555, 240

Charlot S., Fall S. M., 2000, The Astrophysical Journal, 539, 718

Choudhury T. R., Paranjape A., 2018, Monthly Notices of the Royal Astronomical Society, 481, 3821

Cole S., 1991, The Astrophysical Journal, 367, 45

Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, Monthly Notices of the Royal Astronomical Society, 271, 781

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2002, Monthly Notices of the Royal Astronomical Society, 319, 168

Cora S. A., et al., 2018, Monthly Notices of the Royal Astronomical Society, 479, 2

Croton D. J., 2013, Publications of the Astronomical Society of Australia, 30, e052

Croton D. J., et al., 2006, Monthly Notices of the Royal Astronomical Society, 365, 11

Croton D. J., et al., 2016, The Astrophysical Journal Supplement Series, 222, 22

Curtis-Lake E., et al., 2023, Nature Astronomy, 7, 622

Das A., Mesinger A., Pallottini A., Ferrara A., Wise J. H., 2017, Monthly Notices of the Royal Astronomical Society, 469, 1166

Datta A., Bowman J. D., Carilli C. L., 2010, The Astrophysical Journal, 724, 526

Davies F. B., et al., 2018, The Astrophysical Journal, 864, 142

Deboer D. R., et al., 2017, Publications of the Astronomical Society of the Pacific, 129, 045001

Dijkstra M., Haiman Z., Rees M. J., Weinberg D. H., 2004, The Astrophysical Journal, 601, 666

Einstein A., 1915a

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

Einstein A., 1917, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.), 1917, 142

Elahi P. J., Cañas R., Poulton R. J., Tobar R. J., Willis J. S., Lagos C. D. P., Power C., Robotham A. S., 2019a, Publications of the Astronomical Society of Australia, 36, 1

Elahi P. J., Poulton R. J., Tobar R. J., Cañas R., Lagos C. D. P., Power C., Robotham A. S., 2019b, Publications of the Astronomical Society of Australia, 36, 1

Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778

Fialkov A., Barkana R., 2019, Monthly Notices of the Royal Astronomical Society, 486, 1763

Fialkov A., Cohen A., Barkana R., Silk J., 2017, MNRAS, 464, 3498

Field G. B., 1958, PIRE, 46, 240

Fixsen D. J., 2009, The Astrophysical Journal, 707, 916

Fragos T., et al., 2013, The Astrophysical Journal, 764, 41

Friedman A., 1922, Zeitschrift für Physik, 10, 377

Friedman A., 1999, General Relativity and Gravitation, 31, 1991

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, The Astrophysical Journal, 613, 1

Furlanetto S. R., McQuinn M., Hernquist L., 2006a, Monthly Notices of the Royal Astronomical Society, 365, 115

Furlanetto S. R., Peng Oh S., Briggs F. H., 2006b, Physics Reports, 433, 181

Geil P. M., et al., 2016, MNRAS, 462, 804

Ghara R., Choudhury T. R., Datta K. K., 2015, Monthly Notices of the Royal Astronomical Society, 447, 1806

Gnedin N. Y., 2014, The Astrophysical Journal, 793, 29

Gnedin N. Y., Madau P., 2022, Living Reviews in Computational Astrophysics, 8, 3

Granger B. E., Pérez F., 2021, Computing in Science & Engineering, 23, 7

Greig B., Mesinger A., 2015, Monthly Notices of the Royal Astronomical Society, 449, 4246

Greig B., Mesinger A., 2017, Monthly Notices of the Royal Astronomical Society, 472, 2651

Greig B., Mesinger A., 2018, Monthly Notices of the Royal Astronomical Society, 477, 3217

Greig B., Mesinger A., Davies F. B., Wang F., Yang J., Hennawi J. F., 2022a, Monthly Notices of the Royal Astronomical Society, 512, 5390

Greig B., Wyithe J. S. B., Murray S. G., Mutch S. J., Trott C. M., 2022b, Monthly Notices of the Royal Astronomical Society, 516, 5588

Gunn J. E., Peterson B. A., 1965, The Astrophysical Journal, 142, 1633

Guo Q., et al., 2011, Monthly Notices of the Royal Astronomical Society, 413, 101

Guth A. H., 1981, Physical Review D, 23, 347

Harikane Y., et al., 2023, The Astrophysical Journal Supplement Series, 265, 5

Harris C. R., et al., 2020, Nature, 585, 357

Harrison E. R., 1970, Physical Review D, 1, 2726

Hassan S., Davé R., Finlator K., Santos M. G., 2016, Monthly Notices of the Royal Astronomical Society, 457, 1550

Hoag A., et al., 2019, The Astrophysical Journal, 878, 12

Hubble E. P., 1925, The Observatory, 48, 139

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

Hunter J. D., 2007, Computing in Science & Engineering, 9, 90

Hutter A., Dayal P., Yepes G., Gottlöber S., Legrand L., Ucci G., 2021, MNRAS, 503, 3698

Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U. L., 2014, Monthly Notices of the Royal Astronomical Society, 439, 725

Jeans J. H., 1997, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 199, 1

Jung I., et al., 2020, The Astrophysical Journal, 904, 144

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

Kauffmann G., 1996, Mon. Not. R. Astron. Soc, 281, 475

Kauffmann G., White S. D. M., Guiderdoni B., 1993, Monthly Notices of the Royal Astronomical Society, 264, 201

Kaur H. D., Gillet N., Mesinger A., 2020, Monthly Notices of the Royal Astronomical Society, 495, 2354

Kaurov A. A., Gnedin N. Y., 2015, The Astrophysical Journal, 810, 154

Kim H. S., Wyithe J. S. B., Park J., Poole G. B., Lacey C. G., Baugh C. M., 2016, Monthly Notices of the Royal Astronomical Society, 455, 4498

Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, Monthly Notices of the Royal Astronomical Society, p. stx052

Koopmans L. V., et al., 2014, Proceedings of Science, 9-13-June-2014, 1

Kroupa P., 2002, Science, 295, 82

Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, Monthly Notices of the Royal Astronomical Society: Letters, 485, L24

Labbé I., et al., 2023, Nature, 616, 266

Lacey C., Cole S., 1993, Monthly Notices of the Royal Astronomical Society, 262, 627

Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, Monthly Notices of the Royal Astronomical Society, 481, 3573

Leavitt H. S., Pickering E. C., 1912, Harvard College Observatory Circular, 173, 1

Leitherer C., et al., 1999, The Astrophysical Journal Supplement Series, 123, 3

Lemaître G., 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Lemaître A. G., 1931, Monthly Notices of the Royal Astronomical Society, 91, 483

Lemaître G., 1933, Annales de la Société Scientifique de Bruxelles, 53, 51

Lifshitz E. M., 1946, Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki, 16, 587

Longair M. S., 2020, Concepts in Physics: An Alternative View of Theoretical Reasoning in Physics, 3 edn. Cambridge University Press, Cambridge

Longair M. S., 2023, Galaxy Formation, doi:10.1007/978-3-662-65891-8. , https://ui.adsabs.harvard.edu/abs/2023gafo.book.....L

Lynden-Bell D., 1967, Monthly Notices of the Royal Astronomical Society, 136, 101

Ma Q.-B., Ghara R., Ciardi B., Iliev I. T., Koopmans L. V. E., Mellema G., Mondal R., Zaroubi S., 2023, Monthly Notices of the Royal Astronomical Society, 522, 3284

Madau P., Fragos T., 2017, The Astrophysical Journal, 840

Maity B., Choudhury R., 2022a, MNRAS, 511, 2239

Maity B., Choudhury T. R., 2022b, Monthly Notices of the Royal Astronomical Society, 515, 617

Mason C. A., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 3947

Mason C. A., Trenti M., Treu T., 2023a, Monthly Notices of the Royal Astronomical Society, 521, 497

Mason C. A., Muñoz J. B., Greig B., Mesinger A., Park J., 2023b, Monthly Notices of the Royal Astronomical Society, 524, 4711

McGreer I. D., Mesinger A., D'Odorico V., 2015, Monthly Notices of the Royal Astronomical Society, 447, 499

McQuinn M., 2012, Monthly Notices of the Royal Astronomical Society, 426, 1349

McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, The Astrophysical Journal, 653, 815

Mertens F. G., et al., 2020, MNRAS, 493, 1662

Mesinger A., ed. 2019, The Cosmic 21-cm Revolution: Charting the first billion years of our Universe. IOP Publishing, doi:10.1088/2514-3433/ab4a73, https://iopscience.iop.org/book/978-0-7503-2236-2

Mesinger A., Dijkstra M., 2008, Monthly Notices of the Royal Astronomical Society, 390, 1071

Mesinger A., Furlanetto S., Cen R., 2011, Monthly Notices of the Royal Astronomical Society, 411, 955

Mesinger A., Ferrara A., Spiegel D. S., 2013, Monthly Notices of the Royal Astronomical Society, 431, 621

Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, Monthly Notices of the Royal Astronomical Society, 446, 566

Mesinger A., Greig B., Sobacchi E., 2016, Monthly Notices of the Royal Astronomical Society, 459, 2342

Mineo S., Gilfanov M., Sunyaev R., 2012, Mon. Not. R. Astron. Soc, 419, 2095

Miralda-Escude J., Haehnelt M., Rees M. J., 2000, The Astrophysical Journal, 530, 1

Mondal R., Bharadwaj S., Majumdar S., Bera A., Acharyya A., 2015, Monthly Notices of the Royal Astronomical Society: Letters, 449, L41

Mondal R., Bharadwaj S., Majumdar S., 2016, Monthly Notices of the Royal Astronomical Society, 456, 1936

Morales M. F., 2005, The Astrophysical Journal, 619, 678

Morales M. F., Wyithe J. S. B., 2010, Annual Review of Astronomy and Astrophysics, 48, 127

Murray S., Greig B., Mesinger A., Muñoz J., Qin Y., Park J., Watkinson C., 2020, Journal of Open Source Software, 5, 2582

Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, Monthly Notices of the Royal Astronomical Society, 462, 250

Mutch S. J., Greig B., Qin Y., Poole G. B., Wyithe J. S. B., 2023, Monthly Notices of the Royal Astronomical Society, 527, 7924

Muñoz J. B., Loeb A., 2018, Nature, 557, 684

Naidu R. P., et al., 2022, The Astrophysical Journal Letters, 940, L14

Novikov I. D., 1964, Journal of Experimental and Theoretical Physics, 46, 686

Ocvirk P., et al., 2016, Monthly Notices of the Royal Astronomical Society, 463, 1462

Ocvirk P., et al., 2020, MNRAS, 496, 4087

Oñorbe J., Davies F. B., Lukić Z., Hennawi J. F., Sorini D., 2019, Monthly Notices of the Royal Astronomical Society, 486, 4075

O'Raifeartaigh C., O'Keeffe M., Nahm W., Mitton S., 2017, The European Physical Journal H, 42, 431

Pacucci F., Mesinger A., Mineo S., Ferrara A., 2014, Monthly Notices of the Royal Astronomical Society, 443, 678

Page Lyman A. J., Partridge R. B., Peebles P. J. E., eds, 2009, Finding the Big Bang. Cambridge University Press, Cambridge, doi:10.1017/CBO9780511626500, https://www.cambridge.org/core/product/8838C2494586462BC546B340D6377D8C

Park J., Mesinger A., Greig B., Gillet N., 2019, Monthly Notices of the Royal Astronomical Society, 484, 933

Parkinson H., Cole S., Helly J., 2008, Monthly Notices of the Royal Astronomical Society, 383, 557

Parsons A., Pober J., McQuinn M., Jacobs D., Aguirre J., 2012, The Astrophysical Journal, 753, 81

Parsons A. R., et al., 2014, The Astrophysical Journal, 788, 106

Peacock J. A., 1999, Cosmological Physics. https://ui.adsabs.harvard.edu/abs/1999coph.book.....P

Peebles P. J. E., 1980, The large-scale structure of the universe. Princeton University Press, https://ui.adsabs.harvard.edu/abs/1980lssu.book.....P

Peebles P. J. E., 2020, Cosmology's Century. Princeton University Press, doi:10.2307/j.ctvss3zt8, http://www.jstor.org/stable/j.ctvss3zt8

Peebles P. J. E., 2022, The Whole Truth. Princeton University Press, doi:10.2307/j.ctv25wxbnb, http://www.jstor.org/stable/j.ctv25wxbnb

Penzias A. A., Wilson R. W., 1965, \apj, 142, 419

Planck Collaboration et al., 2016, Astronomy and Astrophysics, 594

Planck Collaboration et al., 2020, Astronomy & Astrophysics, 641, A6

Pober J. C., et al., 2013, The Astronomical Journal, 145, 65

Pober J. C., et al., 2014, The Astrophysical Journal, 782, 66

Prelogović D., Mesinger A., 2023, Monthly Notices of the Royal Astronomical Society, p. stad2027

Prelogović D., Mesinger A., 2024, How informative are summaries of the cosmic 21-cm signal?, http://arxiv.org/abs/2401.12277

Press W. H., Schechter P., 1974, The Astrophysical Journal, 187, 425

Pritchard J. R., Furlanetto S. R., 2007, Monthly Notices of the Royal Astronomical Society, 376, 1680

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

Puchwein E., et al., 2023, Monthly Notices of the Royal Astronomical Society, 519, 6162

Qin Y., et al., 2017, Monthly Notices of the Royal Astronomical Society, 472, 2009

Qin Y., Mesinger A., Park J., Greig B., Muñoz J. B., 2020, Monthly Notices of the Royal Astronomical Society, 495, 123

Qin Y., Mesinger A., Greig B., Park J., 2021a, MNRAS, 501, 4748

Qin Y., Mesinger A., Bosman S. E., Viel M., 2021b, Monthly Notices of the Royal Astronomical Society, 506, 2390

Qin Y., Balu S., Wyithe J. S. B., 2023, Monthly Notices of the Royal Astronomical Society, 526, 1324

Qiu Y., Mutch S. J., da Cunha E., Poole G. B., Stuart Wyithe J. B., 2019, Monthly Notices of the Royal Astronomical Society, 489, 1357

Qiu Y., Mutch S. J., Elahi P. J., Poulton R. J. J., Power C., Wyithe J. S. B., 2020, Monthly Notices of the Royal Astronomical Society, 500, 493

Rahimi M., et al., 2021, Monthly Notices of the Royal Astronomical Society, 508, 5954

Robertson H. P., 1935, The Astrophysical Journal, 82, 284

Rosdahl J., et al., 2018, Monthly Notices of the Royal Astronomical Society

Rubin V. C., Ford Jr. W. K., Thonnard N., 1980, The Astrophysical Journal, 238, 471

Santos M. G., Ferramacho L., Silva M. B., Amblard A., Cooray A., 2010, Monthly Notices of the Royal Astronomical Society, 406, 2421

Schaller M., et al., 2018, SWIFT: SPH With Inter-dependent Fine-grained Tasking, https://ui.adsabs.harvard.edu/abs/2018ascl.soft05020S/abstract

Schechter P., 1976, The Astrophysical Journal, 203, 297

Scheuer P. a. G., 1965, Nature, 207, 963

Seiler J., Hutter A., Sinha M., Croton D., 2019, Monthly Notices of the Royal Astronomical Society, 487, 5739

Sheth R. K., Mo H. J., Tormen G., 2001, Monthly Notices of the Royal Astronomical Society, 323, 1

Sims P. H., Pober J. C., 2020, Monthly Notices of the Royal Astronomical Society, 492, 22

Singh S., Subrahmanyan R., 2019, The Astrophysical Journal, 880, 26

Singh S., et al., 2022, Nature Astronomy, 6, 607

Sinha M., Garrison L. H., 2020, Monthly Notices of the Royal Astronomical Society, 491, 3022

Slipher V. M., 1913, Lowell Observatory Bulletin, 2, 56

Sobacchi E., Mesinger A., 2013, Monthly Notices of the Royal Astronomical Society, 432, 3340

Sobacchi E., Mesinger A., 2014, Monthly Notices of the Royal Astronomical Society, 440, 1662

Somerville R. S., Davé R., 2015, Annual Review of Astronomy and Astrophysics, 53, 51

Somerville R. S., Primack J. R., Faber S. M., 2001, Monthly Notices of the Royal Astronomical Society, 320, 504

Song M., et al., 2016, The Astrophysical Journal, 825, 5

Songaila A., Cowie L. L., 2010, The Astrophysical Journal, 721, 1448

Springel V., et al., 2005, Nature, 435, 629

Springel V., et al., 2018, MNRAS, 475, 676

Stefanon M., Bouwens R. J., Labbé I., Illingworth G. D., Gonzalez V., Oesch P. A., 2021, The Astrophysical Journal, 922, 29

Sutherland R. S., Dopita M. A., 1993, The Astrophysical Journal Supplement Series, 88, 253

Tegmark M., Taylor A. N., Heavens A. F., 1997, The Astrophysical Journal, 480, 22

The HERA Collaboration et al., 2022, Improved Constraints on the 21 cm EoR Power Spectrum and the X-Ray Heating of the IGM with HERA Phase I Observations, https://ui.adsabs.harvard.edu/abs/2022arXiv221004912T

Thompson A. R., Moran J. M., Swenson Jr. G. W., 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition. Springer Cham, https://ui.adsabs.harvard.edu/abs/2017isra.book.....T

Thoul A. A., Weinberg D. H., 1996, The Astrophysical Journal, 465, 608

Tingay S. J., et al., 2013, Publications of the Astronomical Society of Australia, 30, e007

Tinker J. L., Kravtsov A. V., Klypin A., Abazajian K., Warren M. S., Yepes G., Gottlober S., Holz D. E., 2008, The Astrophysical Journal, 688, 709

Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, The Astrophysical Journal, 724, 878

Tolman R. C., 1934, Proceedings of the National Academy of Sciences, 20, 169

Trott C. M., et al., 2020, Monthly Notices of the Royal Astronomical Society, 493, 4711

Van Haarlem M. P., et al., 2013, Astronomy & Astrophysics, 556, A2

Van Rossum G., Drake Jr F. L., 1995, Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam

Ventura E. M., Trinca A., Schneider R., Graziani L., Valiante R., Wyithe J. S. B., 2023, Monthly Notices of the Royal Astronomical Society, 520, 3609

Ventura E. M., Qin Y., Balu S., Wyithe J. S. B., 2024, Monthly Notices of the Royal Astronomical Society, 529, 628

Virtanen P., et al., 2020, Nature Methods, 17, 261

Visbal E., Bryan G. L., Haiman Z., 2020, The Astrophysical Journal, 897, 95

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42

Walker A. G., 1937, Proceedings of the London Mathematical Society, s2-42, 90

Wang F., et al., 2020, The Astrophysical Journal, 896, 23

Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, Monthly Notices of the Royal Astronomical Society, 433, 1230

White S. D. M., Frenk C. S., 1991, The Astrophysical Journal, 379, 52

White S. D. M., Rees M. J., 1978, Monthly Notices of the Royal Astronomical Society, 183, 341

Whitler L. R., Mason C. A., Ren K., Dijkstra M., Mesinger A., Pentericci L., Trenti M., Treu T., 2020, Monthly Notices of the Royal Astronomical Society, 495, 3602

Wold I. G. B., et al., 2022, The Astrophysical Journal, 927, 36

Wouthuysen S. A., 1952, AJ, 57, 31

Wyithe J. S. B., Loeb A., 2003, The Astrophysical Journal, 586, 693

Wyithe J. S. B., Loeb A., 2004, Nature, 432, 194

Yeh J. Y.-C., et al., 2023, Monthly Notices of the Royal Astronomical Society, 520, 2757

York D. G., et al., 2000, The Astronomical Journal, 120, 1579

Zeldovich Y. B., 1972, Monthly Notices of the Royal Astronomical Society, 160, 1P

Zwicky F., 1933, Helvetica Physica Acta, 6, 110

van de Hulst H. C., 1945, Nederlandsch Tijdschrift voor Natuurkunde, 11, 210

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (x11) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu. I also acknowledge Lucy Strang for modifying the template to that of the University of Melbourne.